WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS

EFMD EQUIS ACCREDITED

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

Quantitative Research Methods

Measurement

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

Topics 3 & 5
(Book chapters 3 & 5)

[3] **Measurement** **to be presented in Unit 2**

- Quantitative research typically involves measurement.
- Challenge of measuring unobservable (so-called *latent*) variables
- Levels of measurement in the social sciences (according to S.S. Stevens' notion of measurement)
- Quality criteria of measurement (reliability and validity)

[5] **Scaling and Index Construction** **to be presented in Unit 3**

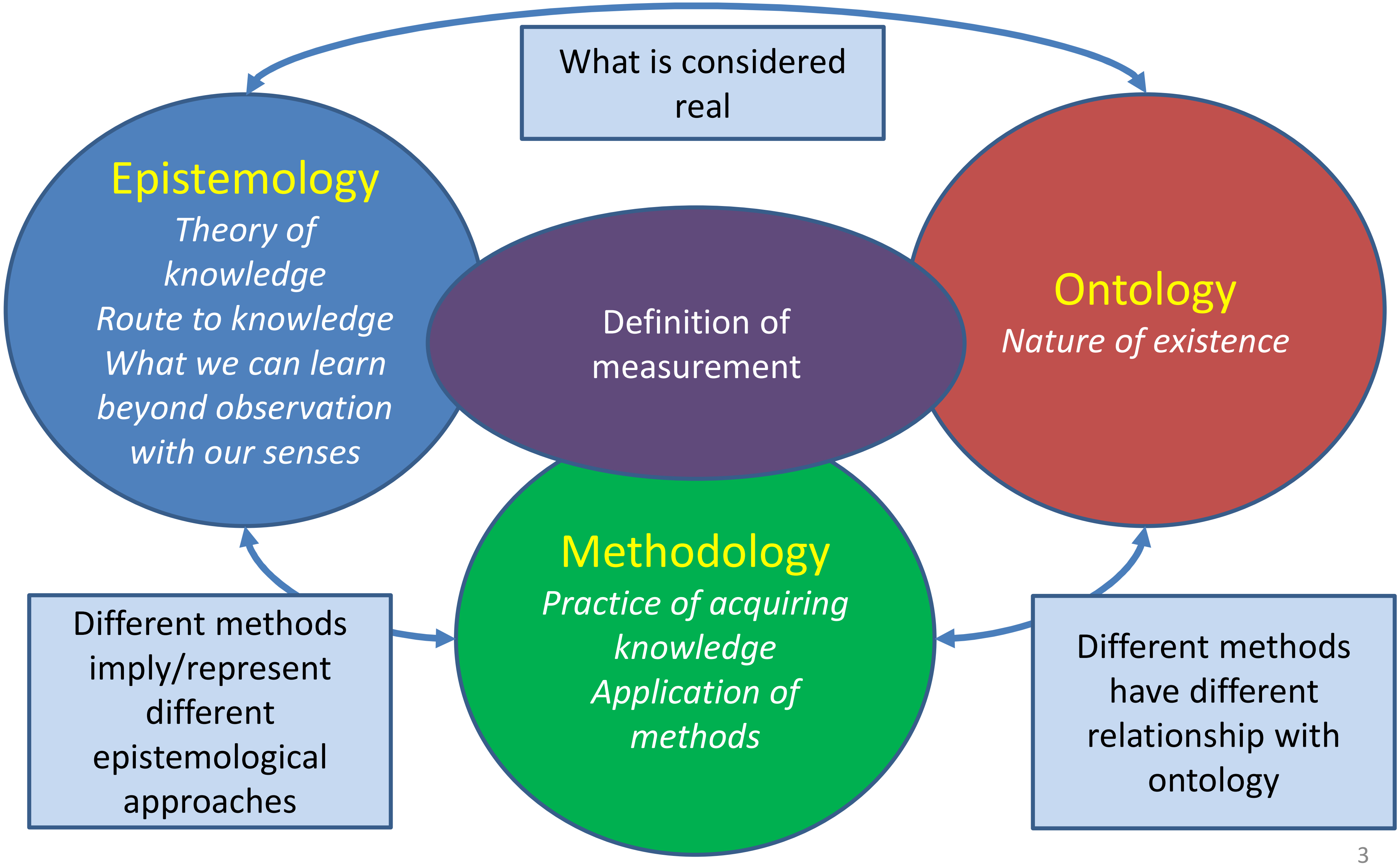
- Selected methods of scaling and index construction
  - Thurstone scaling
  - Likert scaling (additional literature by Likert)
  - Guttman scaling

Topic 4 **to be presented in Unit 3**
(Book chapter 4)

[4] **Survey Research**

- Very widely used method of data collection in the social sciences
- Principles of good survey research
- Types of surveys, how to select a survey method, how to construct a survey, what kind of questions are appropriate, how should they be phrased, how should a response scale look like, what are the pros and cons of survey research

Measurement



Definition of measurement

Prevalent definition of measurement in the social sciences

S.S. Stevens

p.677

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

Thurstone, L. L. (1927). The method of paired comparisons for social values. *The Journal of Abnormal and Social Psychology*, *21*(4), 384.

Best Worst Scaling

Rasch model / Item Response Theory

Thurstone, L. L. (1928). Attitudes can be measured. *American journal of Sociology*, *33*(4), 529-554.

