WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

EFMD EQUIS ACCREDITED

Thomas Salzberger, PD Dr.
thomas.salzberger@wu.ac.at

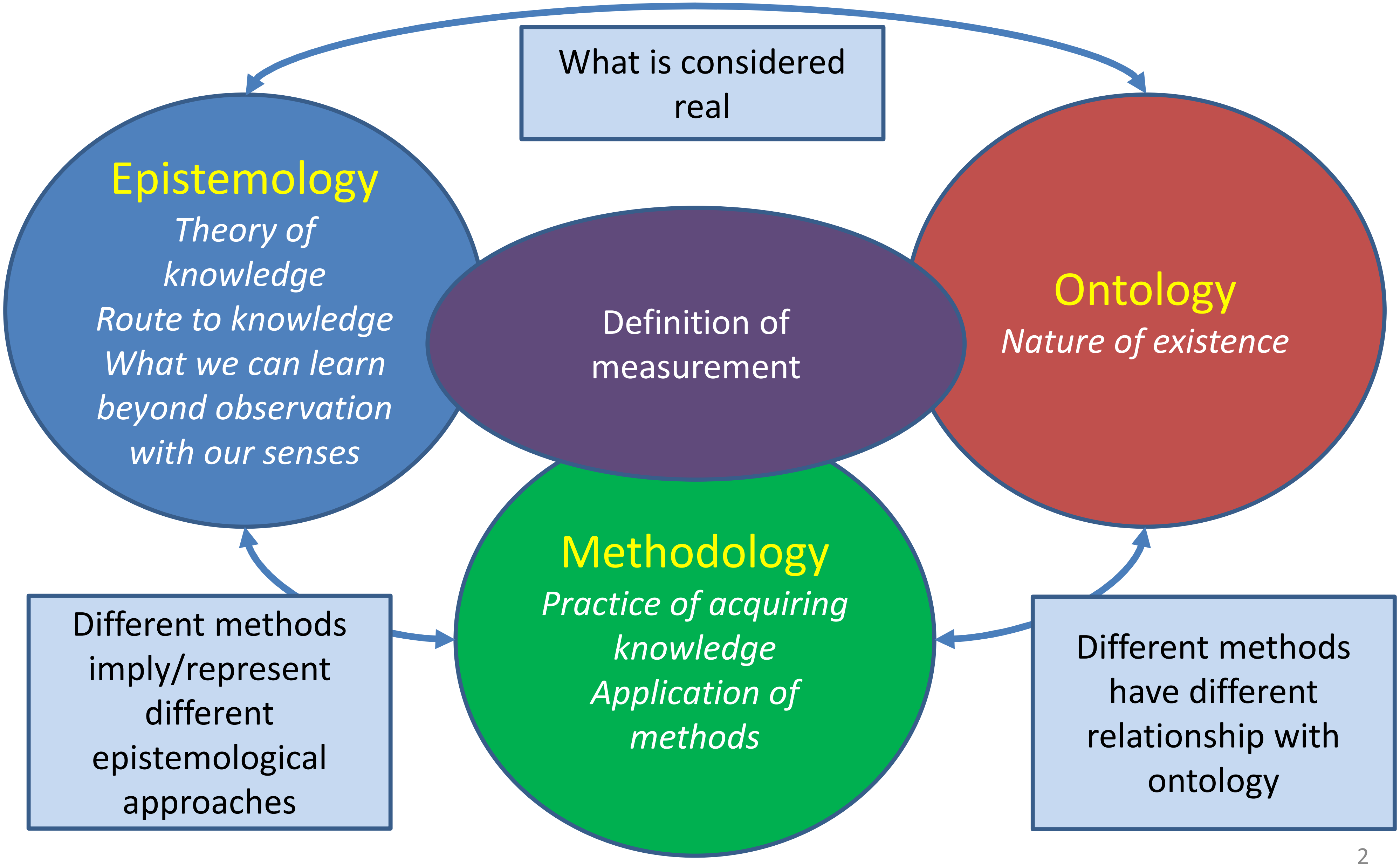
Quantitative Research Methods

Summer term 2023

Measurement

http://statmath.wu.ac.at/courses/m1bw/m1bw_en.html

Measurement



Measurement of some proposed property in social sciences

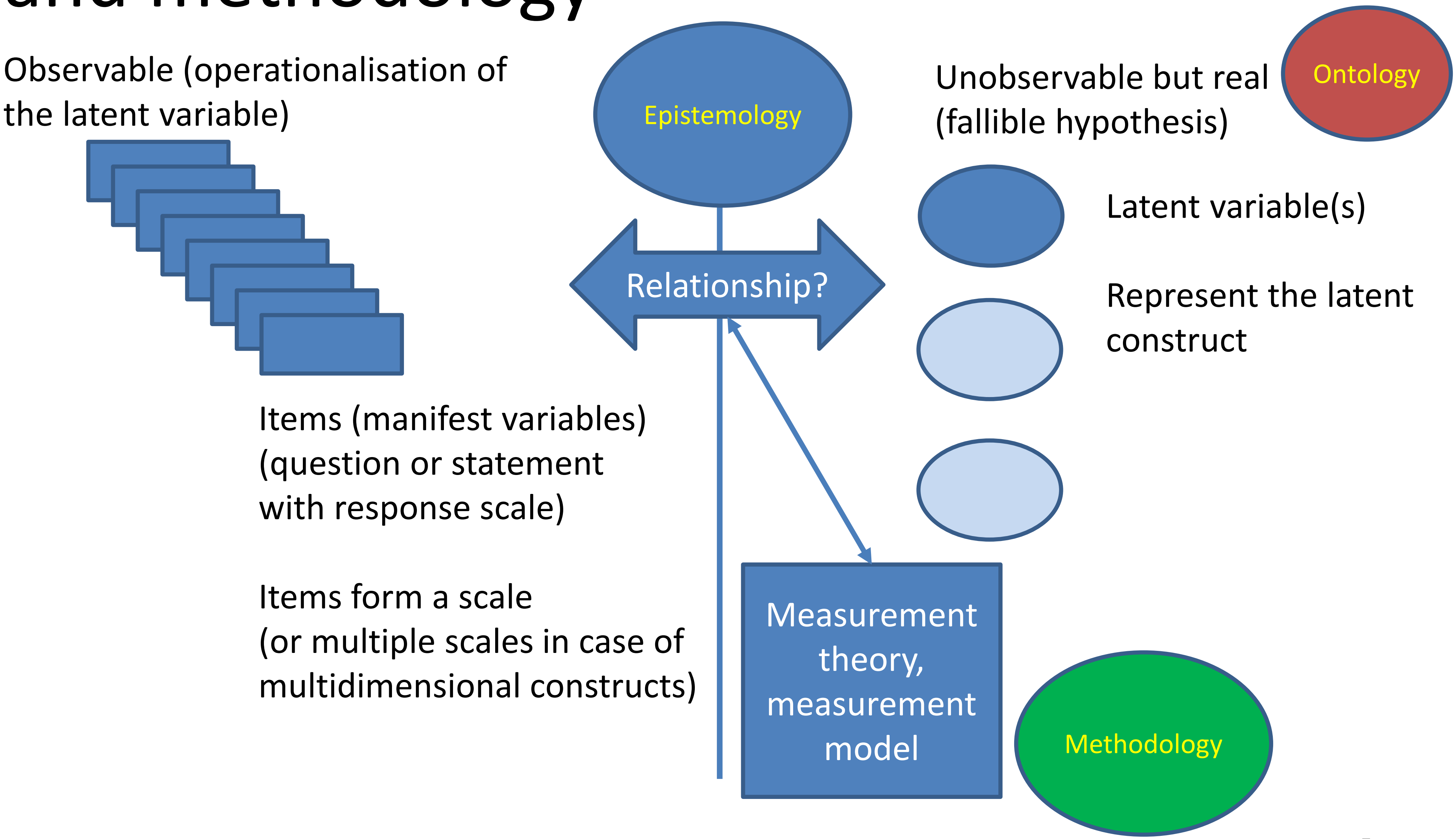
- Quantitative properties
  - More or less, unit of measurement, stable metric
  - Interval scale level (vs. broader concept of "measurement")
- Observable properties (at least in principle)
  - Physical property (e.g., height, blood pressure)
  - Sociodemographic property (e.g., income)
- Unobservable properties
  - Latent properties
    (e.g., constructs such as satisfaction, attitude, trust)

Measurement epistemology

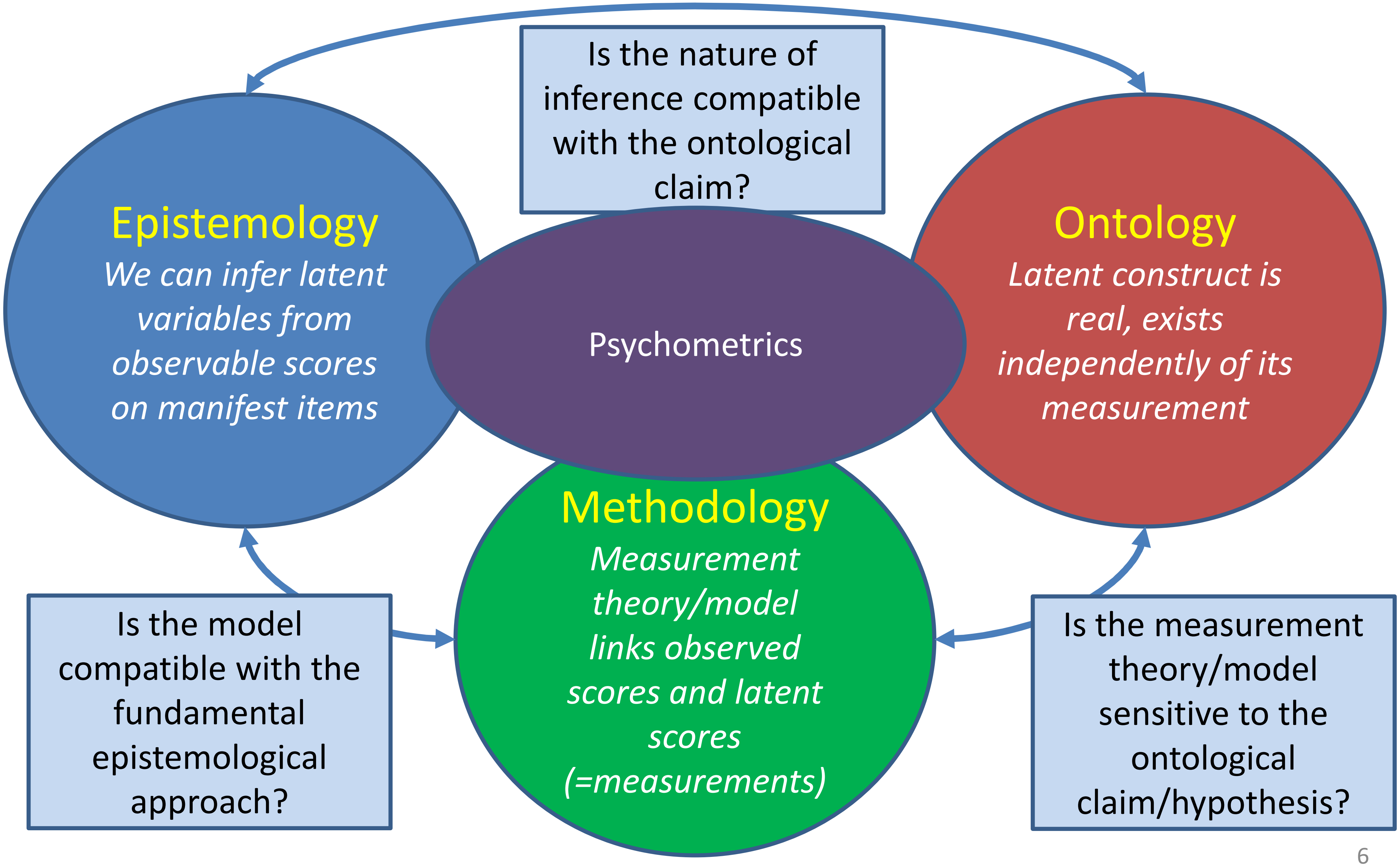
- Observable properties
  - Direct observation (fundamental measurement) or indirect observation (derived measurement)
  - Quantity can be demonstrated by concatenation operation (fundamental measurement)
- Unobservable properties
  - Latent properties cannot be observed directly
  - Require observable manifestations from which we infer the magnitude (value) of the latent property

Measurement epistemology, ontology and methodology

critical realism
The belief that there is an external reality independent of a person's thinking (realism) but that we can never know that reality with perfect accuracy (critical).



Measurement of latent construct



Subjective versus objective

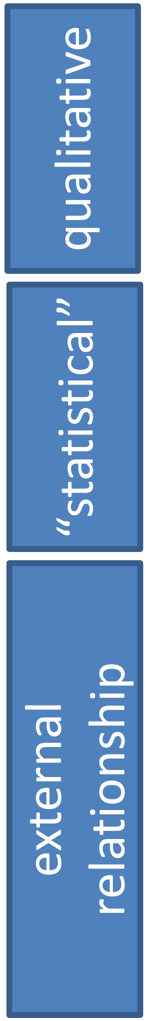
- Observable properties as objective properties
  - Anyone could in principle observe the property directly
- Unobservable properties as subjective properties
  - It arises out of or is identified by means of one's perception of one's own states and processes.
  - Its measurement requires self-perception and active responses to standardized stimuli presented to the respondent.

Subjective versus objective

- Unobservable properties as subjective properties
  - But the proposed construct is considered real, typically of quantitative nature and therefore measurable. (critical realism)
    subjective refers to epistemology, not to ontology
  - The items and the associated response scales (self-report instrument are standardized stimuli thought to be sensitive towards the construct to be measured.
  - The items indicate the latent variable (items as indicators).
  - If the construct is quantitative and the items indicate the construct, it is possible to measure the states by inferring measurements from observed responses.
    Measurement as a hypothesis (fallible, imperfect; critical realism)
  - For such measurements to be expressed in a metric that allows for comparisons across time (repeated measurement), different individuals or groups (mean comparisons), it is pivotal that the items mean the same to all respondents at all times, and that the statistical inference of measurements supposed to be on a linear, interval scale works in the same way for all subjects. In this sense, the measurement ought to be objective.

Assessment of measurement

- Validity (accuracy)
  - Content validity (essentially qualitative) ≠ face validity
  - Construct validity*
    - Structural validity (dimensionality)
    - Fit to measurement model
    - Convergent validity (external)
    - Discriminant validity (external)
    - Known-group validity
  - Criterion-related validity
    - Concurrent validity
    - Predictive validity
  - Other aspects



- Precision
  - Uncertainty when it comes to the inference of the value of the latent variable
  - *Reliability* (consistency of the measure) *is actually related to a particular measurement theory*

* Sometimes construct validity is used as the overarching term.

CHAPTER 3

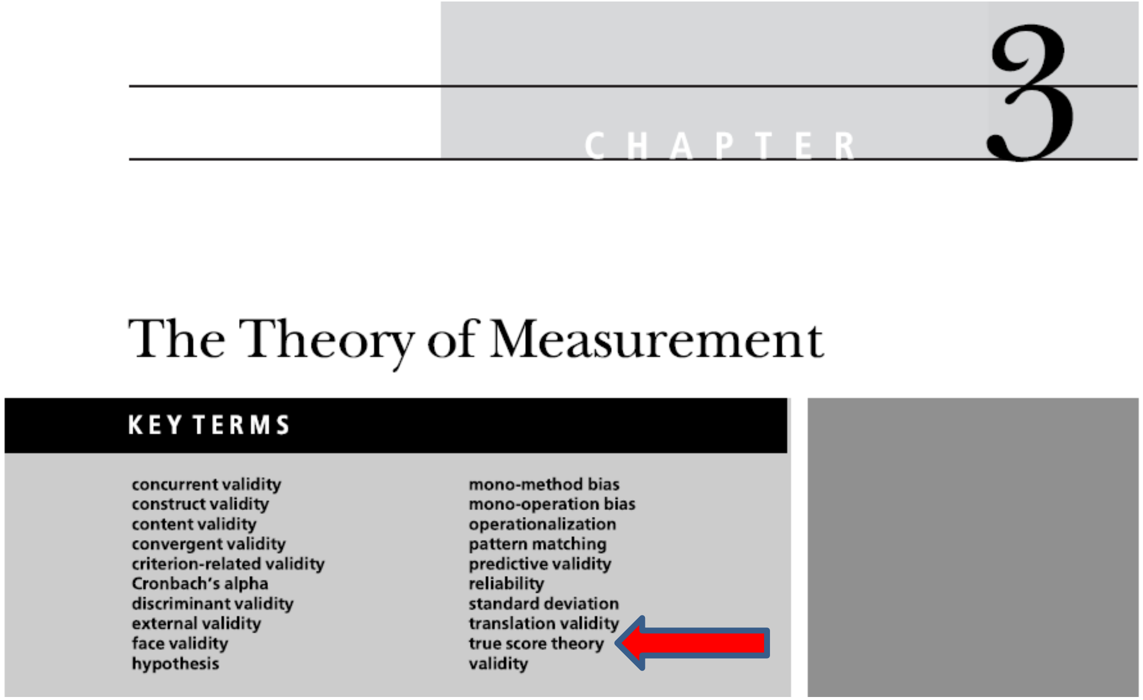
The Theory of Measurement

KEY TERMS

concurrent validity
construct validity
content validity
convergent validity
criterion-related validity
Cronbach's alpha
discriminant validity
external validity
face validity
hypothesis
mono-method bias
mono-operation bias
operationalization
pattern matching
predictive validity
reliability
standard deviation
translation validity
true score theory
validity

- Based on **Classical Test Theory** (CTT; True score theory)
  - Meaningfulness of total score/sum score/mean score
  - Common cause (latent variable) impacting manifest item responses (responses reflect latent variable) – ***latent variable theory***
  - Likert scaling compatible with CTT (→ Topic 5)
  - Factor analysis (congeneric model)
  - Other measurement theories (→ Topic 5: Thurstone's equal appearing interval, Guttman; Item Response Theory, Rasch model)
  - Other concepts of measurement (index formation → Topic 5)
  - CTT still the predominant paradigm in business research
  - Other theories may be based on a different definition of measurement

R. Likert

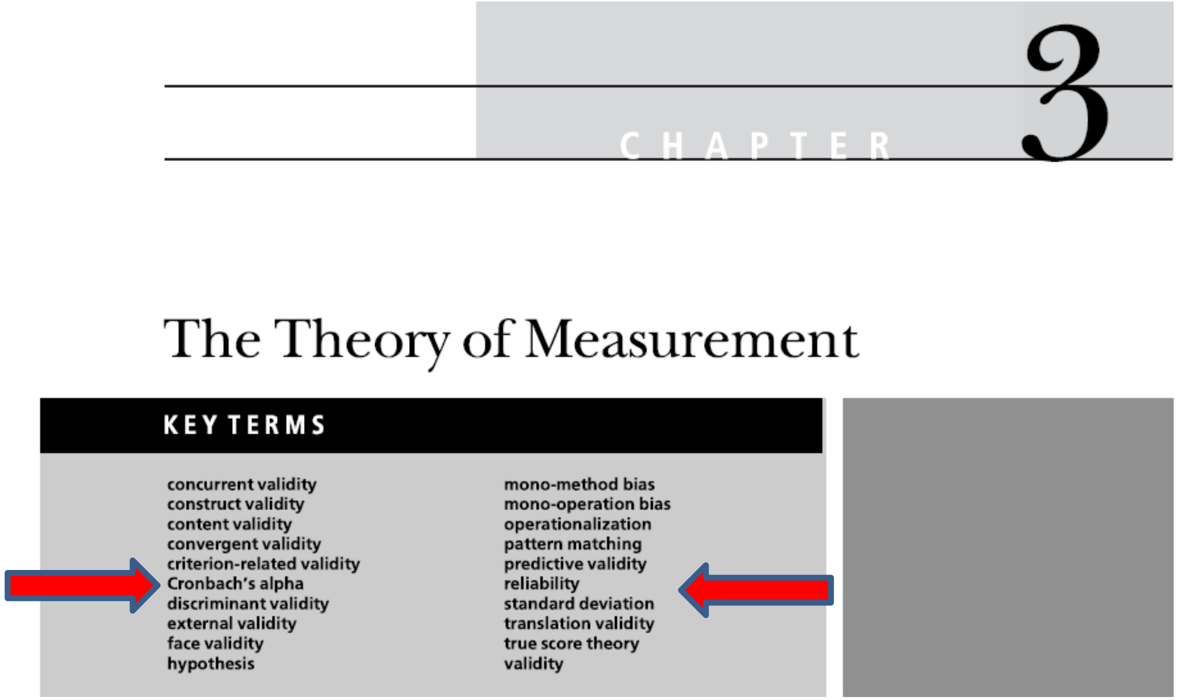

CHAPTER 3

The Theory of Measurement

KEY TERMS

concurrent validity	mono-method bias
construct validity	mono-operation bias
content validity	operationalization
convergent validity	pattern matching
criterion-related validity	predictive validity
Cronbach's alpha	reliability
discriminant validity	standard deviation
external validity	translation validity
face validity	true score theory
hypothesis	validity