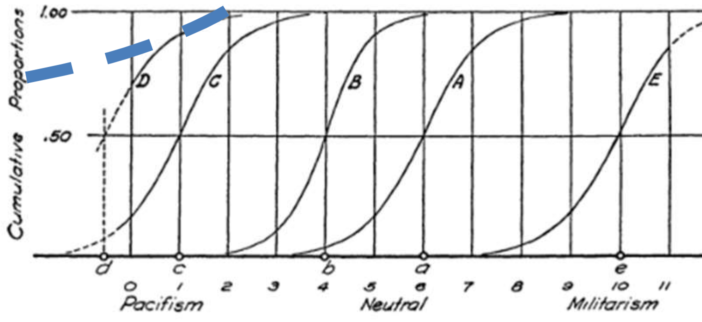


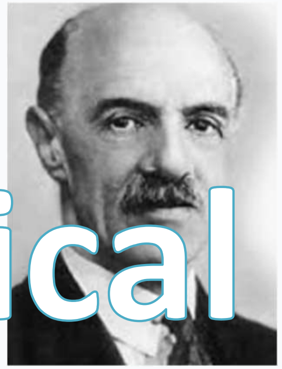
Fig. 2

Guttman, L. (1950), 'The Basis for Scalogram Analysis', in Samuel Andrew Stouffer, Louis Guttman, E.A. Suchman, P.F. Lazarsfeld, S.A. Star and J.A. Clausen (eds), Measurement and Prediction, The American Soldier, Vol. IV, New York: Wiley.

[5] **Scaling**

[3] **Measurement**

Spearman, C. (1904). The proof and measurement of association between two things. Am J Psychol 15: 72–101 [reprinted 1961 and 1987: Spearman, C. (1987). The proof and measurement of association between two things. *The American journal of psychology*, *100*(3/4), 441-471.]



Thurstone, L. L. (1934). The vectors of mind. *Psychological review*, *41*(1), 1.

Tests and Measurements in the Social Sciences. By Truman L. Kelley and A. C. Krey. (Charles Scribner's Sons, New York,1934).

Thurstone, L. L. (1947). *Multiple-factor analysis; a development and expansion of The Vectors of Mind.* University of Chicago Press.

Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley.

Classical Test Theory (CTT)

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*

[5] **Scaling**

Thurstone, L. L. (1929). Fechner's law and the method of equal appearing intervals. *Journal of Experimental Psychology*, *12*(3), 214.

[5] **Scaling**

Thurstone, L. L. (1929). Theory of attitude measurement. *Psychological review*, *36*(3), 222.

Andrich, D. (1989). A probabilistic IRT model for unfolding preference data. *Applied Psychological Measurement*, *13*(2), 193-216.

Some milestones of social measurement

Seminal Papers in Marketing: Churchill (1979), Peter (1979)

Measure and Construct Validity Studies

GILBERT A. CHURCHILL, JR.*

A critical element in the evolution of a fundamental body of knowledge in marketing, as well as for improved marketing practice, is the development of better measures of the variables with which marketers work. In this article an approach is outlined by which this goal can be achieved and portions of the approach are illustrated in terms of a job satisfaction measure.

A Paradigm for Developing Better Measures of Marketing Constructs

Churchill Jr, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of marketing research*, *16*(1), 64-73.

Journal of Marketing Research
Vol. XVI (February 1979), 64-73

Reliability Studies

J. PAUL PETER*

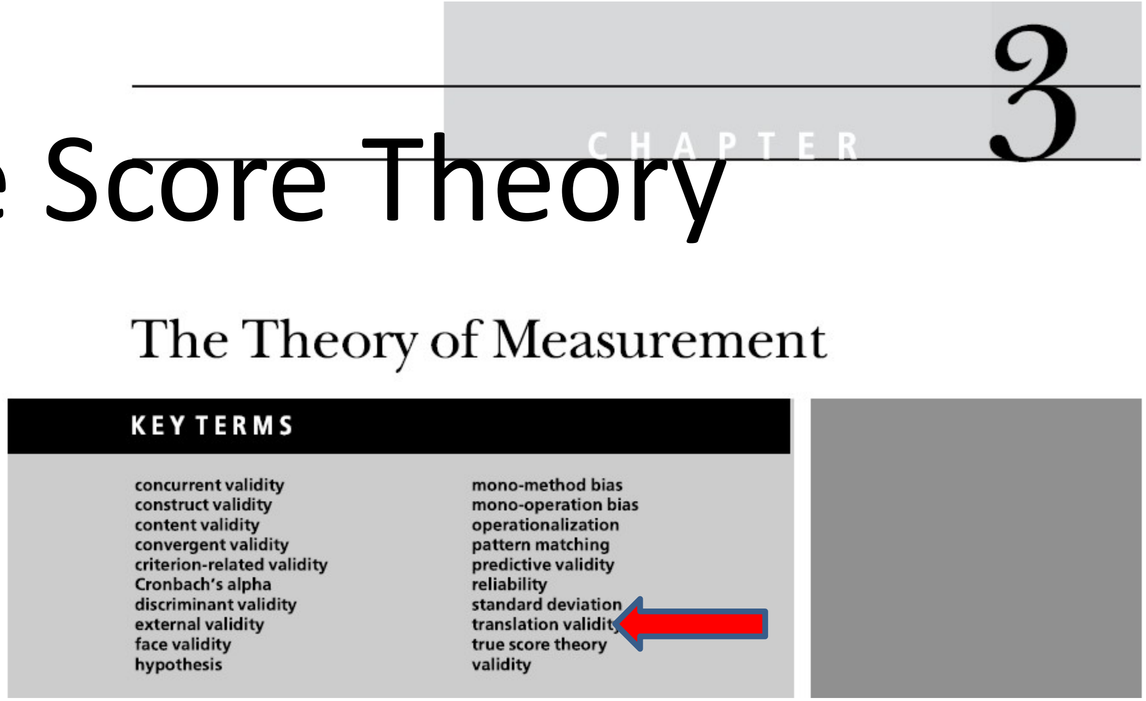
The basic theories and measurement procedures for reliability and the closely related concept of generalizability are reviewed, illustrated, and evaluated for use in marketing research. A critique is given of a subset of previous marketing research studies in which reliability estimates were used and recommendations are made for future research.

Reliability: A Review of Psychometric Basics and Recent Marketing Practices

Peter, J. P. (1979). Reliability: A review of psychometric basics and recent marketing practices. *Journal of marketing research*, *16*(1), 6-17.

Journal of Marketing Research
Vol. XVI (February 1979), 6-17

CTT = True Score Theory



CHAPTER 3

The Theory of Measurement

KEY TERMS

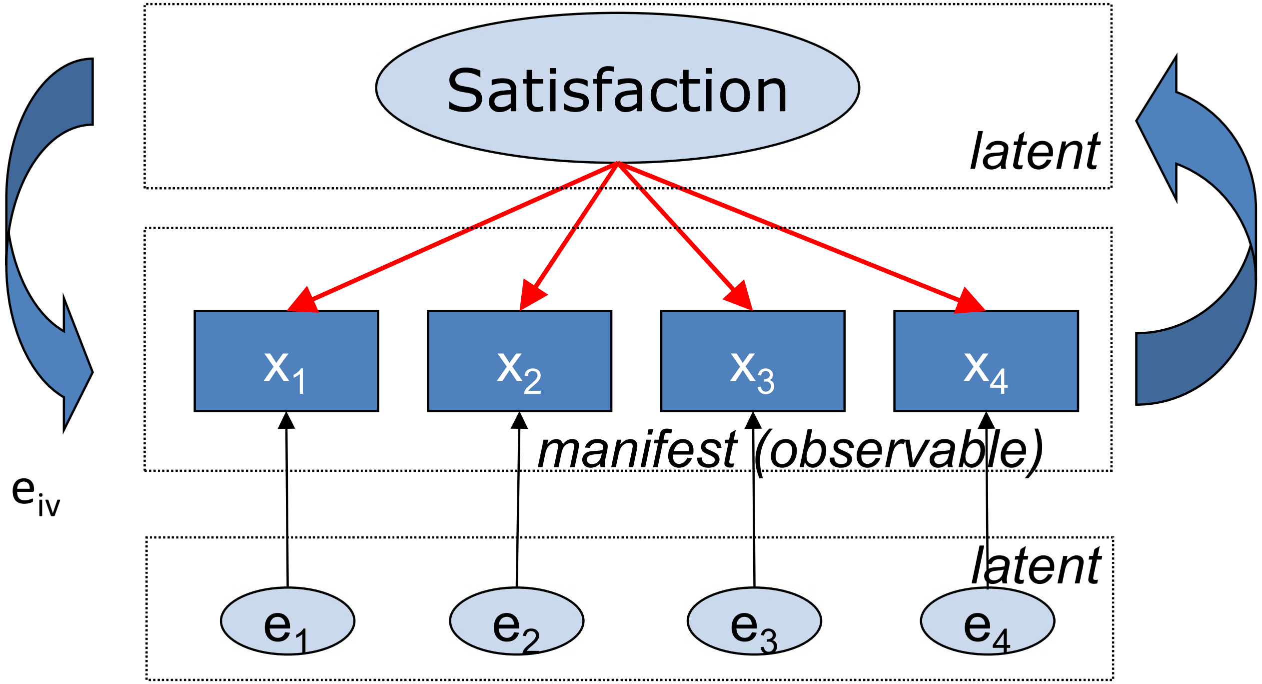
concurrent validity
construct validity
content validity
convergent validity
criterion-related validity
Cronbach's alpha
discriminant validity
external validity
face validity
hypothesis
mono-method bias
mono-operation bias
operationalization
pattern matching
predictive validity
reliability
standard deviation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

- CTT: true score T, observed score X, error score E
  - X_v = T_v + E_v (at the total score level for person v)
  - *No latent variable explicitly accounted for*
- Apply logic of CTT at the item level but introduce a latent variable
  - $x_{iv} = \tau_i + \lambda_i * F_v + e_{iv}$
    (item i; latent variable/factor F) – *latent variable theory*

Factor analytic model / Reflective indicator model / Scaling

- Indicators (=manifest variables) *reflect* the latent variable (e.g., satisfaction)
- The latent variable *causes* the manifest scores
- Manifest scores need to be correlated
  - Item-intercorrelations allow for inferring latent variable score (factor analysis)
- Series of regression analyses (each item regressed on latent variable)