- CTT: true score T, observed score X, error score E
  - $X_v = T_v + E_v$ (at the total score level for person v)
  - $x_{iv} = \tau_i + \lambda_i * F_v + e_{iv}$ (at the item level for item i; latent variable/factor F)


3

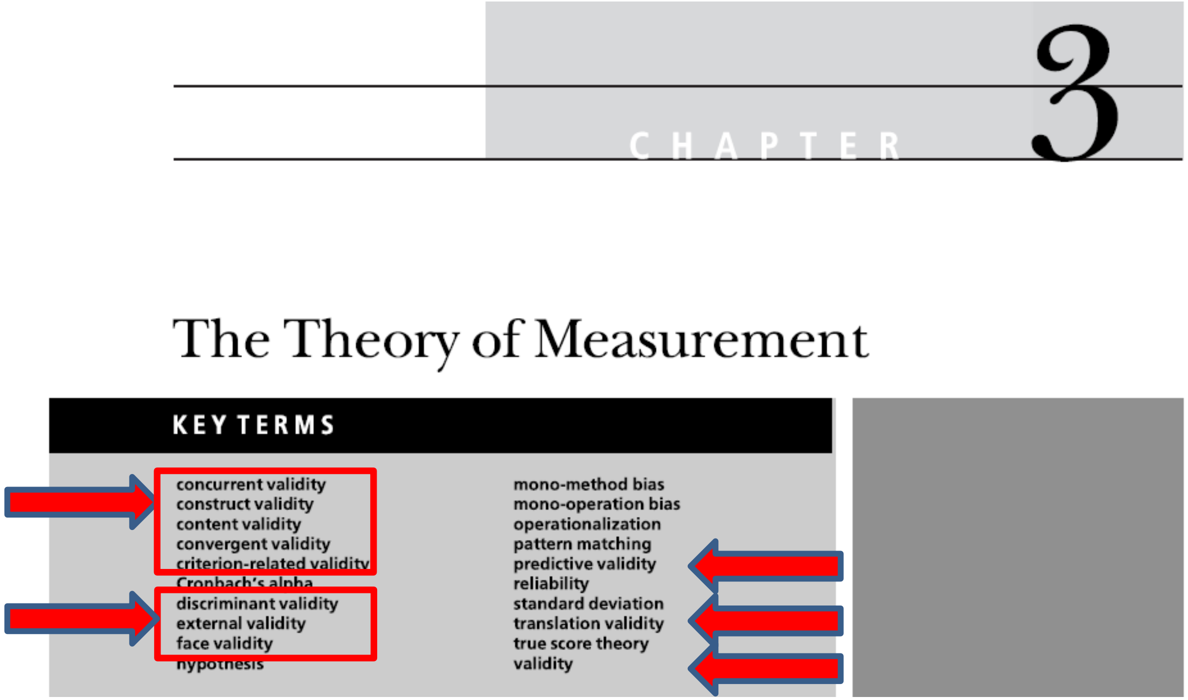
CHAPTER

The Theory of Measurement

KEY TERMS

concurrent validity
construct validity
content validity
convergent validity
criterion-related validity
Cronbach's alpha
discriminant validity
external validity
face validity
hypothesis
mono-method bias
mono-operation bias
operationalization
pattern matching
predictive validity
reliability
standard deviation
translation validity
true score theory
validity

- Some concepts are, by their nature, CTT concepts
  - Reliability (hence also Cronbach's α):
    true score variance / total score variance
  - Often considered a measure of precision
    - Depends on error variance - but also depends on true score variance
    - Internal consistency
    - Can be used to estimate standard error of measurement SEM (uncertainty), but SEM depends on sample properties
    - $SEM = s\sqrt{1 - Rel}$ (s = standard deviation in the sample)
  - Reliability as defined can also be applied outside CTT but has a different meaning (person-separation in Rasch modeling)
    - CTT as an error distribution theory



3

CHAPTER

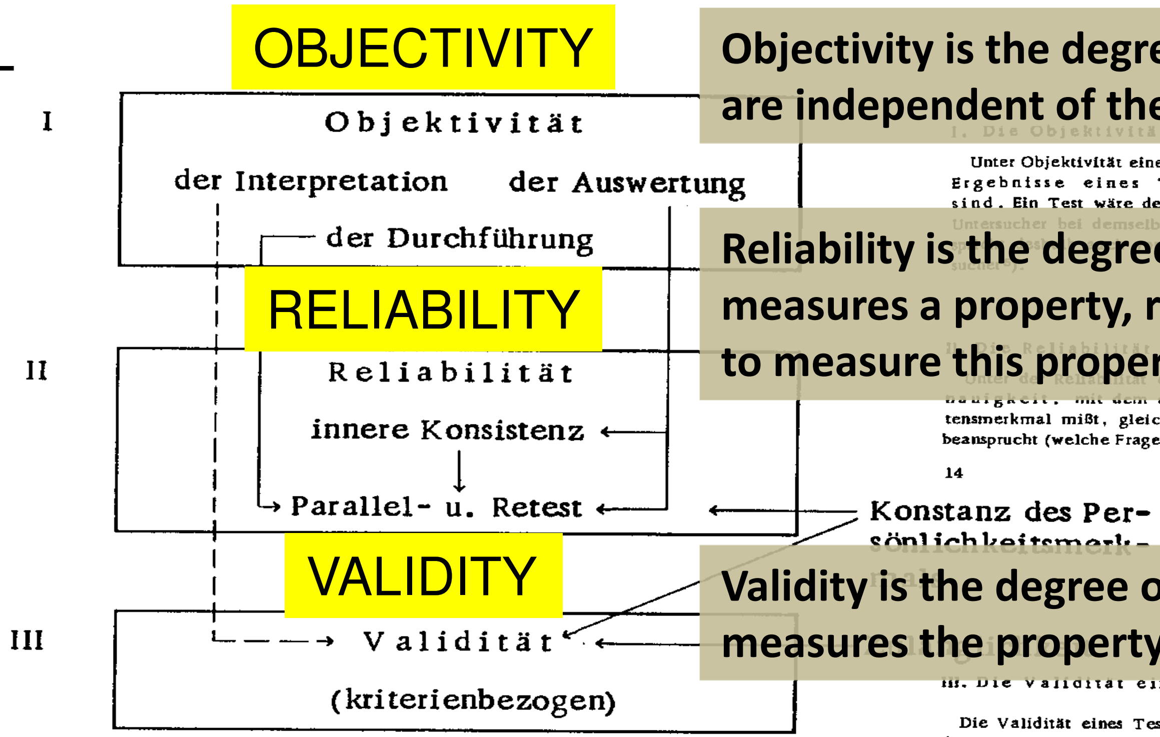
The Theory of Measurement

KEY TERMS

concurrent validity
construct validity
content validity
convergent validity
criterion-related validity
Cronbach's alpha
discriminant validity
external validity
face validity
hypothesis
mono-method bias
mono-operation bias
operationalization
pattern matching
predictive validity
reliability
standard deviation
translation validity
true score theory
validity

- Other concepts are universal: Validity
  - What is validity?
  - How does it differ from reliability?

Quality Criteria [Gütekriterien] according to Lienert (1989)



Objectivity is the degree to which the results of a test are independent of the examiner (investigator)

I. Die Objektivität eines Testes

Unter Objektivität eines Testes verstehen wir den Grad, in dem die Ergebnisse eines Testes unabhängig vom Untersucher sind. Ein Test wäre demnach vollkommen objektiv, wenn verschiedene Untersucher bei demselben Pbn zu gleichen Ergebnissen gelangten.

Reliability is the degree of precision with which the test measures a property, regardless of whether the test is to measure this property

II. Die Reliabilität eines Testes

tensmerkmal mißt, gleichgültig, ob er dieses Merkmal auch zu messen beansprucht (welche Frage ein Problem der Validität ist).

14

Validity is the degree of accuracy with which the test measures the property the test claims to measure

III. Die Validität eines Testes

Die Validität eines Testes gibt den Grad der Genauigkeit an, mit dem dieser Test dasjenige Persönlichkeitsmerkmal oder diejenige Verhaltensweise, das (die) er messen soll oder zu messen vorgibt, tatsächlich mißt.

Ein Test ist demnach vollkommen valide, wenn seine Ergebnisse einen unmittelbaren und fehlerfreien Rückschluß auf den Ausprägungsgrad des zu erfassenden Persönlichkeits- oder Verhaltensmerkmals zulassen, wenn also der individuelle Testpunktwert eines Pb diesen auf der Merkmalsskala eindeutig lokalisiert.

B. DIE GÜTEKRITERIEN EINES TESTES

Ein guter Test soll als Hauptgütekriterien folgende drei Forderungen erfüllen:

1. er soll objektiv,
2. er soll reliabel,
3. er soll valide sein.[4])

Daran schließen sich vier Nebengütekriterien als bedingte Forderungen:

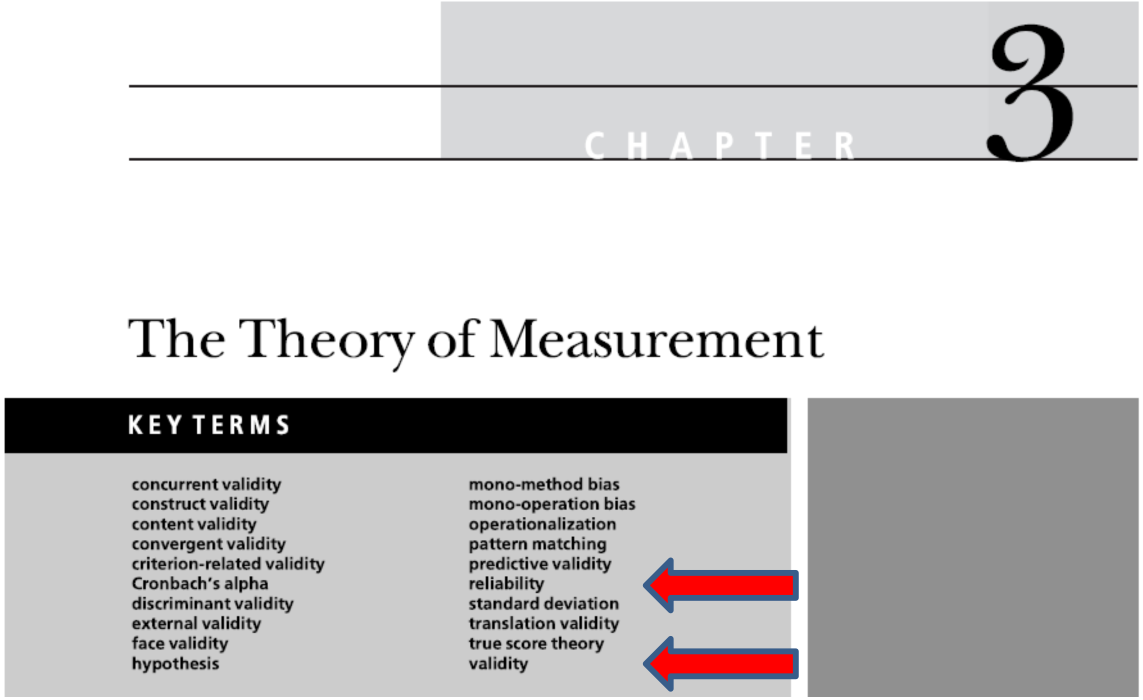
4. er soll normiert,
5. er soll vergleichbar,
6. er soll ökonomisch,
7. er soll nützlich sein.

CIP-Titelaufnahme der Deutschen Bibliothek

Lienert, Gustav A.:
Testaufbau und Testanalyse / Gustav A. Lienert. - 4., neu ausgestattete Aufl. - München : Psychologie-Verl.-Union, 1989
ISBN 3-621-27086-8

Gesamtherstellung: Offsetdruckerei Julius Beltz, Weinheim/Bergstr.
Printed in Germany

ISBN 3-621-27086-8



CHAPTER 3

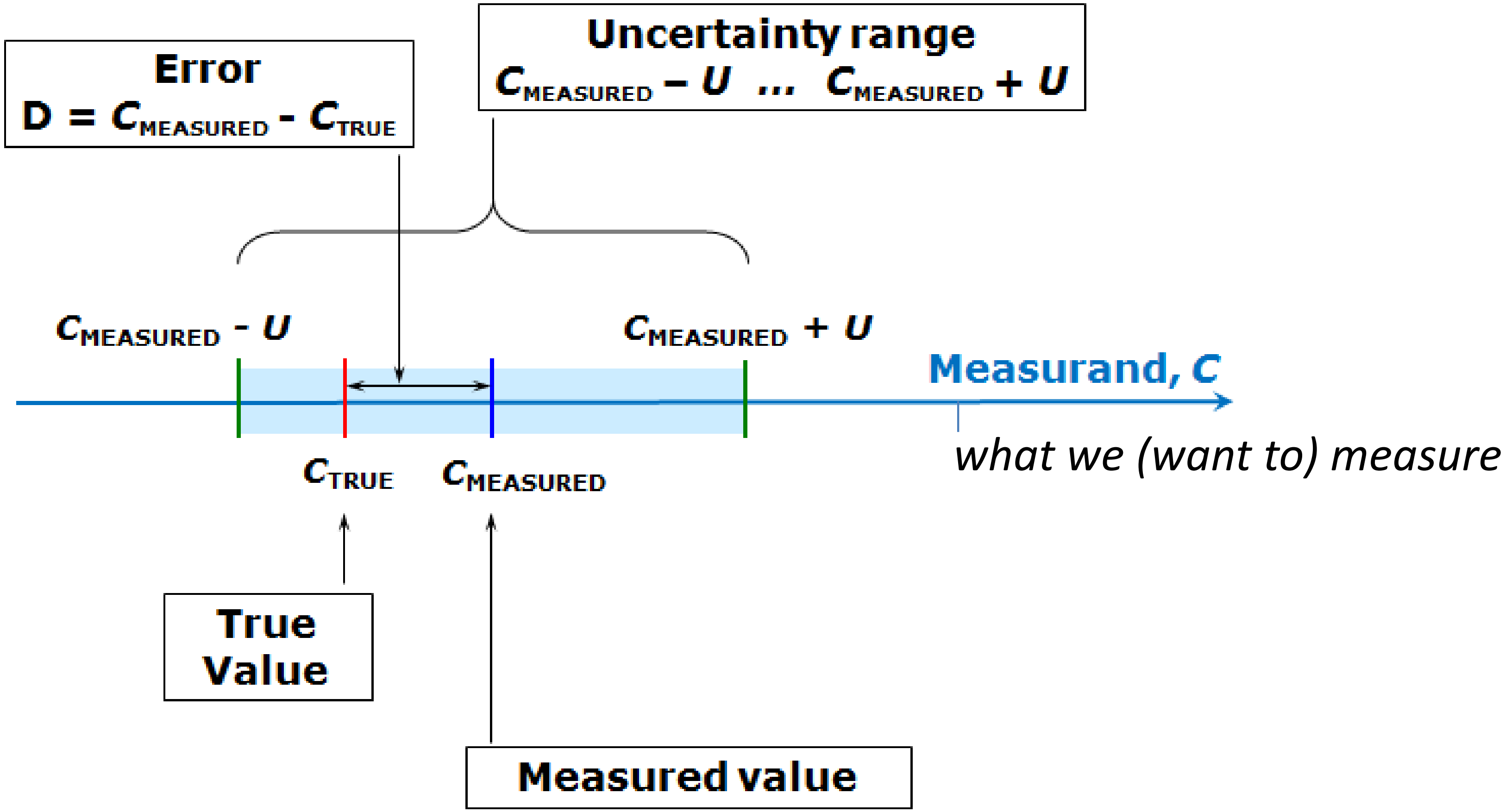
The Theory of Measurement

KEY TERMS

concurrent validity
construct validity
content validity
convergent validity
criterion-related validity
Cronbach's alpha
discriminant validity
external validity
face validity
hypothesis
mono-method bias
mono-operation bias
operationalization
pattern matching
predictive validity
reliability
standard deviation
translation validity
true score theory
validity

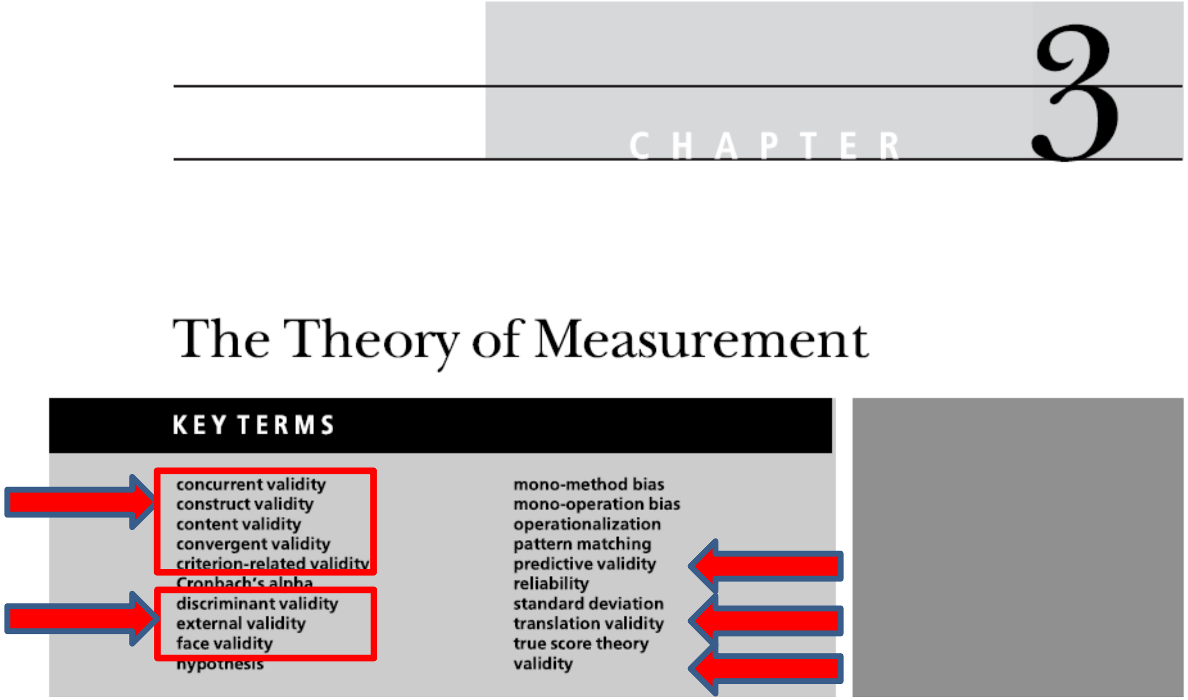
- Other concepts are universal: Validity
  - What is validity?
    - How well/accurately ... → validity comes in degrees
    - Alternatively: validity is whether the test/scale/instrument measures what it is supposed to measure (*think of the ontological claim!*)
  - How does it differ from reliability?
    - Reliability is independent of the content of the items
    - Reliability does not require validity (*but is it meaningful in any way if there is no validity?*)
    - But validity (assessed in degrees) requires reliability (in CTT)

Difference between measurement error (=error score) and standard error of measurement (or rather, in terms of metrology, measurement uncertainty)



Scheme 1.1. Interrelations between the concepts true value, measured value, error and uncertainty.

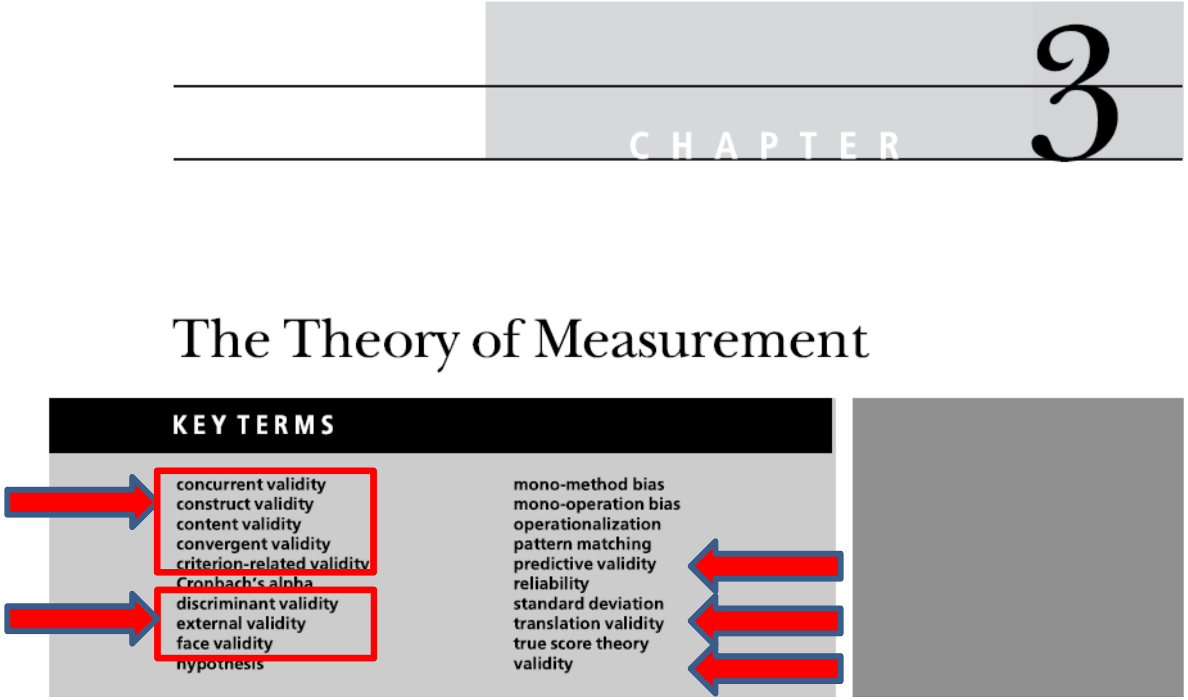
https://sisu.ut.ee/measurement/introduction-concept-measurement-uncertainty



- Validity

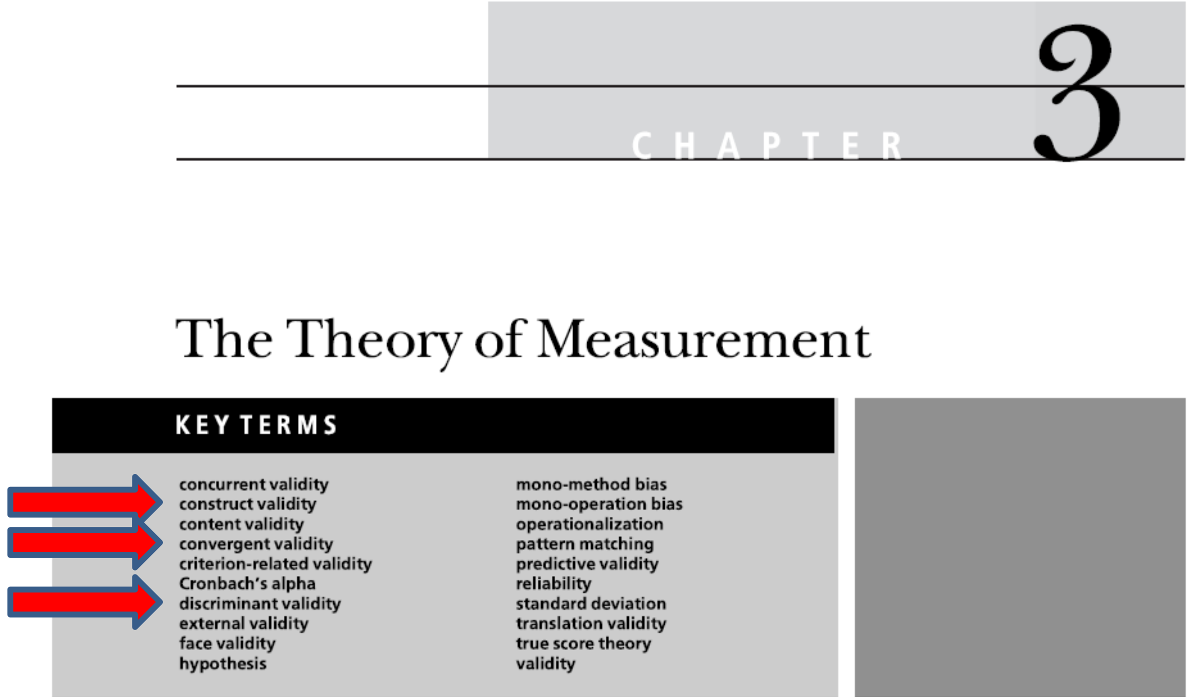
Trochim calls this the definitionalist view

  - On the one hand very simple (we measure what we claim to measure) – and what must exist independently of our measurement (realism; ontological claim)
  - On the other hand, extremely complex
    - when it comes to assessing validity empirically
    - multiplicity of aspects of validity
    - different understanding of what ultimately constitutes validity



- First account of validity credited to Kelley (1927, p.16)
  - "A test is valid if it measures what it purports to measure" (Kelley , 1927, p.16)
  - Focus on content validity and interpretation of test scores
- Later, focus on criterion-related validity *Trochim calls this the relationalist view*
  - Besides a content model providing the basis for scale development
  - High correlation with a criterion Y (r_{XY}) (concurrent, predictive validity)
- Construct validity (Cronbach and Meehl, 1955)
  - When no adequate criterion (gold standard) is available
  - Content validity, concurrent/predictive validity (criterion), construct validity

Kelley, T. L. (1927). *Interpretation of educational measurements*. New York: MacMillan.
Cronbach, L. J., & Meehl, P. E. *Construct validity in psychological tests.* Psychological Bulletin, 52(4), 281, 1955.



CHAPTER 3

The Theory of Measurement

KEY TERMS

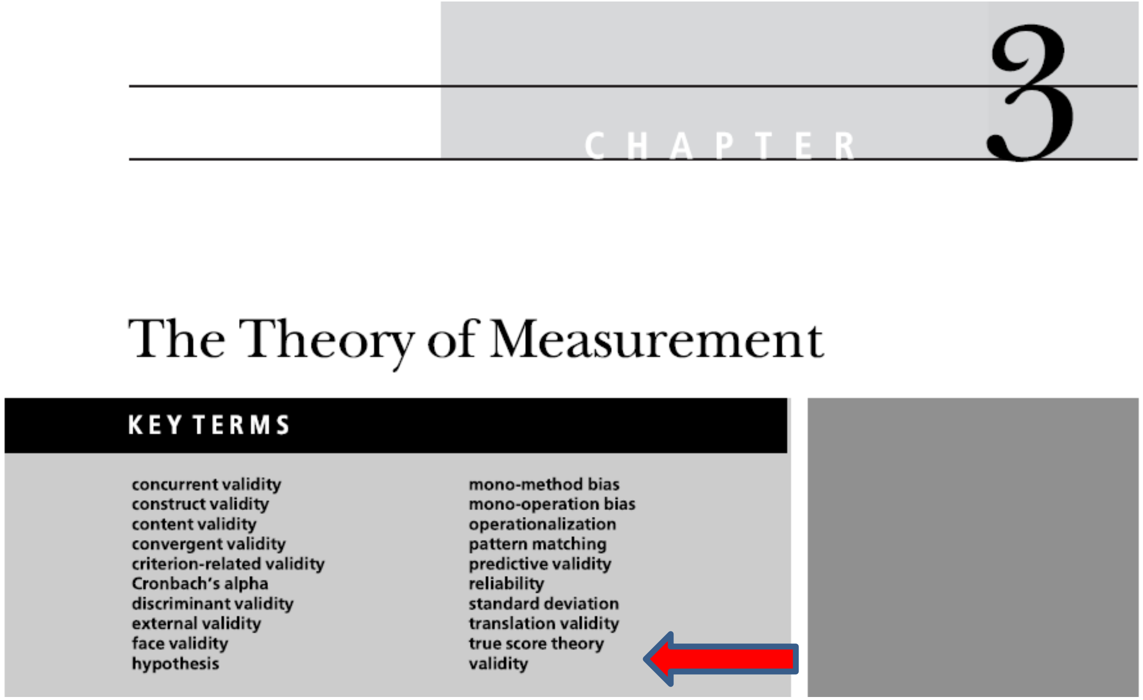
concurrent validity
construct validity
content validity
convergent validity
criterion-related validity
Cronbach's alpha
discriminant validity
external validity
face validity
hypothesis
mono-method bias
mono-operation bias
operationalization
pattern matching
predictive validity
reliability
standard deviation
translation validity
true score theory
validity

- Construct validity at the core
  - Content, concurrent and predictive ad hoc (Loevinger, 1957)
  - *Content validity properly understood and addressed, and ideally linked to construct validity, is of utmost importance*
  - Messick (1989, p.13) defines construct validity as "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other models of assessment" – *expands validity to the use of measurements (consequences, social implications) - controversial*

S. Messick

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3*(3), 635-694.

Messick, S. *Validity*. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). (New York: American Council on Education), 3rd ed., pp. 13-103, 1989.


3

CHAPTER

The Theory of Measurement

KEY TERMS

concurrent validity
construct validity
content validity
convergent validity
criterion-related validity
Cronbach's alpha
discriminant validity
external validity
face validity
hypothesis
mono-method bias
mono-operation bias
operationalization
pattern matching
predictive validity
reliability
standard deviation
translation validity
true score theory
validity

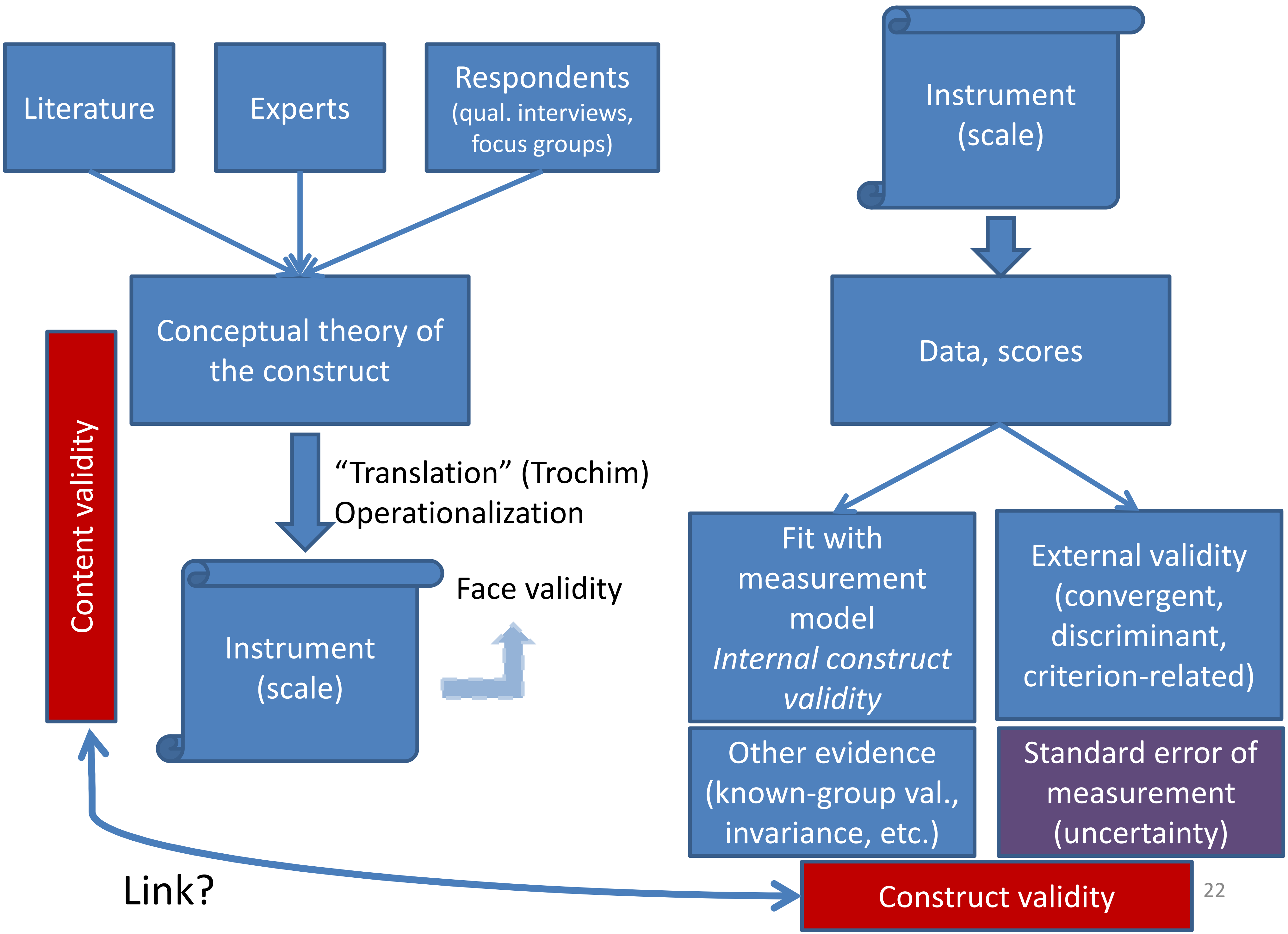
- Aspects of (construct) validity as a unified concept according to Messick (1994)
  - Construct representation (can be too narrow – underrepresentation; can be too broad – construct-irrelevant variance, e.g. reading comprehension in a math test)
  - Aspects: content, substantive, structural, generalizability, external, consequential

Messick, Samuel. "VALIDITY OF PSYCHOLOGICAL ASSESSMENT: VALIDATION OF INFERENCES FROM PERSONS'RESPONSES AND PERFORMANCES AS SCIENTIFIC INQUIRY INTO SCORE MEANING." *ETS Research Report Series* 1994.2 (1994): i-28.

Aspects of (construct) validity as a unified concept according to Messick (1994)

- Content: relevance, representativeness
  - *Think also: concept of interest – context of use*
- Substantive: process-model, how respondents reach a conclusion/response
  - *Why do items work the way they do*
  - *Explanatory model, construct specification equation, concept-driven measurement*
- Structural: "fidelity of the scoring structure to the structure of the construct domain"
  - *Multi-domain constructs (multi-dimensional), proper reflection of dimensional structure*
- Generalizability: extent to which score properties and interpretations generalize to and across population groups and settings
  - *Comparability, fairness, invariance, cross-cultural equivalence*
- External (*i.e. including other measure(ment)s*): convergent and discriminant validity, criterion-related validity, including criterion relevance
  - *Again: concept of interest – context of use*
- Consequential: implications; bias, fairness, distributive judgement

Own comments in italics



Literature
Experts
Respondents
(qual. interviews, focus groups)
Conceptual theory of the construct
Content validity
"Translation" (Trochim)
Operationalization
Instrument (scale)
Face validity
Link?
Instrument (scale)
Data, scores
Fit with measurement model
Internal construct validity
External validity (convergent, discriminant, criterion-related)
Other evidence (known-group val., invariance, etc.)
Standard error of measurement (uncertainty)
Construct validity

Definition of measurement

- Definition of measurement in the social sciences?

SCIENCE

Vol. 103, No. 2684 Friday, June 7, 1946

On the Theory of Scales of Measurement

S. S. Stevens

Director, Psycho-Acoustic Laboratory, Harvard University

A CLASSIFICATION OF SCALES OF MEASUREMENT

Paraphrasing N. R. Campbell (Final Report, p. 340), we may say that measurement, in the broadest sense, is defined as the assignment of numerals to objects or events according to rules. The fact that numerals can be assigned under different rules leads to different kinds of scales and different kinds of measurement. The problem then becomes that of making explicit (a) the various rules for the assignment of numerals, (b) the mathematical properties (or group structure) of the resulting scales, and (c) the statistical operations applicable to measurements made with each type of scale.

Measurement by assignment

Scale levels Permissible statistics

p.677

Definition of measurement

- Definition of measurement in the social sciences:

SCIENCE

Vol. 103, No. 2684 Friday, June 7, 1946

On the Theory of Scales of Measurement

S. S. Stevens

Director, Psycho-Acoustic Laboratory, Harvard University

Most liberal? Measurement is a very rigorous concept that requires an equally rigorous definition, not a liberal one

To the British committee, then, we may venture to suggest by way of conclusion that the most liberal and useful definition of measurement is, as one of its members advised, "the assignment of numerals to things so as to represent facts and conventions about them."

p.680

Reference to *Representational Measurement Theory*

"isomorphism between numeral series and what we can do with the objects"

(in fact, Stevens' definition boils down to Operationalism, also called operationism)

Scaling and indexes

SCIENCE

Vol. 103, No. 2684 — Friday, June 7, 1946

On the Theory of Scales of Measurement

S. S. Stevens

Director, Psycho-Acoustic Laboratory, Harvard University

- Stevens' measurement by (merely) assigning numerals allows for various, and very different, attempts at measurement
  - Scaling: reflective indicator model, common cause (latent factor/variable) explains observed responses (correlations)
  - Index formation: formative indicator model, set of indicators (items) define/form/create a (purportedly) latent variable
- *Beware: some scales are called an Index, and some indices are called scales ..., others are called inventories …*
- *Sometimes it is just semantics, sometimes is misspecification*

Scaling (CTT) – *e.g. Likert Scaling*

CTT: X = T + E (observed score = true score + error score)

T = E(X)

Adding up all item scores to get the test score X (Likert)

- CTT does not explain how measurement is achieved
- No latent variable referred to in the equation above
- Error distribution theory (reliability=r^2_{XT})

- Different CTT models* (k items; i subjects):

Parallel: $X_{ik} = T_i + E_i$

Tau-Equivalent: $X_{ik} = T_i + E_{ik}$

Essentially Tau-Equivalent: $X_{ik} = (\alpha_K + T_i) + E_{ik}$

Congeneric: $X_{ik} = [\alpha_k + \beta_k(T_i)] + E_{ik}$

Cronbach‘s α is only a lower bound estimate of reliability!

* Birney, D. P., Beckmann, J. F., Beckmann, N., & Stemler, S. E. (2022). Sophisticated Statistics Cannot Compensate for Method Effects If Quantifiable Structure Is Compromised. *Frontiers in psychology.*, *13*(812963).

Scaling (CTT) – Congeneric model

CTT: $X = T + E$ (observed score = true score + error score)

$T = E(X)$

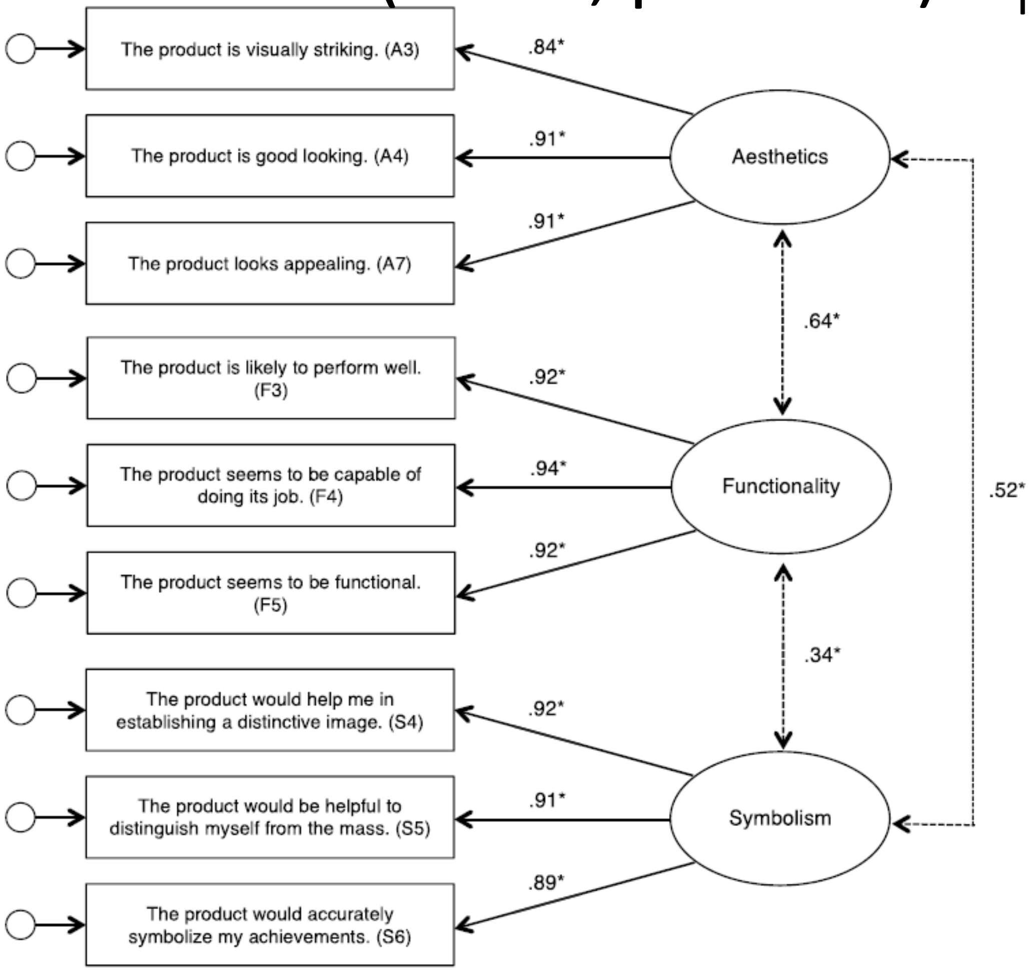
$$X_{ik} = [\alpha_k + \beta_k(T_i)] + E_{ik}$$

CTT at the item level (item i, person v): $x_{iv} = \tau_i + \lambda_i * F_v + e_{iv}$

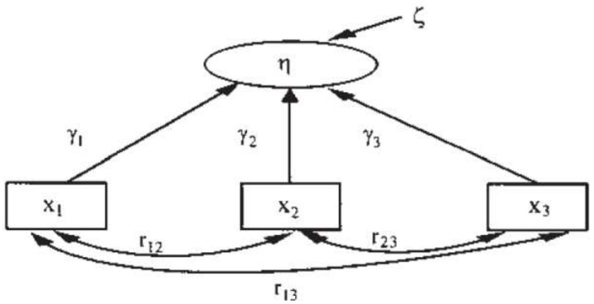
- Congeneric model, factor analytic model
- Maximum adaptation of model to the data
  fitting the model to the data (statistical approach)
- Latent variable F, which is the common cause for item responses
- Reflective indicator model (item responses reflect/are caused by the latent variable), “scale”
- λ_i can be estimated, test of fit (confirmatory factor analysis)

Scaling (CTT)

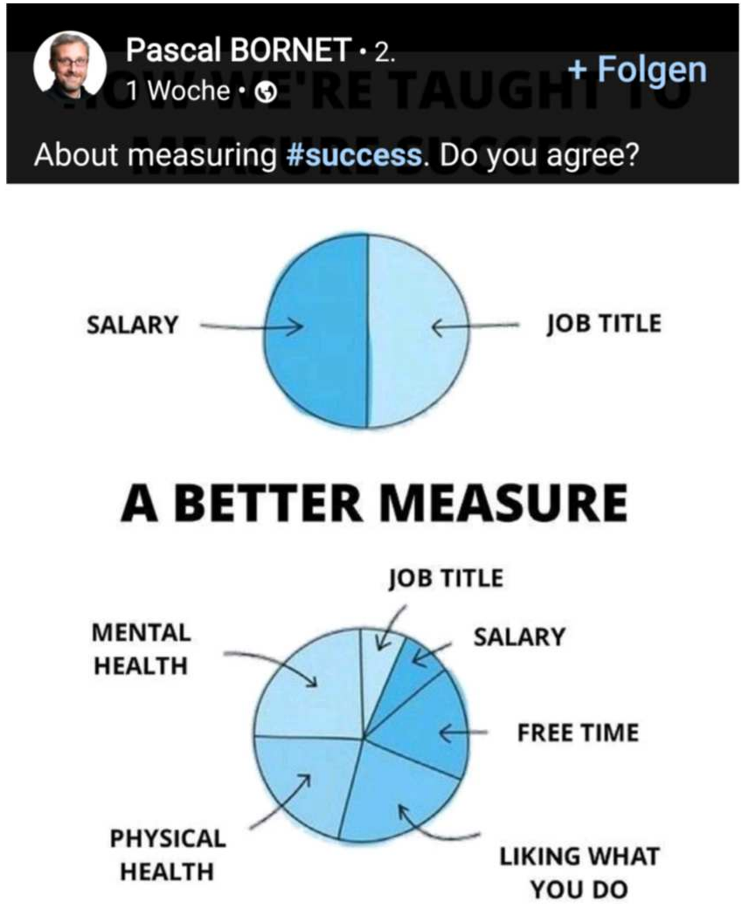
CTT at the item level (item i, person v): $x_{iv} = \tau_i + \lambda_i * F_v + e_{iv}$



Homburg, C., Schwemmle, M., & Kuehnl, C. (2015). New product design: Concept, measurement, and consequences. *Journal of Marketing*, *79*(3), 41-56.