Satisfaction becomes manifest in its indicators



Satisfaction is inferred from its indicators

$x_{iv} = \tau_i + \lambda_i * F_v + e_{iv}$

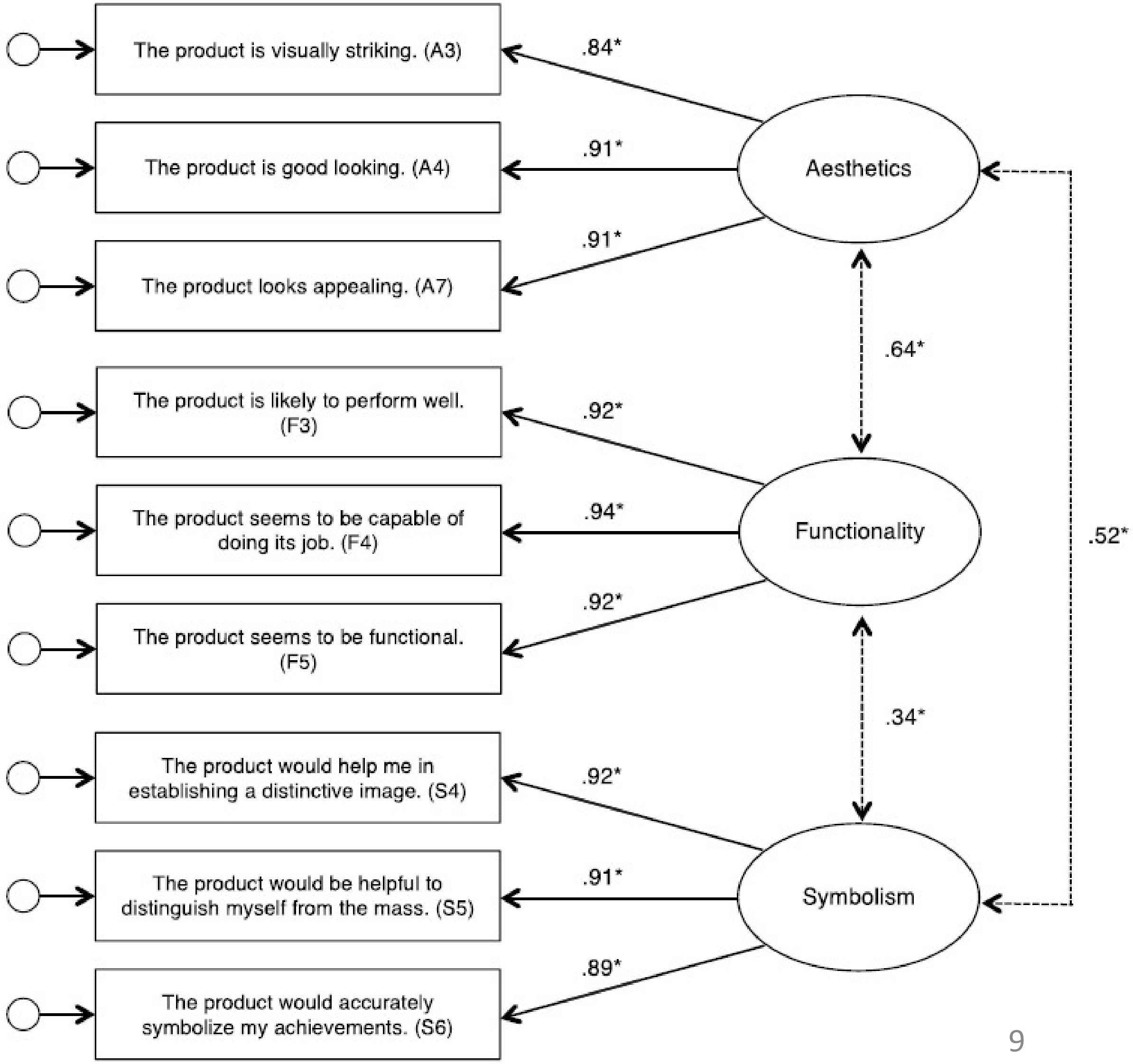
e: error score (residual)

Scaling (example)

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

Multidimensional construct

i.e. three interrelated unidimensional domains



Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*.