Index



Another branch of measurement (Trochim)

Index: Formative indicator model (items define the latent variable): $\eta_v = \gamma_i{*}x_{iv} + \gamma_j{*}x_{jv} + \ldots + \gamma_k{*}\, x_{kv}$

– pure operationalism (social constructivism)
– no need for item intercorrelations (multicollinearity!)
– γ cannot be estimated (unless there are also reflective indicators available [MIMIC model] or further dependent latent variables with reflective indicators)
– measurement error?

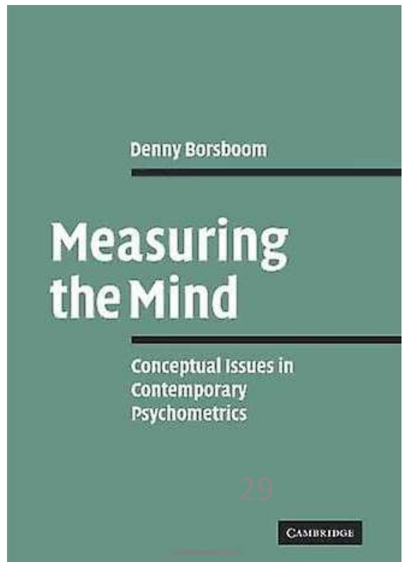
Measurement of a latent variable from a realist perspective

– *Ontological claim: latent variable exists as a quantitative property independent of our measurement*

Index/formative model: latent variable is defined by the indicators

Any such definition creates a “latent“ variable

– No latent variable from a realist perspective
Incompatible with the idea of a latent variable (Borsboom, 2005)
– Index as a summary of measurements (useful for some purpose)



Formative Model: Theoretical foundation

Strict operationalism (concept measured is defined in terms of how we measure it)

The origins of the formative perspective can be traced back to the "operational definition" model. Under strict operationalism, "a concept becomes its measure and has no meaning beyond that measure.… [T]he entire meaning of a theoretical concept is assigned to its measurement and any theoretical concept has one and only one measurement" (Bagozzi 1982, p. 15). Thus, if η represents the concept (i.e., latent variable) in question and x is an empirical measure (i.e., observed or manifest variable), then

$$\eta \equiv x. \tag{1}$$

- Diamantopoulos, A., & Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of marketing research*, *38*(2), 269-277.
- 1960s/1970s: Single-Item-Measurement
- Revival? Rossiter's C-OAR-SE approach (2001)

Formative Model: Theoretical foundation

- Construct defined by its measurements
  (so, item scores are measurements!)

A more contemporary view, which allows the possibility of multiple measures, x_i, ($i = 1, 2, ..., n$), suggests that "a concept is assumed to be defined by, or to be a function of, its measurements" (Bagozzi and Fornell 1982, p. 34).

Formative Model: Theoretical foundation

According to this latter definition, a formative specification implies the following relationship:

(2) $$\eta = \gamma_1 x_1 + \gamma_2 x_2 + \ldots + \gamma_n x_n,$$

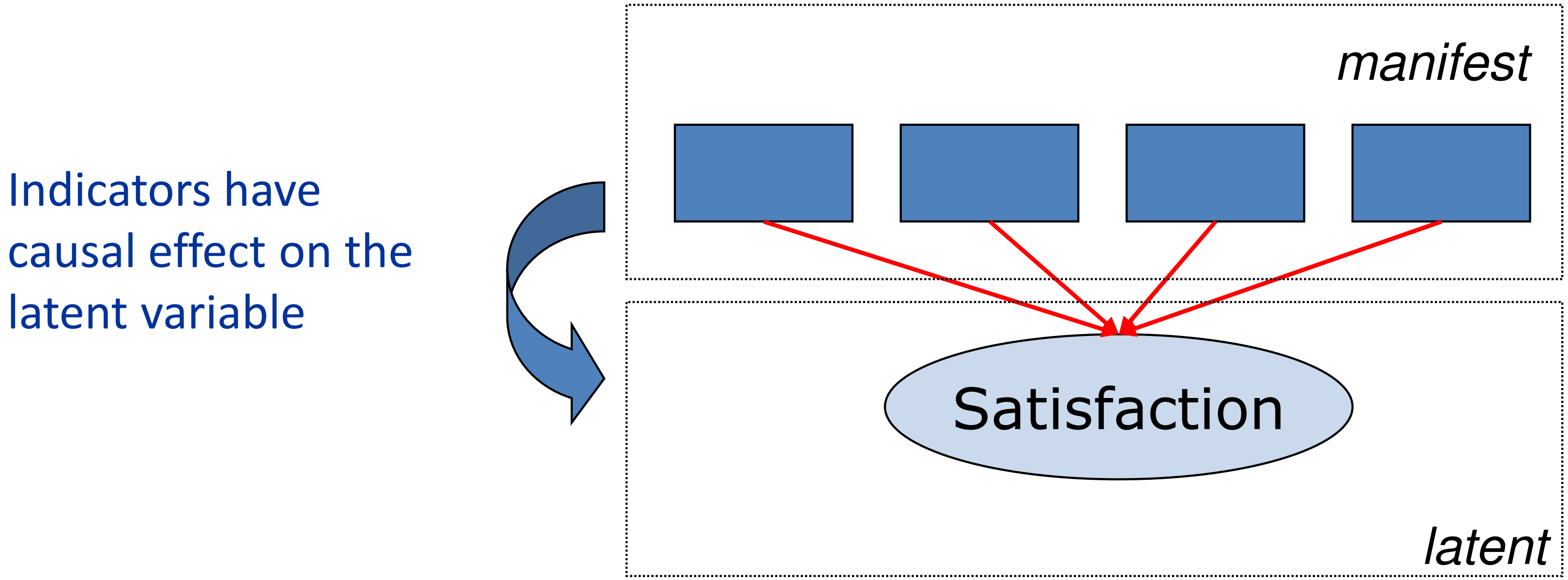
where γ_i is a parameter reflecting the contribution of x_i to the latent variable η.

Another formative specification (shown in Figure 1 for n = 3) is provided by Bollen and Lennox (1991, p. 306):

(3) $$\eta = \gamma_1 x_1 + \gamma_2 x_2 + \ldots + \gamma_n x_n + \zeta.$$

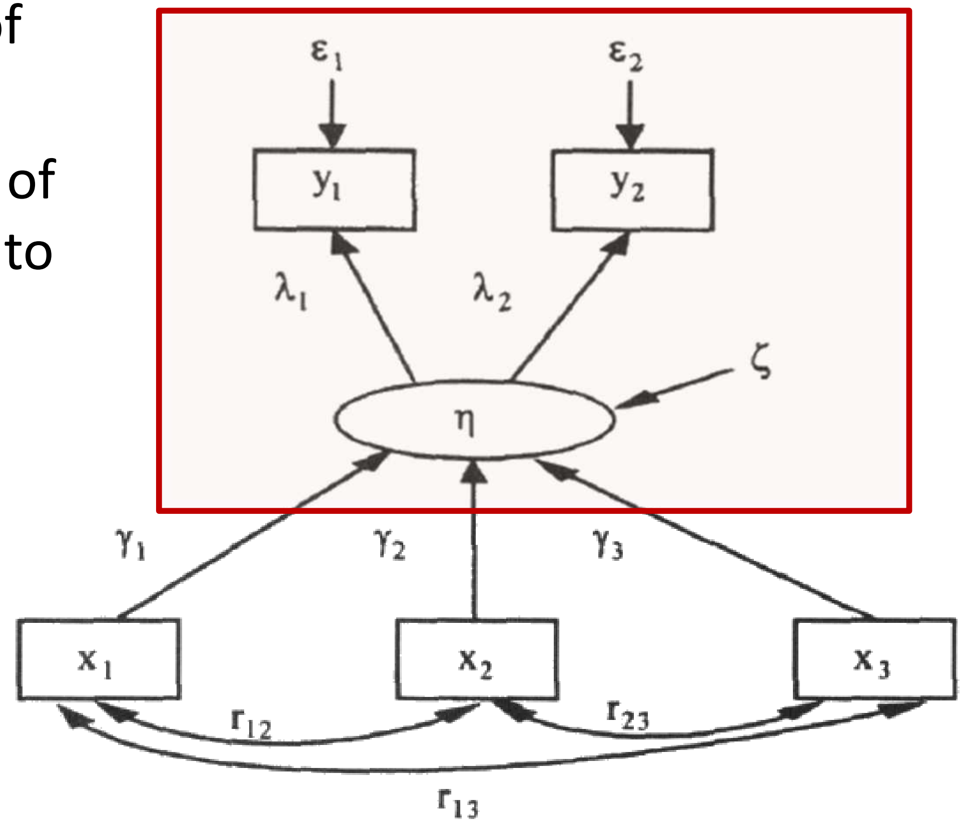
Formative Model

- Landmark article by Diamantopoulos & Winklhofer (2001)
- Scale development (reflective indicators)
- Alternative: index construction
  - Items as formative indicators
  - Indicators have causal effect on latent variable (reversed causality)



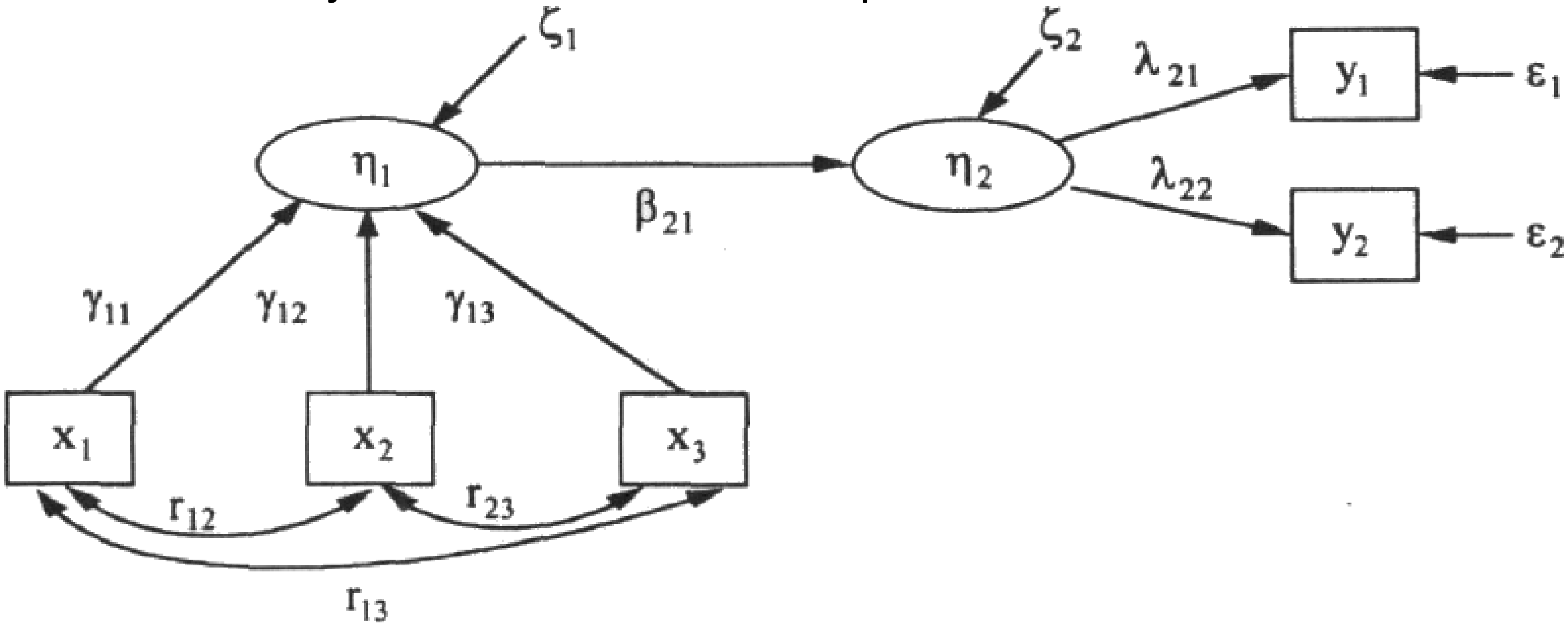
Estimation of Formative Model: MIMIC-Model

- MIMIC: Multiple Indicators Multiple Causes
  - Combination of (reflective) indicators (y) and (formative) “causes" (x)
  - Fit of whole model evidence of validity of γ-coefficients
  - MIMIC-Model solves problem of identification and contributes to validation



Estimation of Formative Model: Identification via Consequential Constructs

- Alternative to MIMIC-Modell
- Dependent construct(s), measured by reflective indicators
  - Two-indicator rule
  - Theoretical justification of relationship

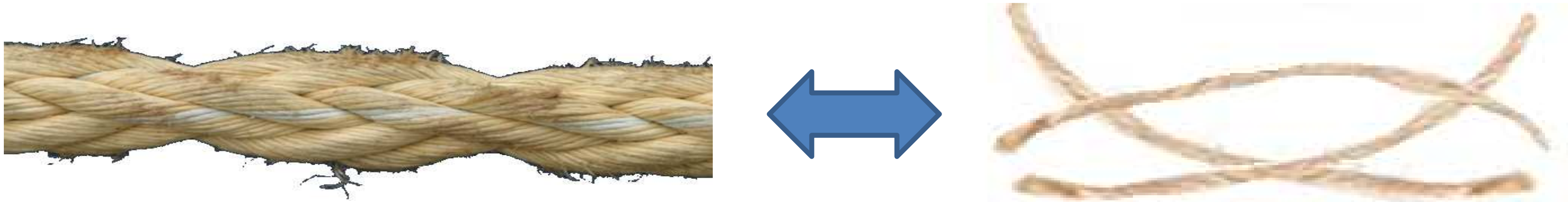


Key question

- Do indexes measure a latent variable?
  - Realism?
  - Ontological claim?

- Indexes summarize multiple variables, or combine them into one composite variable
  - If this is what you want to achieve, then indexes are fine.

- Transition from latent variable to composite index variable can be fuzzy



PLS – Partial Least Squares

A miracle of measurement or accidental constructivism? How PLS subverts the realist search for truth

John W. Cadogan
School of Business and Economics, Loughborough University, Loughborough, UK and School of Business and Management, LUT University, Finland and Business School, University of Eastern Finland, Kuopio, Finland, and

Nick Lee
Warwick Business School, Warwick University, Coventry, UK

Received 27 August 2020
Revised 22 March 2021
20 April 2021
11 June 2021
14 June 2021
Accepted 28 February 2022

Abstract

Purpose – This study aims to determine whether partial least squares path modeling (PLS) is fit for purpose for scholars holding scientific realist views.

Design/methodology/approach – The authors present the philosophical foundations of scientific realism and constructivism and examine the extent to which PLS aligns with them.

Findings – PLS does not align with scientific realism but aligns well with constructivism.

Research limitations/implications – Research is needed to assess PLS's fit with instrumentalism and pragmatism.

Practical implications – PLS has no utility as a realist scientific tool but may be of interest to constructivists.

Originality/value – To the best of the authors' knowledge, this study is the first to assess PLS's alignments and mismatches with constructivist and scientific realist perspectives.

Keywords Composites, PLS partial least squares, Structural equation models, Antirealism, instrumentalism, and pragmatism, Unobservable conceptual variables, Latent variables, Theory, Scientific realism, Constructivism, Causality, Truth and facts

Paper type Research paper

This paper examines partial least squares path modeling (PLS) for its alignments with two ontological stances: scientific realism [1] and constructivism [2]. Realism and constructivism are at odds with each other, built on fundamentally diverging beliefs about the nature of knowledge and how it is generated and justified. This conflict is seen clearly when one looks at the "science wars," which began in the early 1980s, and are the disagreements between constructivists, who argue that it makes little sense to claim that there is objective truth, since all facts are constructed by humans, not discovered (Collin, 2017), and scientific realists, the self-styled "defenders of the 'objective truth' derived from scientific investigation [. . .] [and] of rationality and realism" (Linker, 2001, p. 59).

Thanks to the two anonymous reviewers for their insights, and to Mikko Rönkkö and Cameron McIntosh for commenting on an early version of the manuscript.

European Journal of Marketing
© Emerald Publishing Limited
0309-0566
DOI 10.1108/EJM-08-2020-0637

Standard reflective model

A miracle of measurement or accidental constructivism? How PLS subverts the realist search for truth

John W. Cadogan
School of Business and Economics, Loughborough University, Loughborough, UK and School of Business and Management, LUT University, Finland and Business School, University of Eastern Finland, Kuopio, Finland, and

Nick Lee
Warwick Business School, Warwick University, Coventry, UK

Received 27 August 2020
Revised 22 March 2021
20 April 2021
11 June 2021
14 June 2021
Accepted 28 February 2022

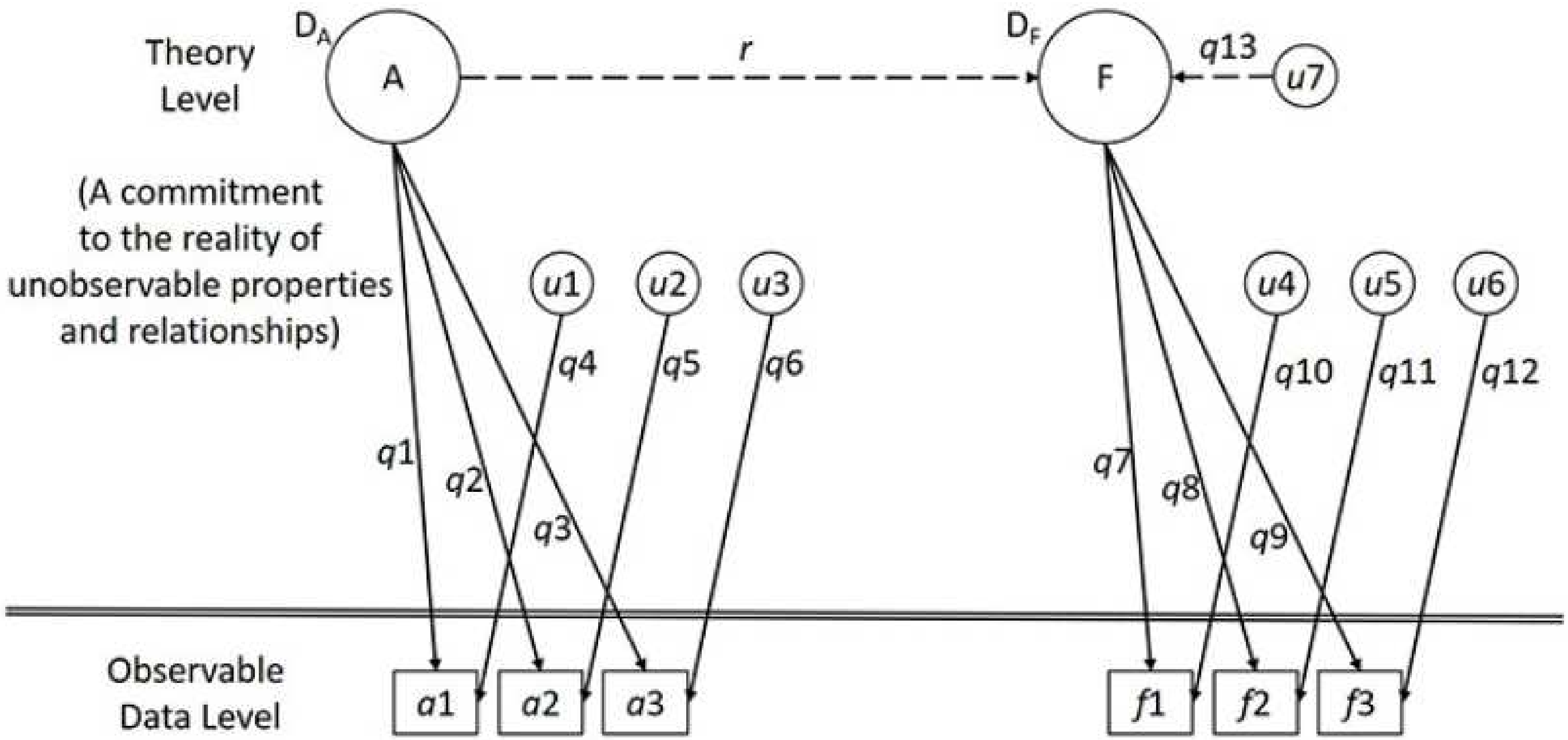


Figure 1. **(a)**

PLS – Partial Least Squares

A miracle of measurement or accidental constructivism? How PLS subverts the realist search for truth

John W. Cadogan
School of Business and Economics, Loughborough University, Loughborough, UK and School of Business and Management, LUT University, Finland and Business School, University of Eastern Finland, Kuopio, Finland, and
Nick Lee
Warwick Business School, Warwick University, Coventry, UK

Received 27 August 2020
Revised 22 March 2021
20 April 2021
11 June 2021
14 June 2021
Accepted 28 February 2022

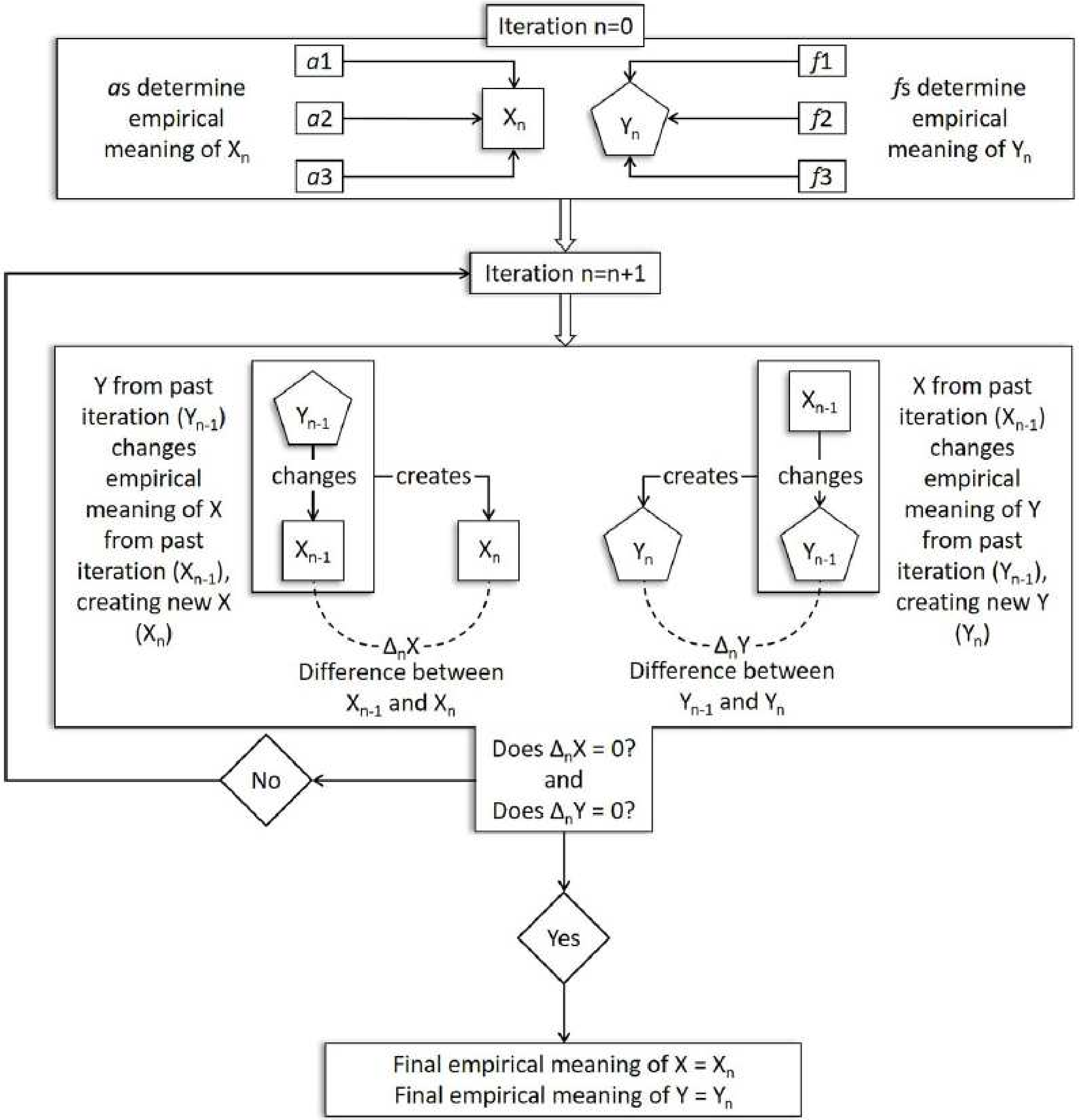


Figure 2.
How PLS gives its composites empirical meanings

A miracle of measurement or accidental constructivism? How PLS subverts the realist search for truth

John W. Cadogan
School of Business and Economics, Loughborough University, Loughborough, UK and School of Business and Management, LUT University, Finland and Business School, University of Eastern Finland, Kuopio, Finland, and
Nick Lee
Warwick Business School, Warwick University, Coventry, UK

Received 27 August 2020
Revised 22 March 2021
20 April 2021
11 June 2021
14 June 2021
Accepted 28 February 2022

PLS – Partial Least Squares

- First, in iteration $n = 0$, PLS creates X_0, a composite [5] of $a1$, $a2$ and $a3$ using unit weights.

$$X_0 = a1 + a2 + a3 \quad (1)$$

Similarly, PLS creates Y_0, a composite of $f1$, $f2$ and $f3$ using unit weights.

- In iteration $n = 1$, PLS calculates the relationships between $a1$, $a2$, $a3$ and Y_0 [6], either using correlations (called *Mode A*) or using regression coefficients (called *Mode B*). These relationships become the weights that are used to create a new X variable (X_1). For instance, if the relationship between $a1$ and Y_0 is 0.90, between $a2$ and Y_0 is 0.20 and between $a3$ and Y_0 is 0.01, X_1 is calculated as:

$$X_1 = 0.90{*}a1 + 0.20{*}a2 + 0.01{*}a3 \quad (3)$$

Similarly, PLS calculates the relationships between $f1$, $f2$, $f3$ and X_0, and these relationships become the weights that are used to create a new Y variable (Y_1). For instance, if the relationship between $f1$ and X_0 is 0.80, between $f2$ and X_0 is 0.15 and between $f3$ and X_0 is 0.02, Y_1 is calculated as:

$$Y_1 = 0.80{*}f1 + 0.15{*}f2 + 0.02{*}f3 \quad (4)$$

PLS – Partial Least Squares

A miracle of measurement or accidental constructivism? How PLS subverts the realist search for truth

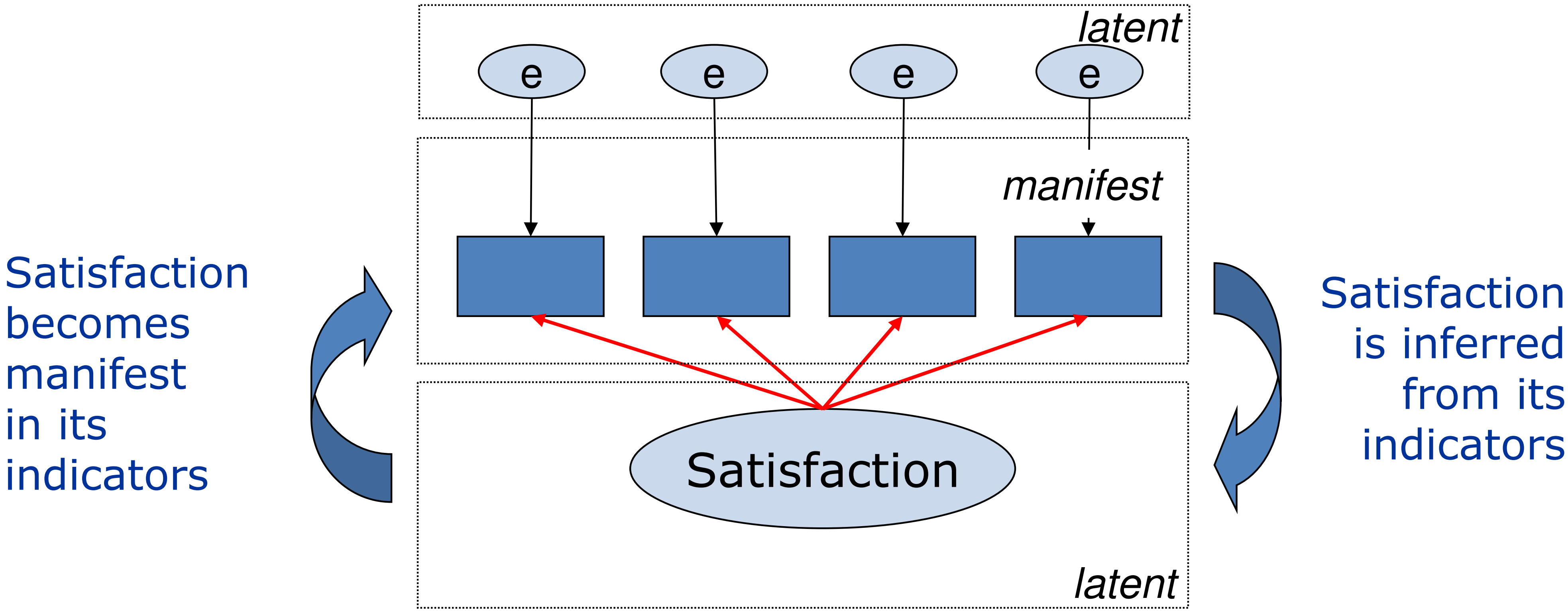
John W. Cadogan
School of Business and Economics, Loughborough University, Loughborough, UK and School of Business and Management, LUT University, Finland and Business School, University of Eastern Finland, Kuopio, Finland, and
Nick Lee
Warwick Business School, Warwick University, Coventry, UK

Received 27 August 2020
Revised 22 March 2021
20 April 2021
11 June 2021
14 June 2021
Accepted 28 February 2022

- PLS then examines the change in the scores produced by the new and old X (Δ_nX) and the new and old Y (Δ_nY) and asks "do the new X and Y variables have the same empirical meanings as the old X and Y scores?"
- If the answer is no, PLS goes through a second iteration (n = 2). It calculates the relationships between *a*1, *a*2, *a*3 and Y_1 and creates a new X variable (X_2) using the new relationships between the *a*s and Y_1 as weighting values. Likewise, it calculates the relationships between *f*1, *f*2, *f*3 and X_1 and creates a new Y variable (Y_2) using the new relationships between the *f*s and X_1 as weighting values.
- The process of repeating iterations continues until the empirical meanings of the new versions of X and Y do not change from iteration n-1 to iteration n. At this stage, the final empirical meanings of X and Y are Xn and Yn, respectively.

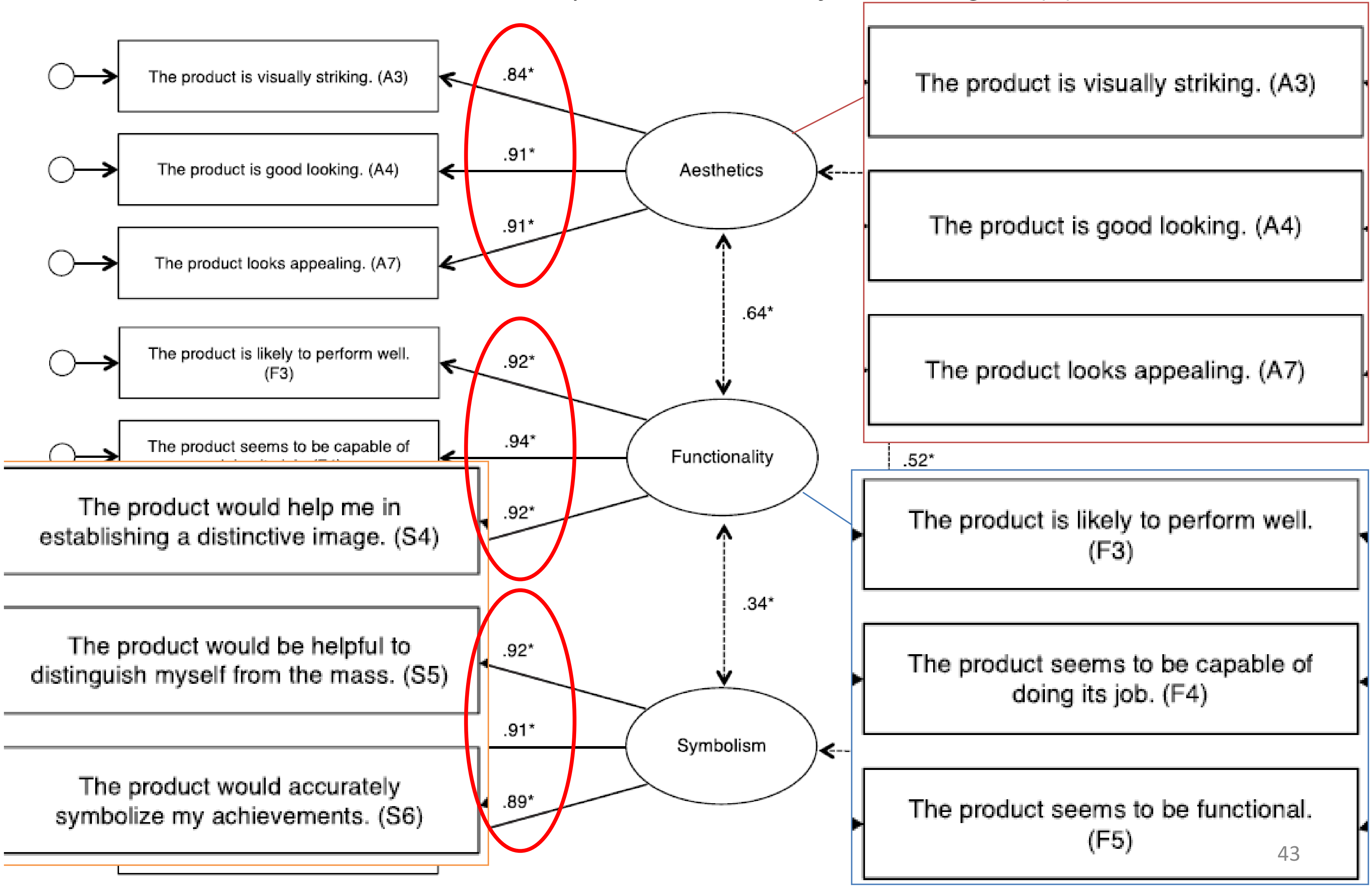
Reflective indicator model

- Indicator variables reflect latente variable
  - Reflective indicators



$AVE_{aesthetic} = .79;\ AVE_{functional} = .86;\ AVE_{symbolic} = .83.$

Homburg, C., Schwemmle, M., & Kuehnl, C. (2015). New product design: Concept, measurement, and consequences. *Journal of Marketing*, *79*(3), 41-56.



Homburg, C., Schwemmle, M., & Kuehnl, C. (2015). New product design: Concept, measurement, and consequences. *Journal of Marketing*, *79*(3), 41-56.

TABLE 3
Items, Means, and Standard Deviations (Study 1)

Item No.	The product...	M	SD
A3	...is visually striking	2.26	1.13
A4	...is good looking	2.08	1.07
A7	...looks appealing	2.10	1.07
F3	...is likely to perform well	1.86	.88
F4	...seems to be capable of doing its job	1.76	.85
F5	...seems to be functional	1.78	.86
S4	...would help me in establishing a distinctive image	3.34	1.29
S5	...would be helpful to distinguish myself from the mass	3.34	1.30
S6	...would accurately symbolize or express my achievements	3.35	1.29

Notes: All items assessed on a five-point scale (1 = "strongly agree," and 5 = "strongly disagree"). Only items included in the final scale are reported. For items F3–5 and S4–6, the questionnaire indicated to the respondent to judge the product "only from looking at it." Item numbers are consistent with Web Appendix W8.

M
2.26
2.08
2.10
1.86
1.76
1.78
3.34
3.34
3.35

Measurement

Measurement:

Claim that a latent variable exists as a quantitative property

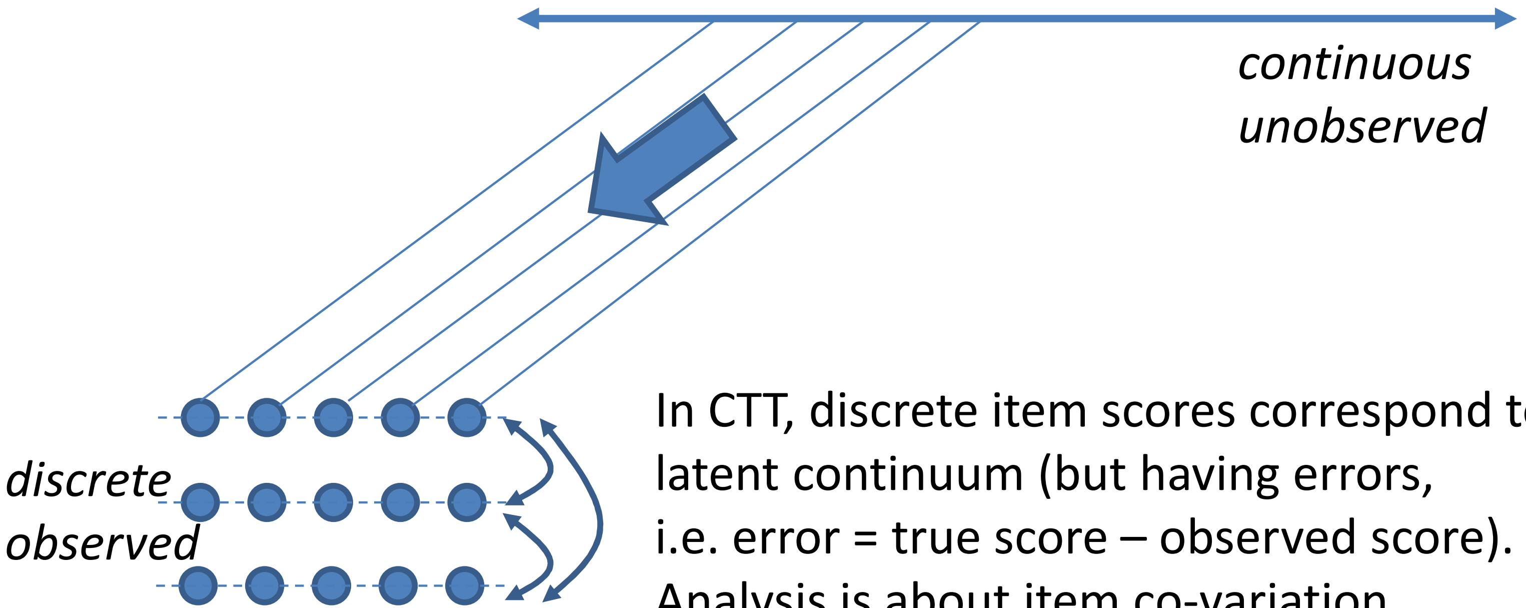
(Stevens opened the flood gates for various "types" of measurement, i.e. classification, and ordering)

Inferring from observed behaviour (item responses) the amount of the property the person possesses

Property = what we want to measure = measurand (metrology)