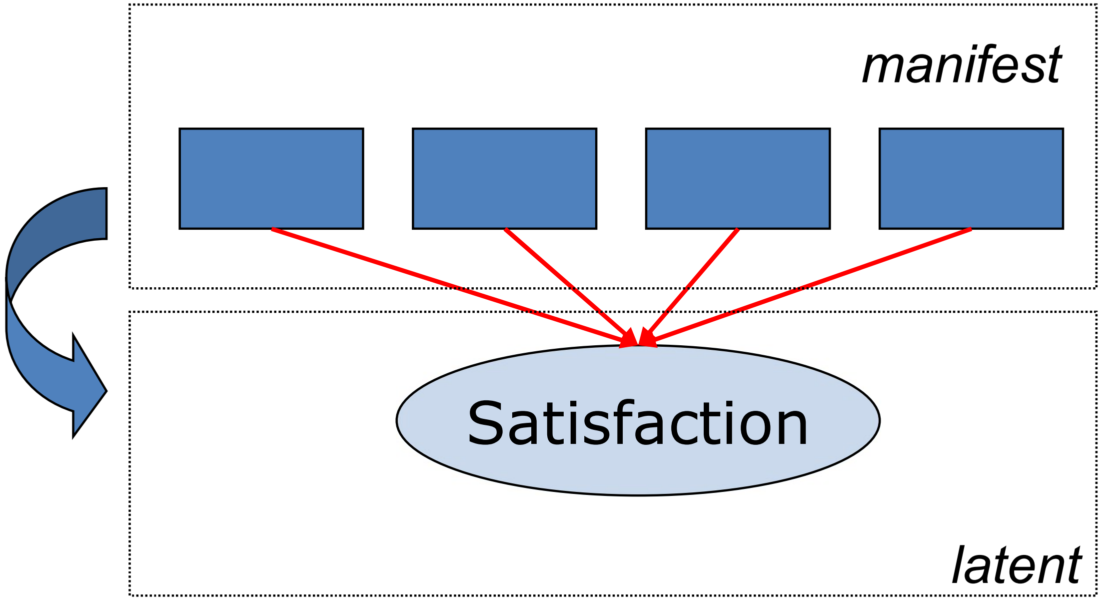
[5] **Scaling**

Formative indicators / Index formation

- Indicators (=manifest variables) *form* (constitute, define) the ‘latent’ variable
- The manifest scores *cause* the ‘latent’ variable
- Manifest scores are not required to be correlated (and in fact, should not be)
  - Item-intercorrelations cannot be explained by latent latent variable score
- One multiple regression analysis (latent variable regressed on all items)
- Alternative to scale development (*to be discussed*)

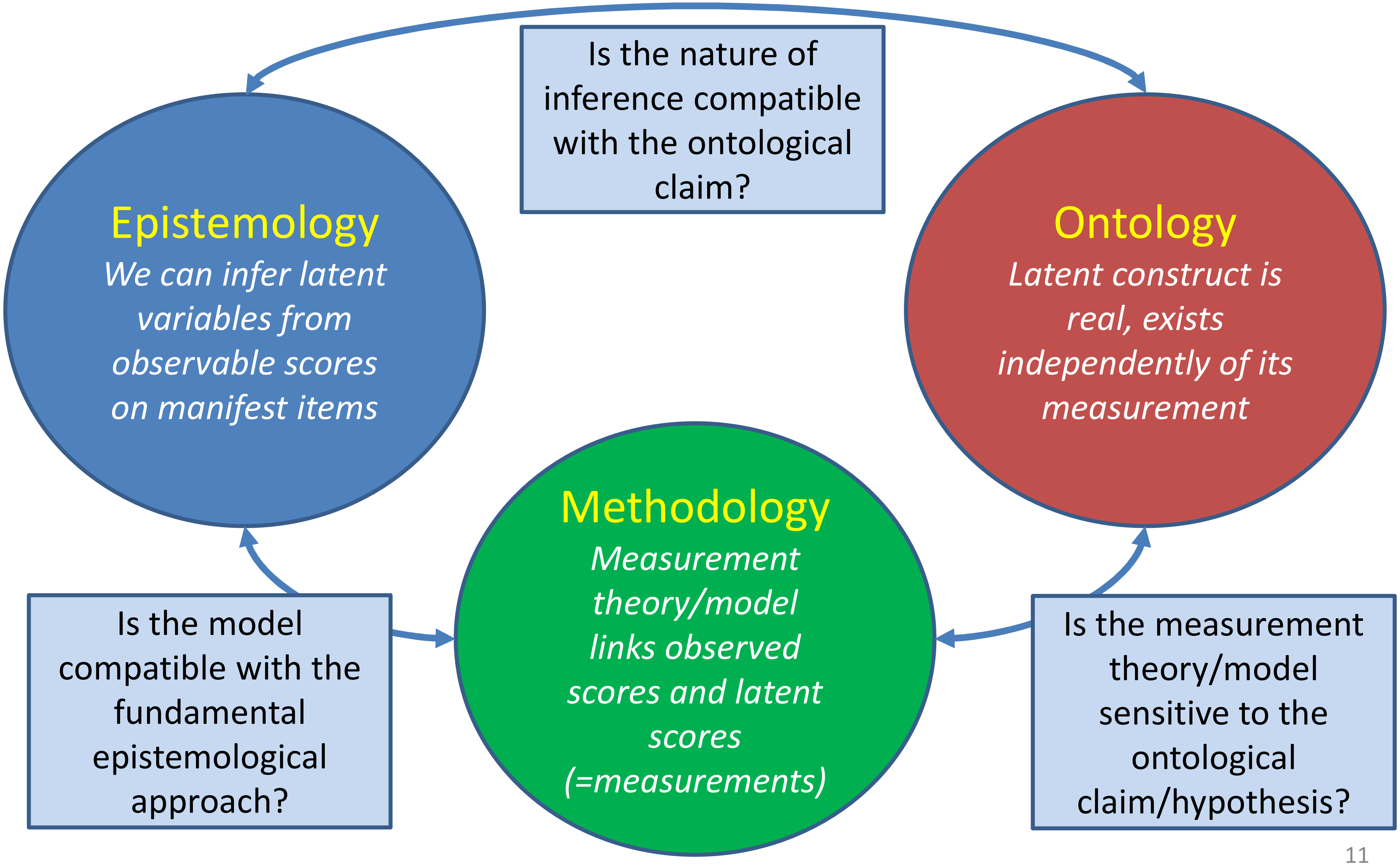
[1] **Theoretical Foundations**

Indicators have causal effect on the latent variable



[5] **Scaling and Index Construction**

Measurement of latent construct



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).