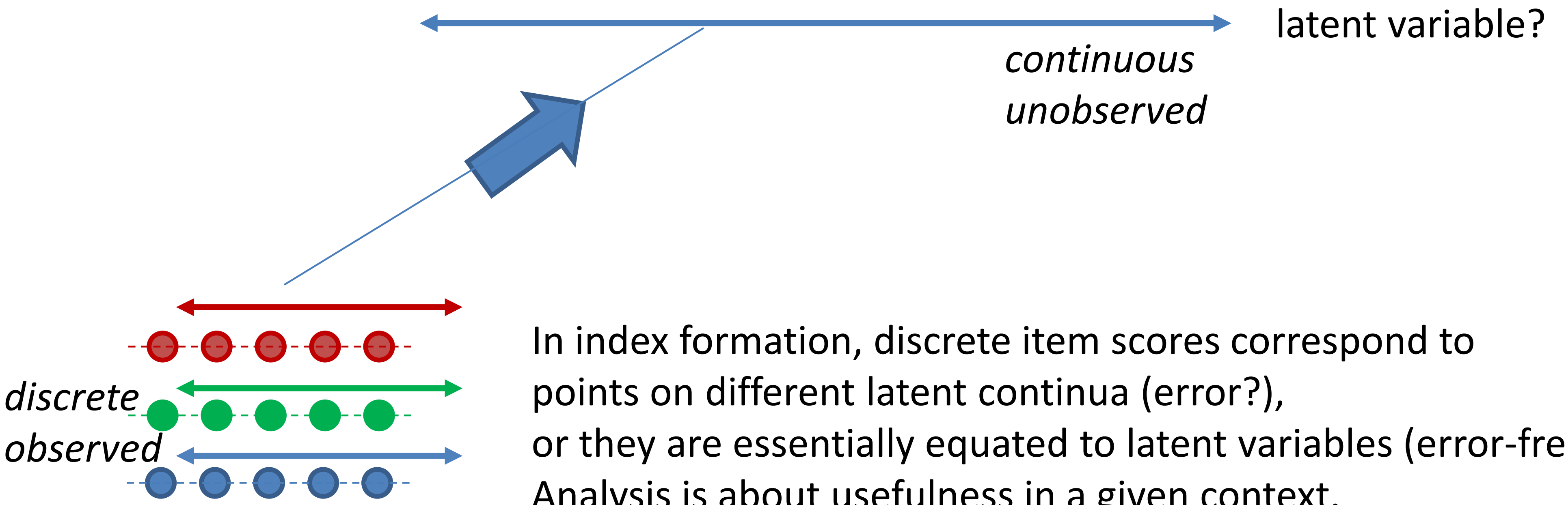
(Stevens focuses on objects to which numbers/numerals are assigned, but we are measuring attributes of objects, not objects themselves)

Elements in Measurement I: CTT



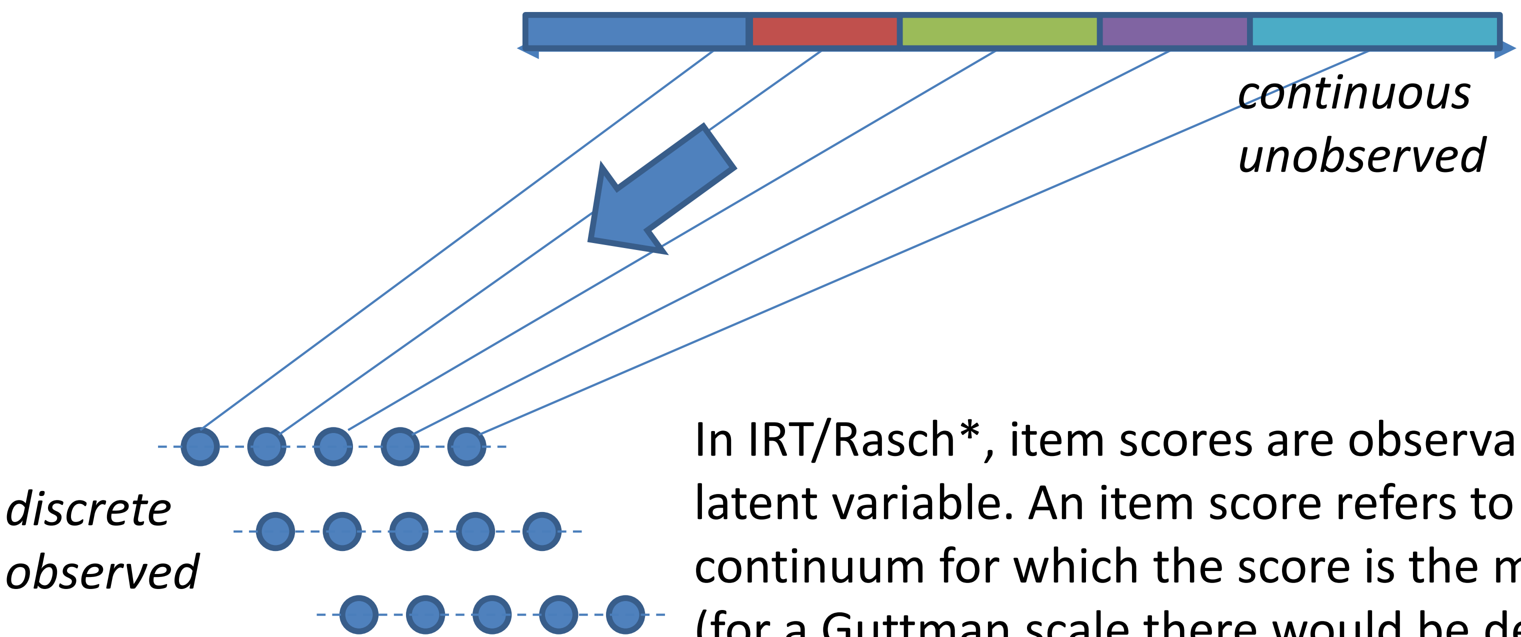
In CTT, discrete item scores correspond to points on the latent continuum (but having errors, i.e. error = true score – observed score).
Analysis is about item co-variation.
Maximum co-variation with identical means.

Elements in Measurement II: Index



In index formation, discrete item scores correspond to points on different latent continua (error?), or they are essentially equated to latent variables (error-free). Analysis is about usefulness in a given context.

Elements in Measurement III: IRT

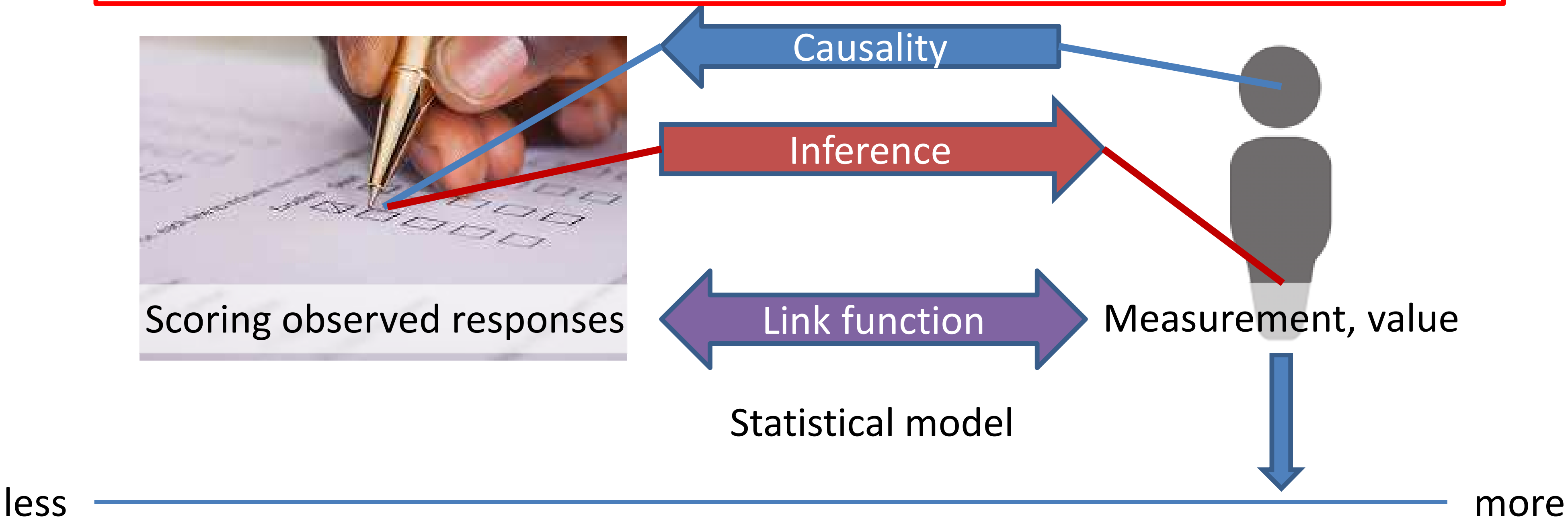


In IRT/Rasch*, item scores are observable consequences of the latent variable. An item score refers to a range on the latent continuum for which the score is the most likely score (for a Guttman scale there would be deterministic certainty).
Analysis is about expected structure in the data.
Items should not have identical means, as the items represent different levels of the construct.

** From a Rasch point of view, the Rasch model is not an IRT model, from an IRT perspective, it is.*

Measurement

How can we be sure that what we measure is what we want/claim to measure?
Content validity: items represent the latent variable
(conceptual definition, conceptual framework, substantive theory)



CTT: $x_{iv} = \tau_i + \lambda_i * F_v + e_{iv}$ (linear function)

Quantitative variable requires a quantitative substantive theory
Theory that explains/predicts properties of the items
Majority of substantive theories are qualitative (at best ordinal, very rarely quantitative)

Measurement

CTT: no invariant item properties (τ depends on sample)
Requires assumption of normal distributions: of e_{iv} (generally not a problem), of F_v (sample dependence, might not be true)

Scoring observed responses

Causality

Inference

Link function

Measurement, value

Statistical model

less — more

CTT: $x_{iv} = \tau_i + \lambda_i * F_v + e_{iv}$ (linear function)

Qualitative substantive theory predicts which items are part of the universe of potential indicators of the latent variable
(in practice, almost every time items are discarded; item purification)
Items are assumed to be equal (same τ, same λ, same VAR[e]; thus same means and standard deviations) and perfectly interchangeable

Measurement

CTT: no invariant item properties (τ depends on sample)
Requires assumption of normal distributions: of e_{iv} (generally not a problem), of F_v (sample dependence, might not be true)

Scoring observed responses

Causality

Inference

Link function

Measurement, value

Item scores are at best ordinal

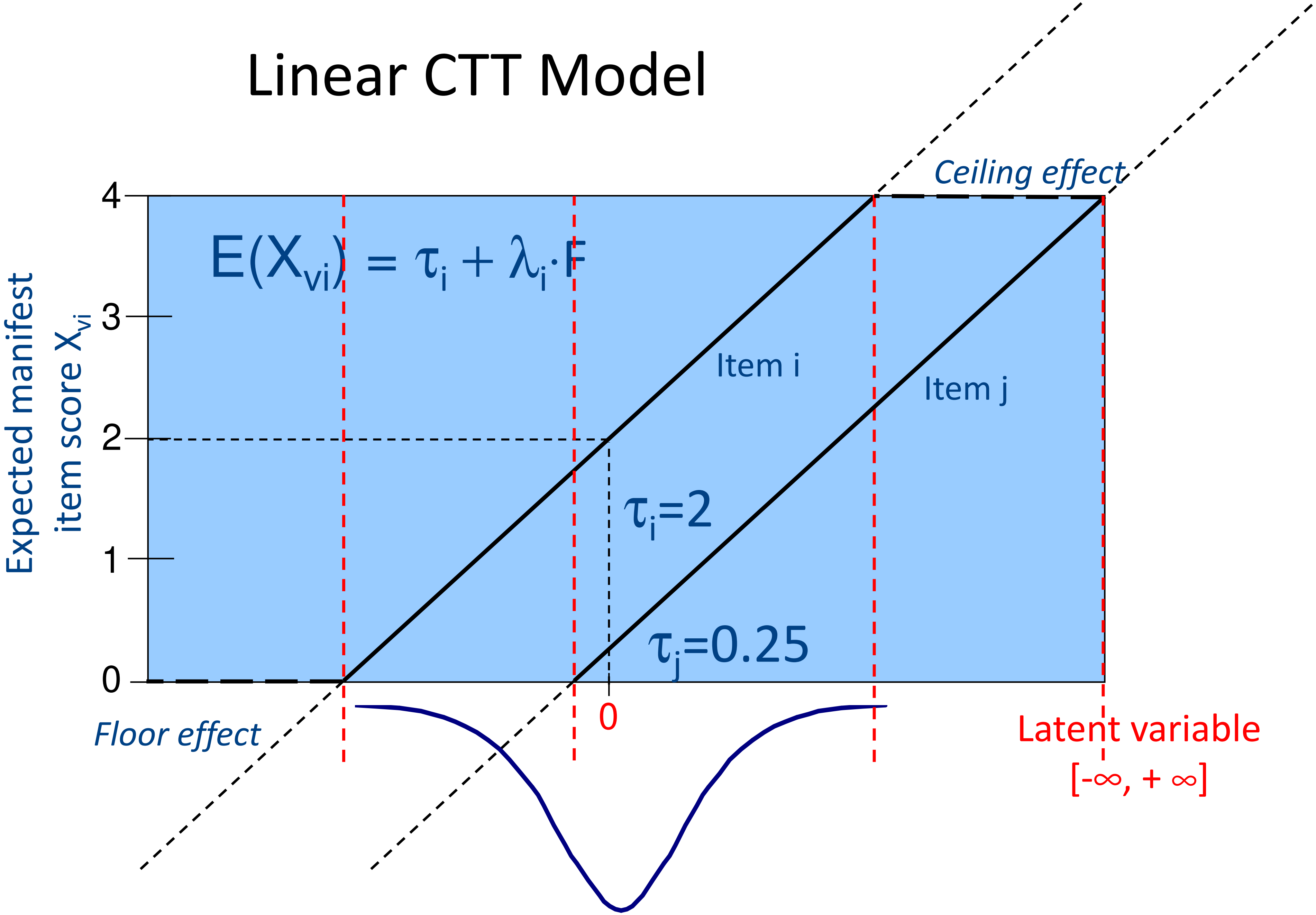
Statistical model

less — more

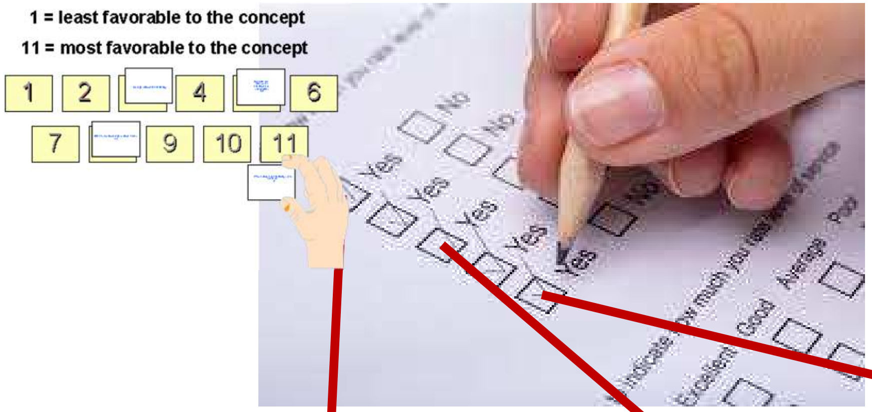
CTT: $x_{iv} = \tau_i + \lambda_i * F_v + e_{iv}$ (linear function)

Linear link between ordinal item score and linear, interval scaled true value F
Item scores must also be linear, interval scaled
Thus, item scores are already measures
Sum score across items informative (cumulative model)
all Y or all N makes most sense (but is, in practice, seen as an indication of straight-lining)

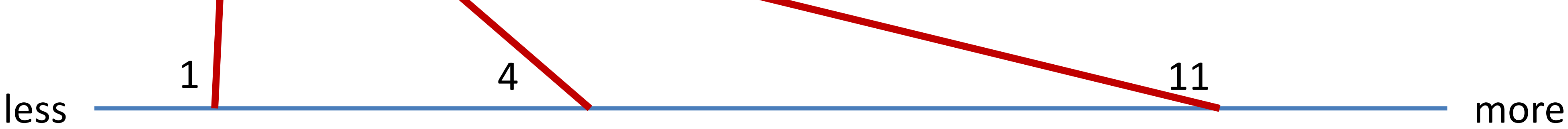
Linear CTT Model



Thurstone scaling (equal-appearing interval)



Theory-based item properties
(Could be tested empirically by means of unfolding models, which disentangle negative responses due to "too much" and "too little")



Y N N → 1; N Y N → 4; N N Y → 11 (most plausible patterns)
Y Y N → 2.5; N Y Y → 7.5
Y Y Y → 5.3
Y N Y → 6 (least plausible pattern)
Sum score is not informative! (all three patterns in line 1 have total score of 1)
Non-cumulative/additive model
Scale level?
Interval scale for 1 to 11 is speculative
Some sort of ordinal (implicit) theory of the latent variable required
Two reasons for disagreement (ideal point) → Unfolding model

Guttman scaling

1 2 3 4 ...

less ——— more

Theory-based hierarchy of items (from "easy" to "difficult")

- Each item requires a particular level of the latent variable to agree
- A positive response to a given item implies a positive response to all easier items (cumulative, deterministic)
- *(The rating of items suggested by Trochim is no part of traditional Guttman scaling but could, in principle, be used to define the item hierarchy)*

Guttman patterns:

N N N N → 0; Y N N N → 1; Y Y N N → 2; Y Y Y N → 3; Y Y Y Y → 4

Sum score is informative!

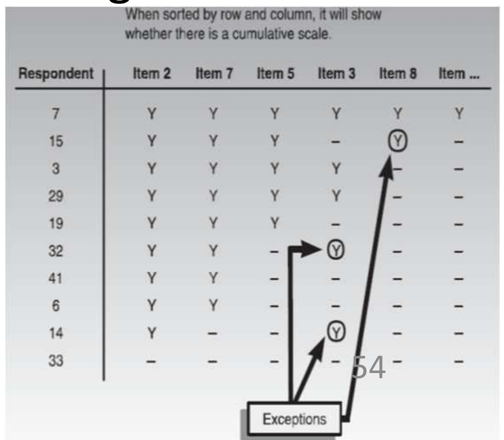
But there is no way to determine the distance between successive integer sum scores

Hence, ordinal scale only

all other patterns are so-called Guttman errors

e.g., N Y N N → (1); N N N Y → (1)

In practice, approx. 15% Guttman errors are tolerated



When sorted by row and column, it will show whether there is a cumulative scale.

Respondent	Item 2	Item 7	Item 5	Item 3	Item 8	Item ...
7	Y	Y	Y	Y	Y	Y
15	Y	Y	Y	–	(Y)	–
3	Y	Y	Y	Y	–	–
29	Y	Y	Y	Y	–	–
19	Y	Y	Y	–	–	–
32	Y	Y	–	(Y)	–	–
41	Y	Y	–	–	–	–
6	Y	Y	–	–	–	–
14	Y	–	–	(Y)	–	–
33	–	–	–	–	–	–

Rasch Measurement

Hierarchy of items (from “easy” to “difficult”)
Locations of items and persons are placed on the same latent continuum
(Think of weights and weightlifters)
Sum score (score across items, yes =1, no = 0) is a sufficient statistic (there is no additional information in the pattern;
Y N N Y N = Y Y N N N = 2; *but different person fit!*)

item 1 item 2 item 3 item 4

less more

Probability of positive response depends on “item difficulty” and “person ability”
(how much of the property the item represents and how much the person possesses)
Probabilistic version of the Guttman model

Guttman patterns have highest likelihood
Non-Guttman patterns have smaller likelihood but are possible (non-zero prob.)
Y Y N N is more likely than Y N Y N, but both score 2
N N Y Y is least likely (given a total score of 2), questions measurement (poor person fit)

Rasch Measurement

Hierarchy of items (from “easy” to “difficult”)
Locations of items and persons are placed on the same latent continuum
(Think of weights and weightlifters)
Sum score (score across items, yes =1, no = 0) is a sufficient statistic (there is no additional information in the pattern;
Y N N Y N = Y Y N N N = 2 ; *but different person fit!*)

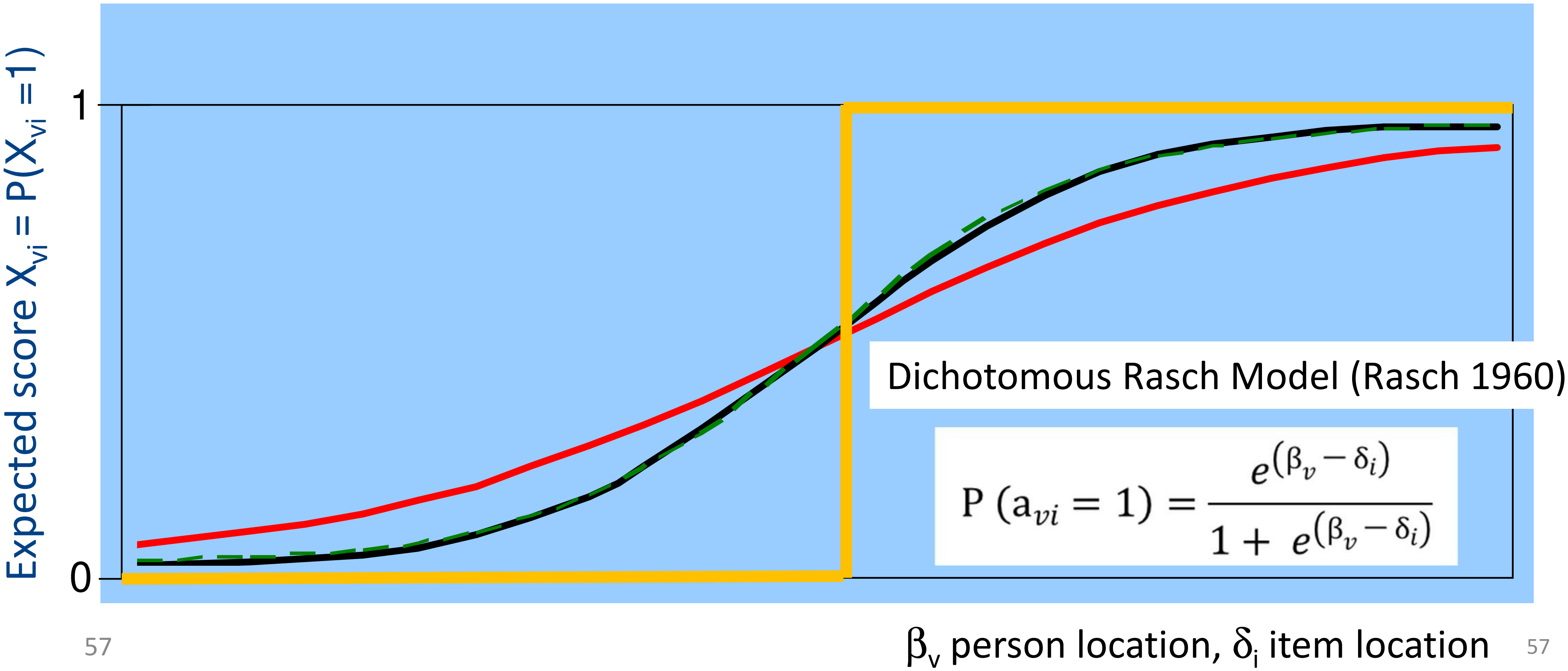
less — item 1 — item 2 — item 3 — item 4 — more