Ontology

The concept/construct
Unobservable but real
(fallible hypothesis)

Latent variable(s)

Represent the latent construct

Epistemology

Observable (operationalisation of the latent variable)

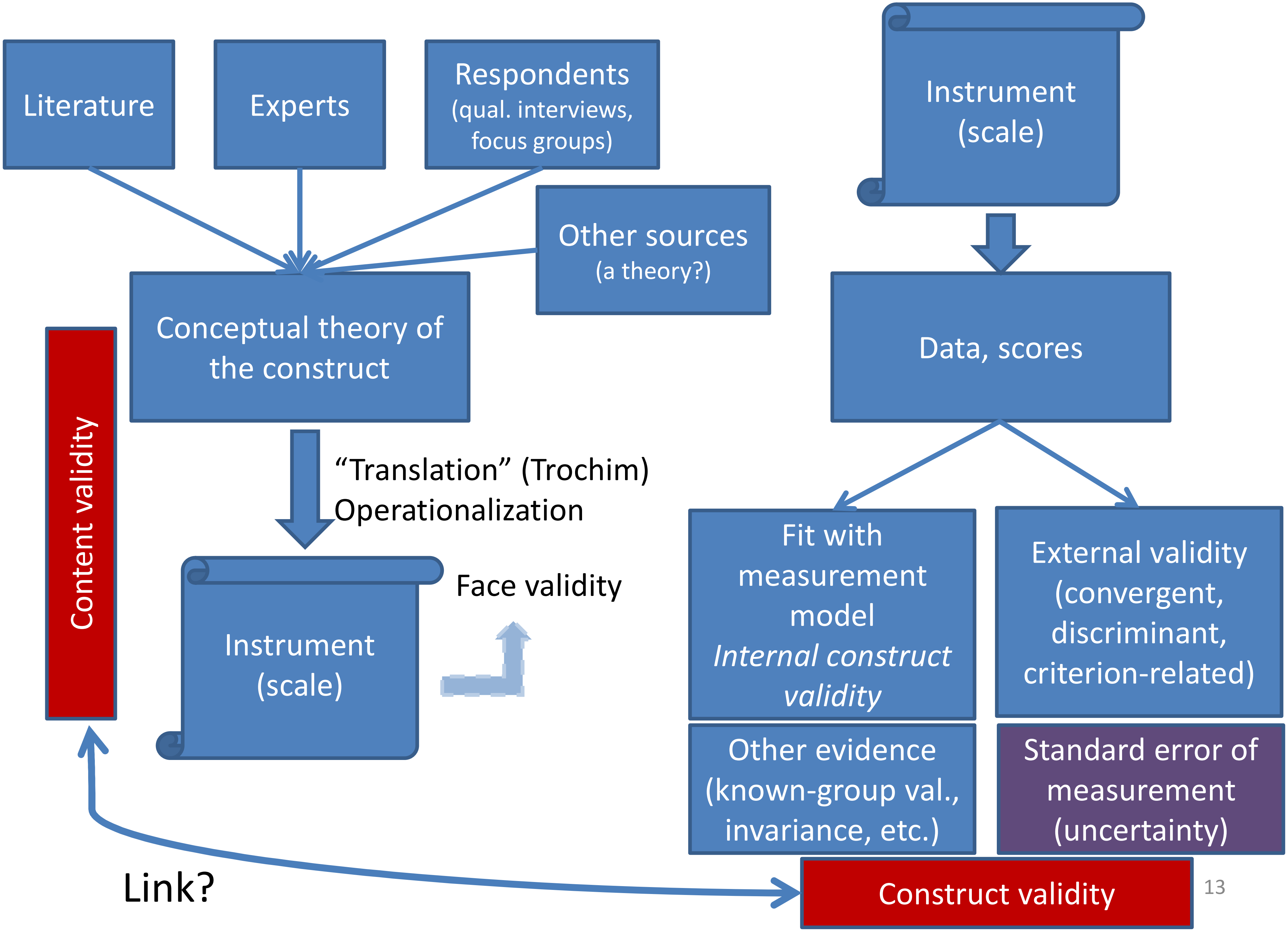
Relationship?

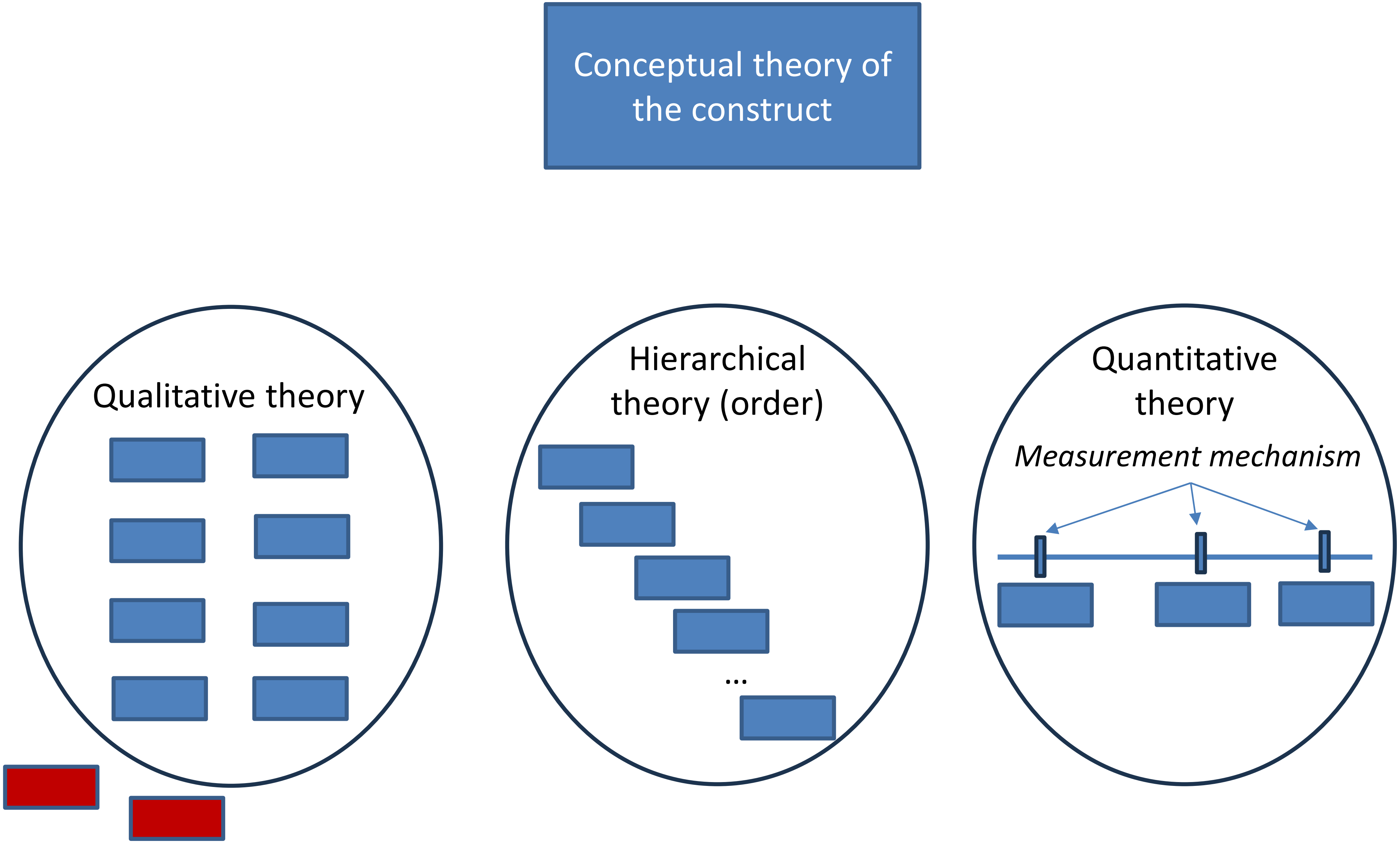
Items (manifest variables)
(questions or statements with response scale)

Measurement theory, measurement model

Items form a scale
(or multiple scales in case of multidimensional constructs)

Methodology


Literature
Experts
Respondents
(qual. interviews, focus groups)
Other sources
(a theory?)
Conceptual theory of the construct
Content validity
“Translation” (Trochim)
Operationalization
Instrument
(scale)
Face validity
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)
Link?
Construct validity


Conceptual theory of the construct
Qualitative theory
Hierarchical theory (order)
...
Quantitative theory
Measurement mechanism

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 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)

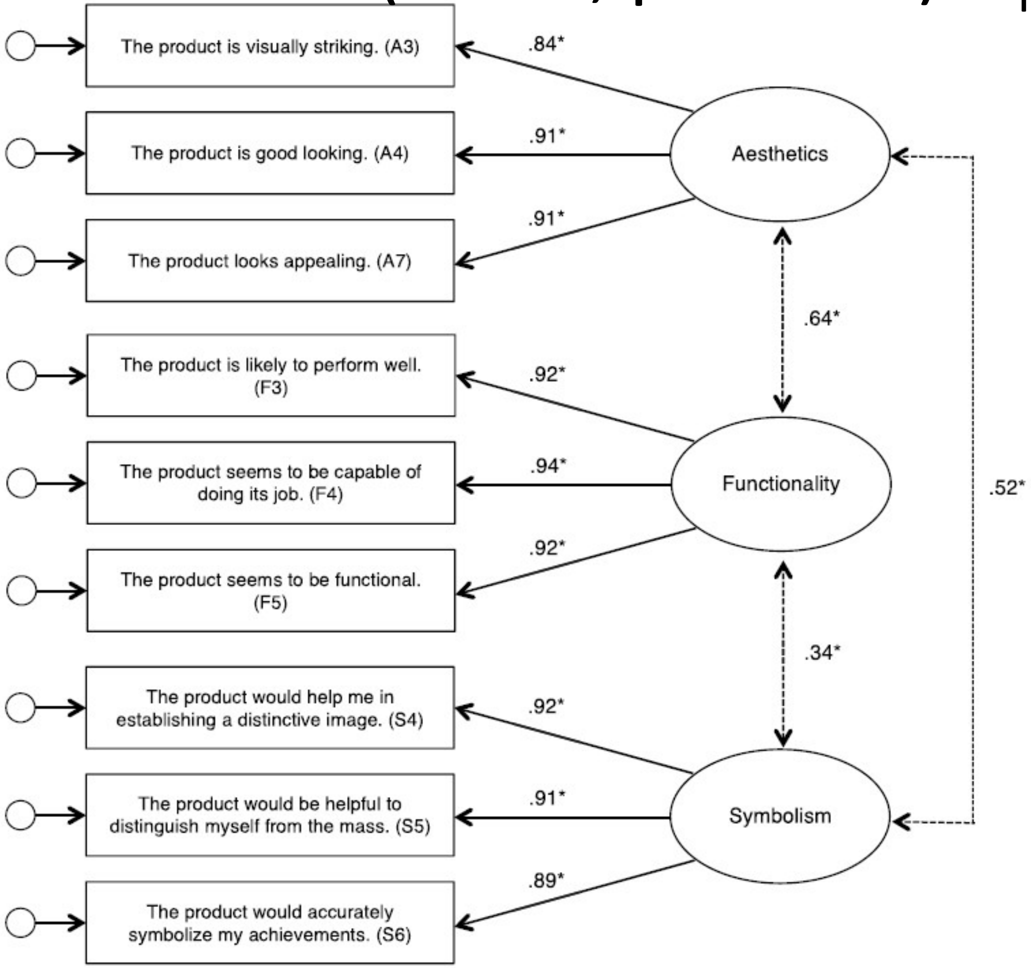
$$X_{ik} = [\alpha_k + \beta_k(T_i)] + E_{ik}$$