Say, the probability of a positive response to item 1 = 0.7
for item 2 = 0.5, for item 3 = 0.4 and for item 4 = 0.1
Probability for Y N N N = 0.01, Y Y N N = 0.19, Y Y Y N = 0.13, Y Y Y Y = 0.01
Y N Y N = 0.13, N Y Y N = 0.05, Y N N Y = 0.02, N Y N Y = 0.01, N N Y Y < 0.01

Modelling the probability of a positive response as a function of the item location and the person location (both are unknown and estimated from the data)
Non-linear link function between item response (bound between 0 and 1, or 0 and k for polytomous items) and person measure

Rasch Measurement

- Cumulative normal function (normal ogive, early Item Response Theory models)
- Logistic function: $e^{d}/(1 + e^{d})$
- Logistic with scaling constant: $e^{1.7 \cdot d}/(1 + e^{1.7 \cdot d})$
- Guttman



57

Rasch Measurement

- Specific objectivity (Rasch, 1961, 1977; invariance) – unique for Rasch model
  - Item properties are independent of persons used to estimate them
  - Person properties are independent of items used
  - Rasch model is, in principle, sample-independent
- Raw score sufficiency
  - requires equal discrimination of all items
  - otherwise there are no sufficient statistics (score for persons would depend on item discrimination, the estimation of which requires distributional assumptions of persons, hence item properties would not be independent of persons)
- Item score is a count of thresholds passed
  - Non-linear transformation of raw score to linear measurement
- Broad range of tests and criteria of fit
  - item fit, person fit, unidimensionality, invariance, local independence
- Applicable to dichotomous and polytomous items
- Rasch model is prescriptive (requires structure in the data)
  - compatible with axioms of quantity
- Other IRT models are descriptive (aim at best describing given data)
  - estimate item discrimination

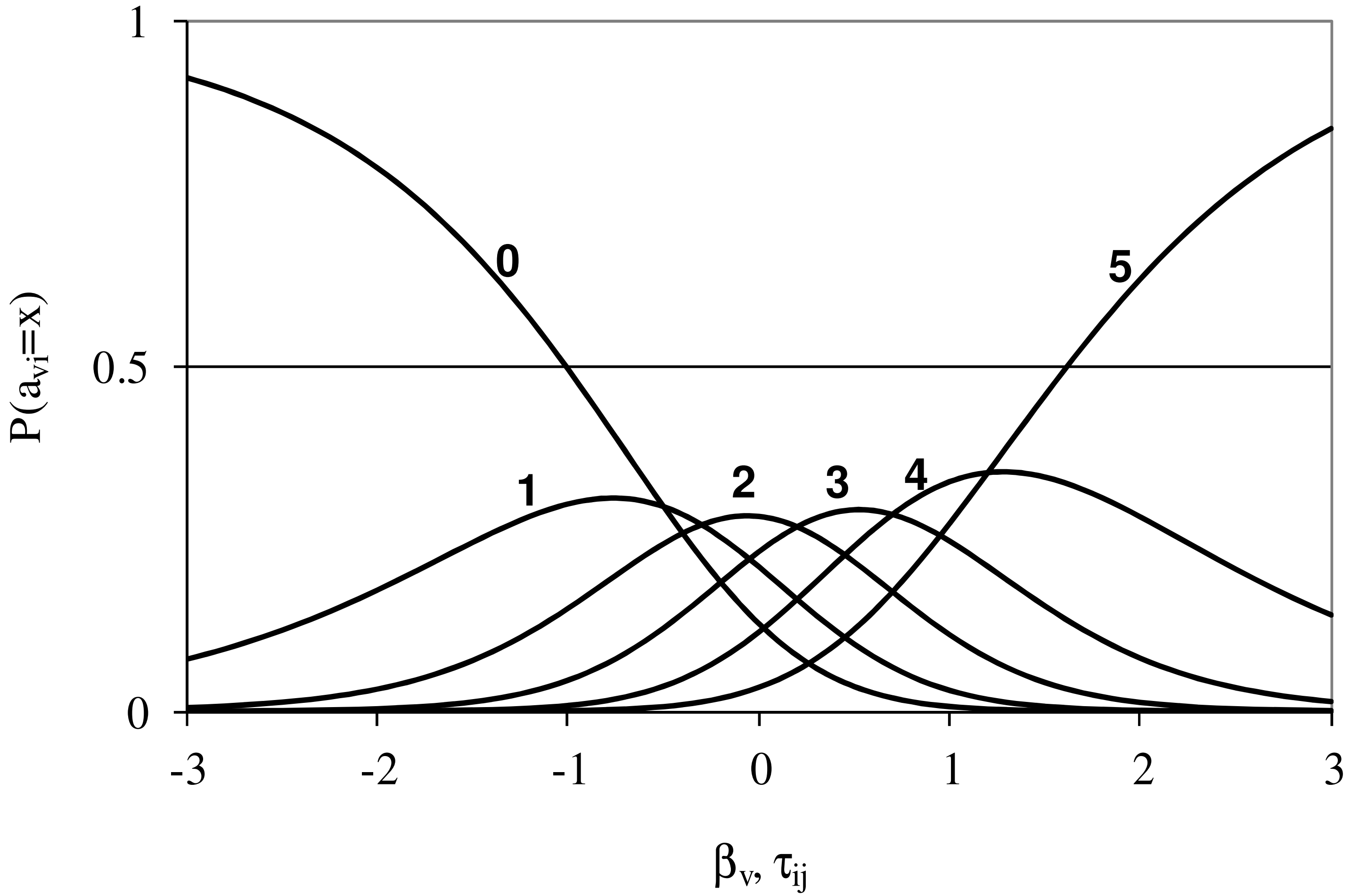
Polytomous Rasch model

- Partial Credit Model (Masters, 1982; Andrich, 1988)
  - Structure of threshold (τ) distances is different for each item

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

$$\gamma = 1 + \sum_{k=1}^{m} e^{\left(\sum_{j=1}^{x} -\tau_{ij}\right) + k \cdot (\beta_v - \delta_i)}$$

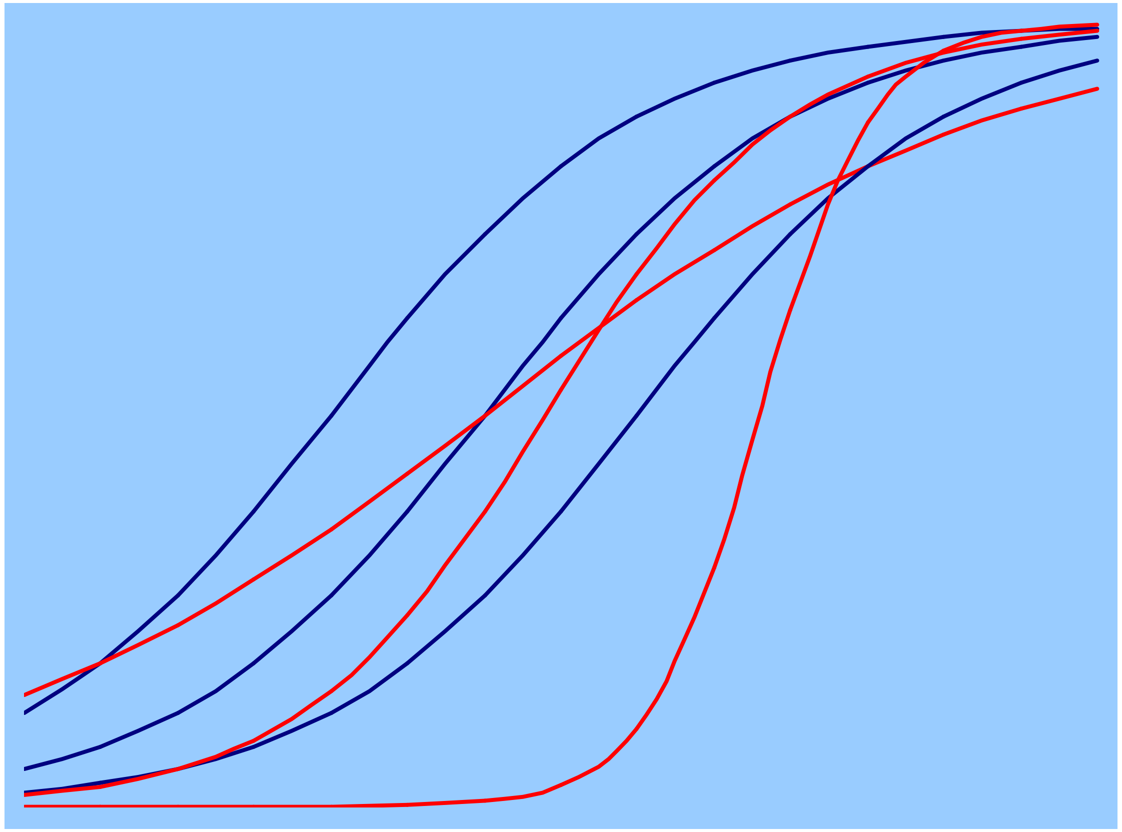
Polytomous Rasch Model



60

Different IRT-Models

- Rasch model (in blue), called 1pl in iRT
- Birnbaum model, 2pl (in red)

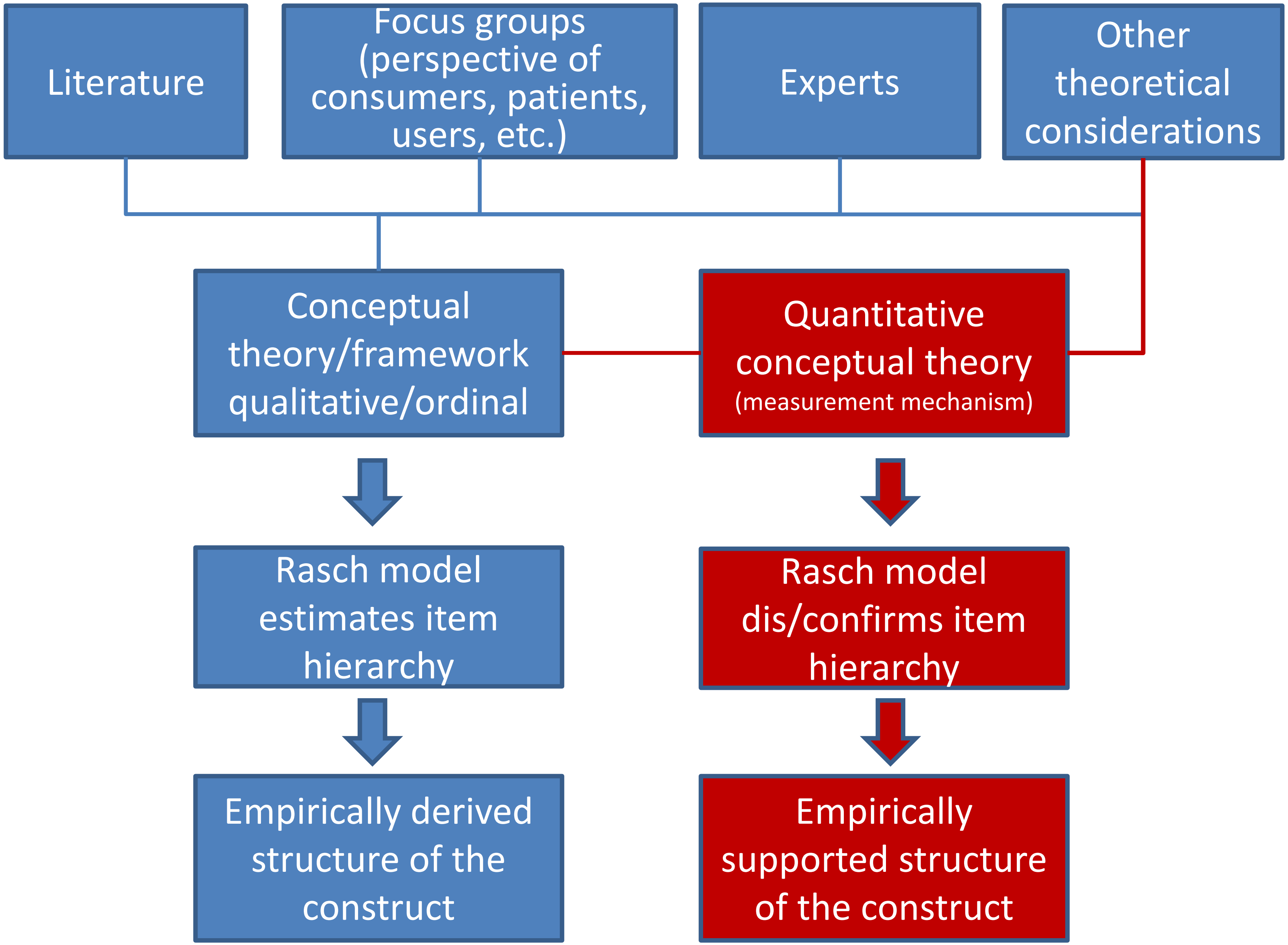


$$P(a_{vi} = 1) = \frac{e^{\beta_v - \delta_i}}{1 + e^{\beta_v - \delta_i}}$$

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

61

Rasch Measurement



Conclusions

- Social measurement requires a strong substantive theory of the construct (latent variable)
  - Ideally a quantitative theory (after all, we claim to measure a quantitative variable!) that exposes a measurement mechanism, which tells us why a given item has a particular location

Concept-driven measurement

Quaglia, M., Pendrill, L., Melin, J., Cano, S., & 15HLT04 NeuroMET Consortium. (2016-2019). Innovative measurements for improved diagnosis and management of neurodegenerative diseases (EMPIR NeuroMET). Teddington, Middlesex, UK:. EURAMET. https://www.lgcgroup.com/our-programmes/empir-neuromet/neuromet-landing-page/ (36 pp.)

Quaglia, M., Pendrill, L., Melin, J., Cano, S., & 18HLT09 NeuroMET2 Consortium. (2019-2022). Publishable Summary for 18HLT09 NeuroMET2: Metrology and innovation for early diagnosis and accurate stratification of patients with neurodegenerative diseases (EMPIR NeuroMET). Teddington, Middlesex, UK:. EURAMET. https://www.lgcgroup.com/our-programmes/empir-neuromet/neuromet-landing-page/ (5 pp.)

Measurement mechanism (Causal Rasch Models)

Stenner, A. J. (1996). Measuring Reading Comprehension with the Lexile Framework. [https://files.eric.ed.gov/fulltext/ED435977.pdf]

Stenner, A. J., Burdick, H., Sanford, E. E., & Burdick, D. S. (2006). How accurate are Lexile text measures?. *Journal of Applied Measurement*, *7*(3), 307.

Stenner, A. J., Fisher Jr, W. P., Stone, M., & Burdick, D. (2013). Causal Rasch models. *Frontiers in psychology*, *4*, 536.

A. Jackson Stenner, Mark H. Stone, Donald S. Burdick: The concept of a measurement mechanism Rasch Measureent Transactions [http://www.rasch.org/rmt/rmt232b.htm]

Conclusions

Lexile framework

https://metametricsinc.com/parents-and-students/lexile-for-parents-and-students/lexile-for-reading/

Conclusions

- Without a strong substantive theory of the construct ...
- ... measurement could just be a statistical exercise
  - Risk of treating measurement models as a technology to produce measurements (Michell, 2017)
  - Traditional content validity is insufficient (confirmation bias, post hoc?)
- Measurement model must incorporate principles (axioms) of quantity that have to be present in the data
  (Hölder, 1901; Michell and Ernst, 1996, 1997; Michell, 1999).
- Psychometrics rather than (only) statistics
  - Psychometric paradigm: prescriptive
  - Statistical paradigm: descriptive

Michell, J. (2017). On substandard substantive theory and axing axioms of measurement: A response to Humphry. *Theory & Psychology*, *27*(3), 419-425.

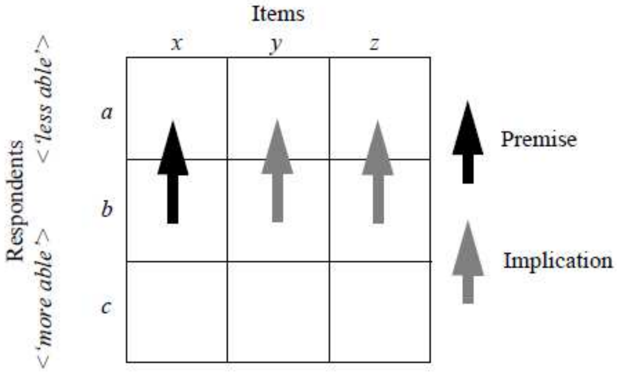
Hölder, O. (1901), 'Die Axiome der Quantität und die Lehre vom Mass', *Berichte über die Verhandlungen der Königlichen Sächsischen Gesellschaft der Wissenschaften zu Leipzig*, Mathemathische-Physische Classe, Leipzig: Hirzel, 35, pp. 1−64.

Michell, J. and C. Ernst (1996), 'The Axioms of Quantity and the Theory of Measurement, Part I, An English Translation of Hölder (1901)', Journal of Mathematical Psychology, 40, 235−252.

Michell, J. and C. Ernst (1997), 'The Axioms of Quantity and the Theory of Measurement, Part II, An English Translation of Hölder (1901)', Journal of Mathematical Psychology, 41, 345−356.

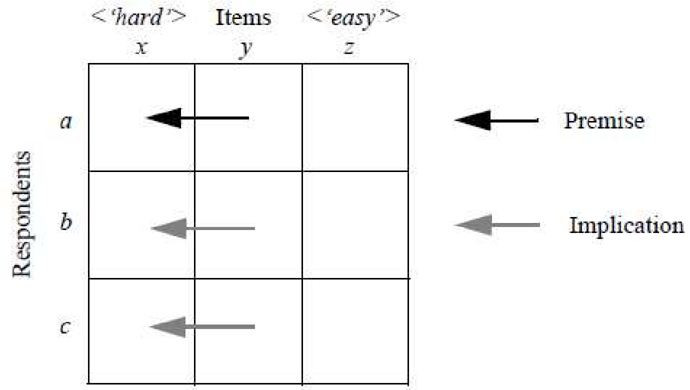
Michell, J. (1999), *Measurement in Psychology – a Critical History of a Methodological Concept*, Cambridge: Cambridge University Press.

Cancellation conditions



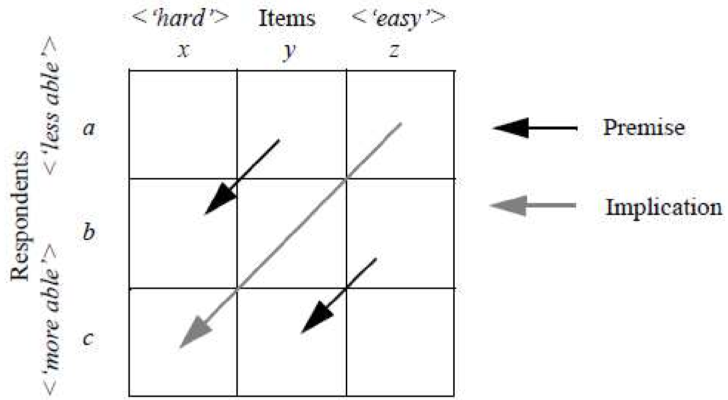
Source: adapted from Karabatsos (2001, p.399)

Figure 4.1 The weak respondent (column) independence axiom



Source: adapted from Karabatsos (2001, p.399)

Figure 4.3 The weak item (row) independence axiom



Source: adapted from Karabatsos (2001, p.402)

Figure 4.5 The double cancellation axiom

Salzberger, T. (2009). *Measurement in marketing research: An alternative framework*. Edward Elgar.

A naturalistic, realist account of measurement

Measurement is ***not*** the assignment of numerals to objects

“[measuring mental attributes] is more myth-based technology than science” (Michell, 2017)

Operationalism is incompatible with realism, it is a “technology paradigm” (Humphry, 2017; Michell, 2017)

Numbers are ***not*** to represent something that is in the object (or rather the attribute)

Numerical relationships are inherent in the attributes (provided they are quantitative)

Humphry, S. M. (2017). Psychological measurement: Theory, paradoxes, and prototypes. Theory & Psychology, 27, 407–418.

Michell, J. (2017). On substandard substantive theory and axing axioms of measurement: A response to Humphry. Theory & Psychology, 27(3), 419-425.

Michell, J. (2021). Representational measurement theory: Is its number up?. *Theory & Psychology*, *31*(1), 3-23.

A naturalistic, realist account of measurement

The fallacy of representational measurement

According to representational theory, measurement is possible only because the empirical system represented and the numerical system representing it, possess the same *mathematical* structure. This is the basis of the theory, necessary to sustain surrogative reasoning. However, if the empirical and numerical systems have the same mathematical structure, it follows that **mathematical structure is present in empirical systems**. (page 11)

... if mathematical structure is present in empirical situations, there is no ontological divide between mathematical structure and empirical situations. (page 12)

Michell, J. (2021). Representational measurement theory: Is its number up?. *Theory & Psychology*, *31*(1), 3-23.

A naturalistic, realist account of measurement

... position that mathematics is the science of structure with many of the structures investigated by mathematicians, including **number systems, being instantiated in the world around us**. Page 13.

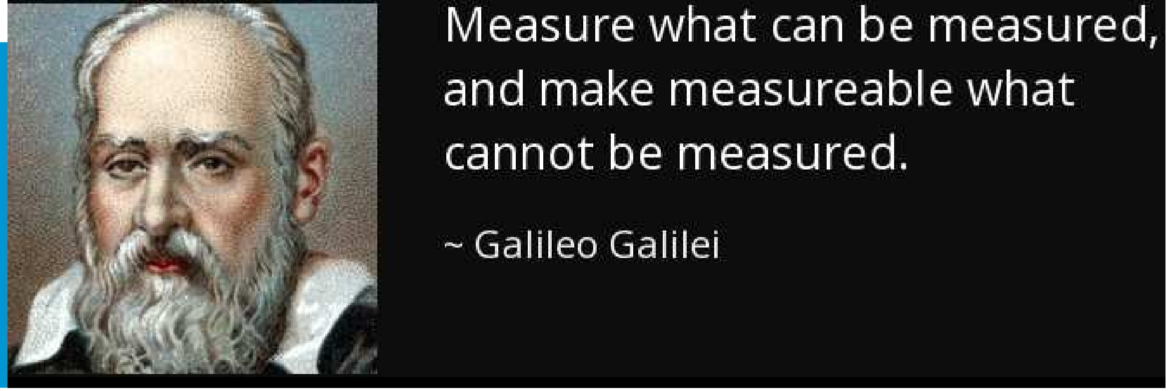
[T]he logic of measurement is not representation; it is instantiation. Page 16.

instantiate: to represent (an abstraction) by a concrete instance

heroes instantiate ideals— W. J. Bennett

Michell, J. (2021). Representational measurement theory: Is its number up?. *Theory & Psychology*, *31*(1), 3-23.

All is Quantitative
Is it?



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

*

"All is number."

Pythagoras

Explaining the world using mathematics.

Mathematics: "that which is learned"

(from *manthanein* "to learn")

Mathematics tells us lessons.

EFMD EQUIS ACCREDITED

A naturalistic, realist account of measurement

Three concepts framed the traditional paradigm: *quantity, magnitude*, and *ratio*. A *quantity* is an attribute of some kind, such as length, mass, or velocity, possessing internal structure sufficient for measurement, and the expression *quantitative attribute* is a synonym. A *magnitude* of a quantity is a specific degree of a quantitative attribute, such as the length of a football field. It is not a number. A magnitude is a specific, instantiated attribute of something. **A *ratio* is a kind of relation holding between two magnitudes of the same kind.** It is the relation of *relative magnitude*. (Michell, 2021, page 5)

Measurement 'is the process of discovering ratios' (Michell, 1999, p.14). The measure of any other magnitude of the same quantitative attribute is just its ratio to the unit of measurement' (Michell, 1999, p.13).

Michell, J. (1999), Measurement in Psychology – a Critical History of a Methodological Concept, Cambridge: Cambridge University Press.

A naturalistic, realist account of measurement

Scientific task of measurement

Evidence that the attribute exists as a quantitative property

Instrumental task of measurement

Devising instruments to measure the property

We need a model that simultaneously addresses either task of measurement.

Michell, J. (1999), Measurement in Psychology – a Critical History of a Methodological Concept, Cambridge: Cambridge University Press.

A naturalistic, realist account of measurement

In summary, Michell (1999, pp.222f) provides the following definitions relevant to measurement,

- *Quantity* – an attribute possessing ordinal and additive structure,
- *Quantification* – the process of (i) showing that an attribute is quantitative
- and (ii) devising procedures to measure it,
- *Measurement* – the discovery or estimation of the ratio of a magnitude of a quantity to a unit of the same quantity,
- *Unit* – a specific magnitude of a quantity relative to which measurements are made.

Clifford (1882) : “Every quantity is measured by the ratio which it bears to some fixed quantity, called the unit” (p. 525).

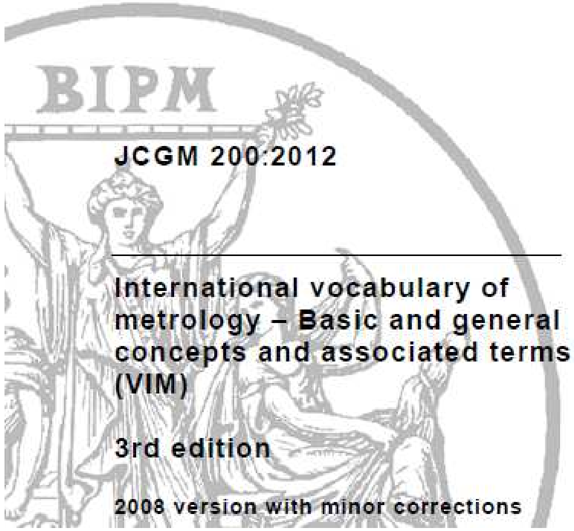
Clifford, W. K. (1882). Lecture notes. In R. Tucker (Ed.), Mathematical papers by William Kingdom Clifford (pp. 524–530). Macmillan. Cited in Michell (2021)

Unified measurement
(natural and social sciences)

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**

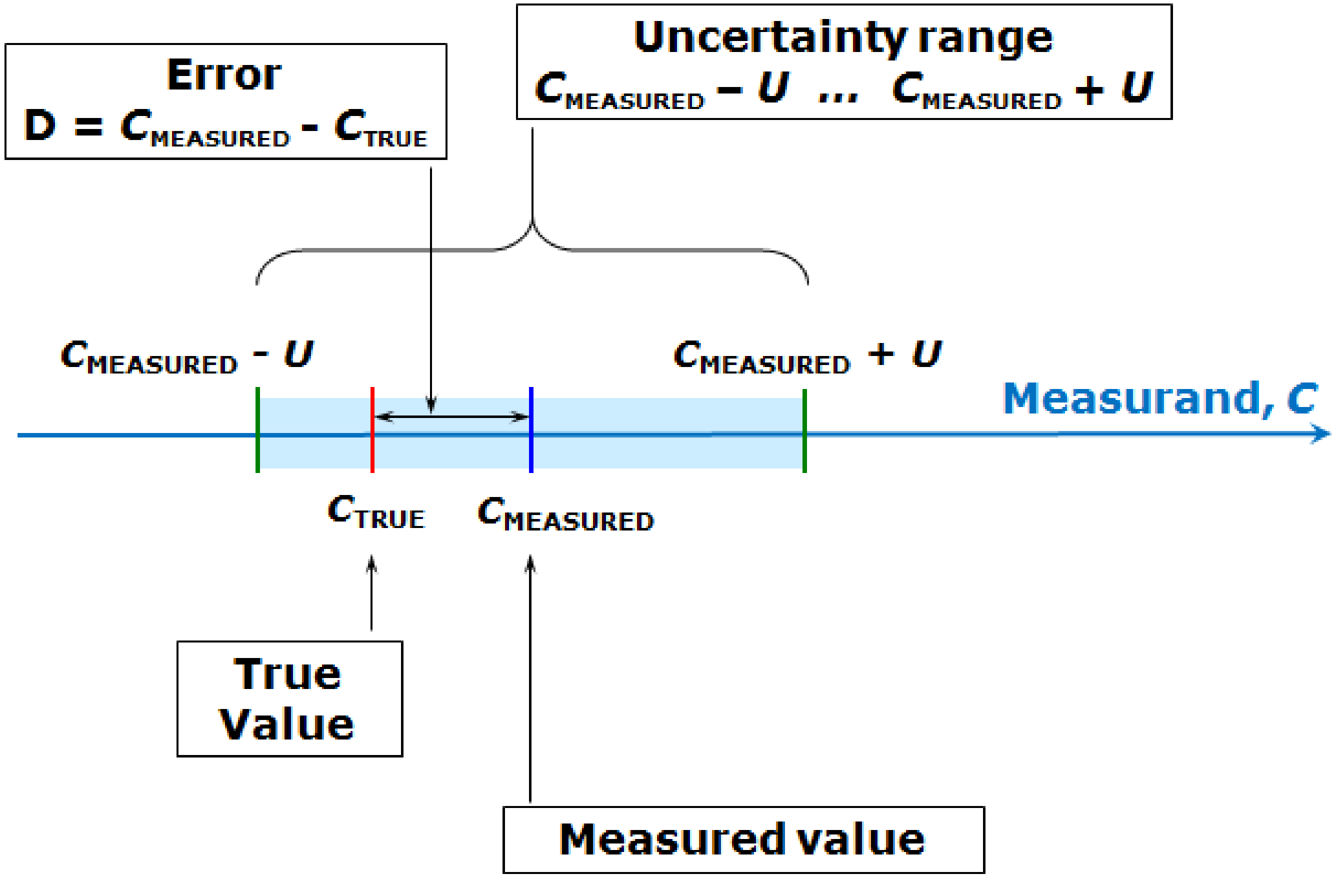
NOTE 1 For this definition, a 'reference' can be a definition of a **measurement unit** through its practical realization, or a **measurement procedure** including the measurement unit for a non-**ordinal quantity**, or a **measurement standard**.



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

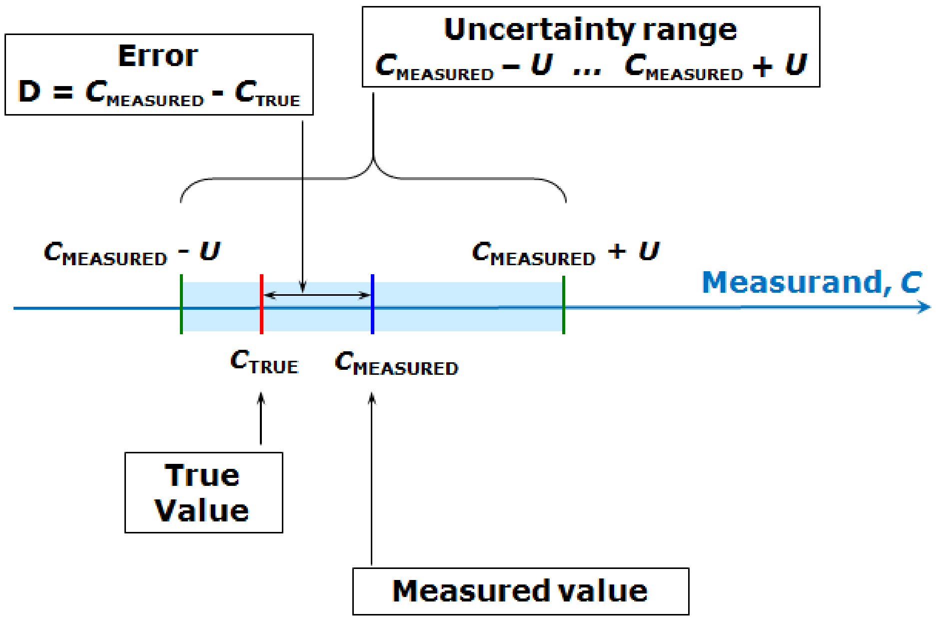
Metrology



Scheme 1.1. Interrelations between the concepts true value, measured value, error and uncertainty.

https://sisu.ut.ee/measurement/introduction-concept-measurement-uncertainty

Metrology



Scheme 1.1. Interrelations between the concepts true value, measured value, error and uncertainty.

low **uncertainty** (high accuracy) = **trueness** + **precision**

- **trueness** = (lack of) **systematic error**
- **precision** = (lack of) **random error**

until recently accuracy = trueness
accuracy was understood as a qualitative problem
(one either measures what one wants to measure or not)
trueness inherits this meaning

- trueness has a qualitative component (e.g. a scale that measures image but claims to measure satisfaction has no trueness – this cannot be corrected)
- trueness has a quantitative component (systematic error; can be corrected, bias* as an estimate of the systematic error)

reliability captures only precision
standard error of measurement (SEM) based on reliability inappropriate $SEM = S\sqrt{1 - r_{XX}}$

what is validity? accuracy? trueness? qual/quant?

* in the social sciences bias usually is the systematic error, not its correction

Measurement System and Uncertainty

Guide to the expression of uncertainty in measurement — Part 6: Developing and using measurement models

JCGM GUM-6:2020

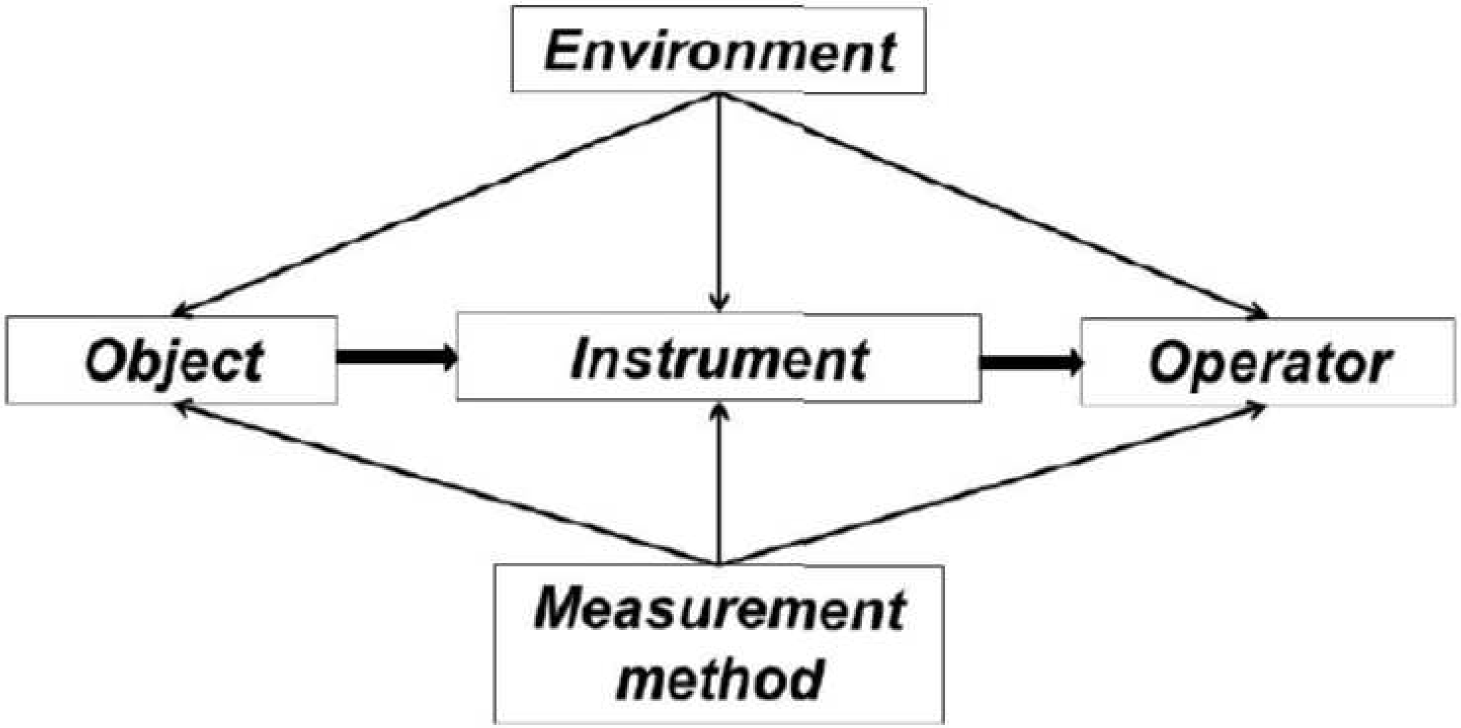
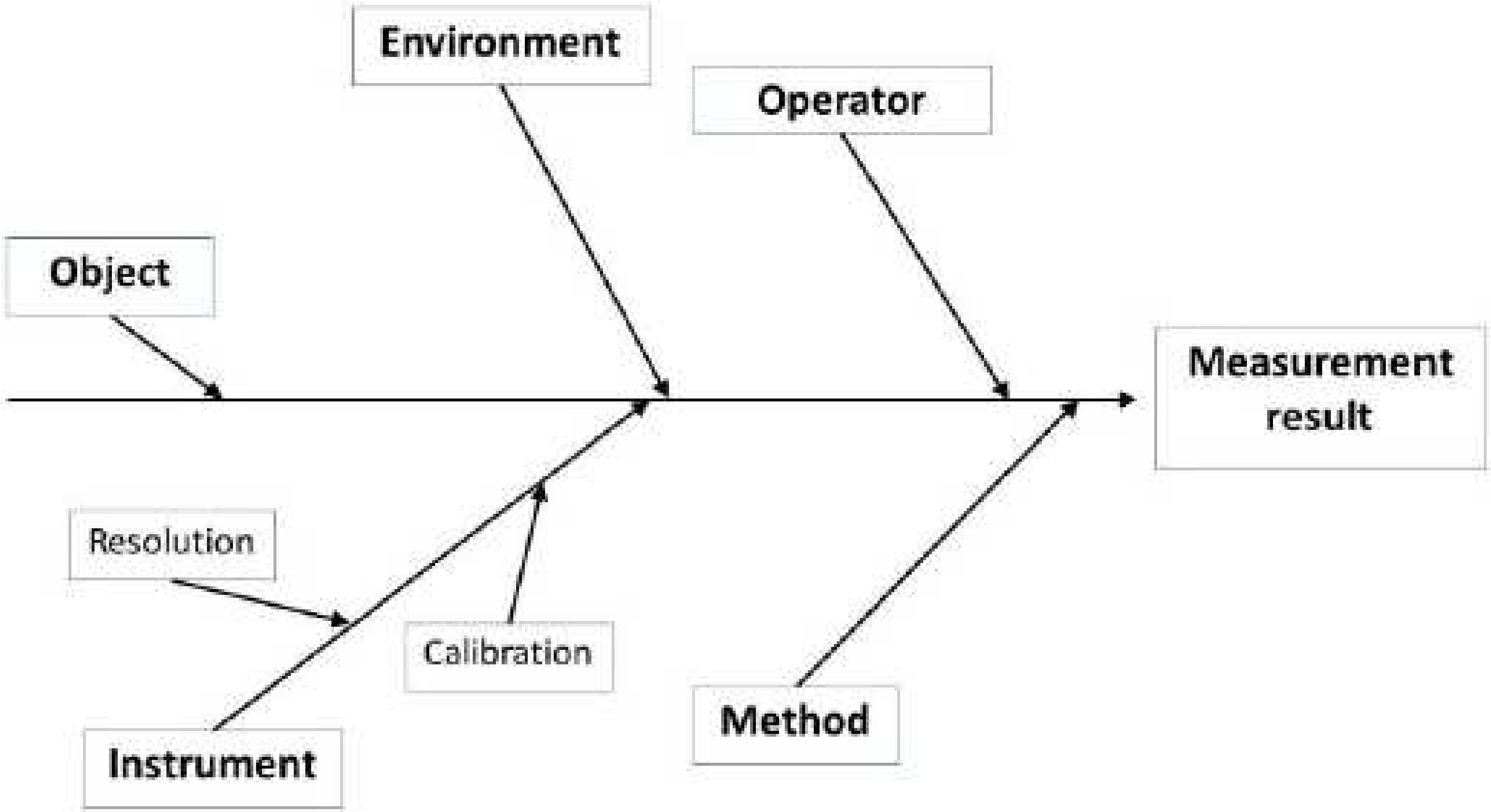


Figure E.3 — Measurement system analysis

Measurement System and Uncertainty



[1] Robinson Research Institute, University of Adelaide

[2] California State University, Northridge

Cite this as: *BMJ* 2022;376:o702

http://dx.doi.org/10.1136/bmj.o702

Published: 16 March 2022

The illusion of evidence based medicine

Evidence based medicine has been corrupted by corporate interests, failed regulation, and commercialisation of academia, argue these authors

Jon Jureidini, [1] Leemon B. McHenry[2]