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)

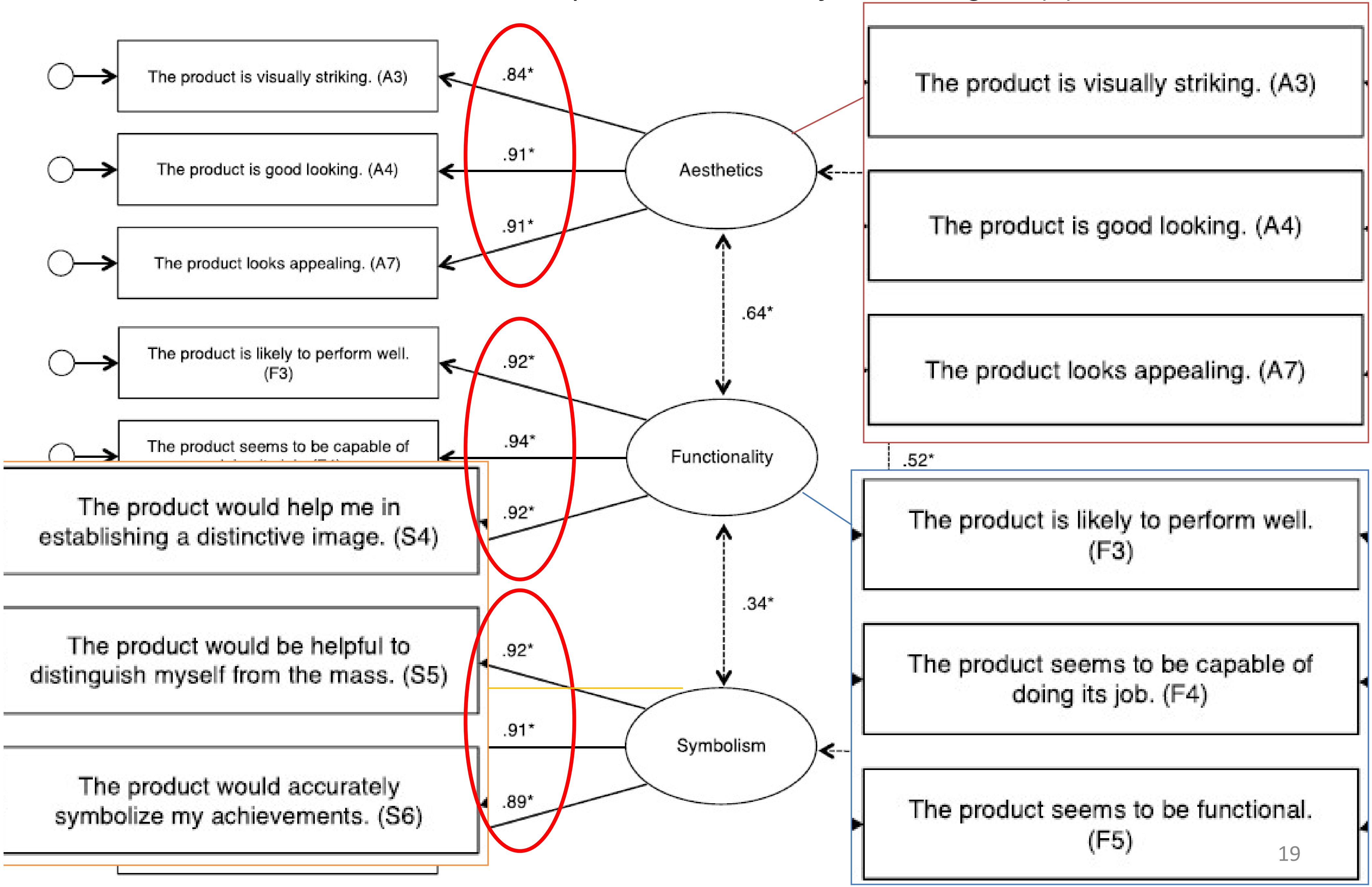
CTT at the item level (item i, person v): $x_{iv} = \tau_i + \lambda_i * F_v + e_{iv}$



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

Average variance extracted: $AVE_{aesthetic} = .79$; $AVE_{functional} = .86$; $AVE_{symbolic} = .83$.

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

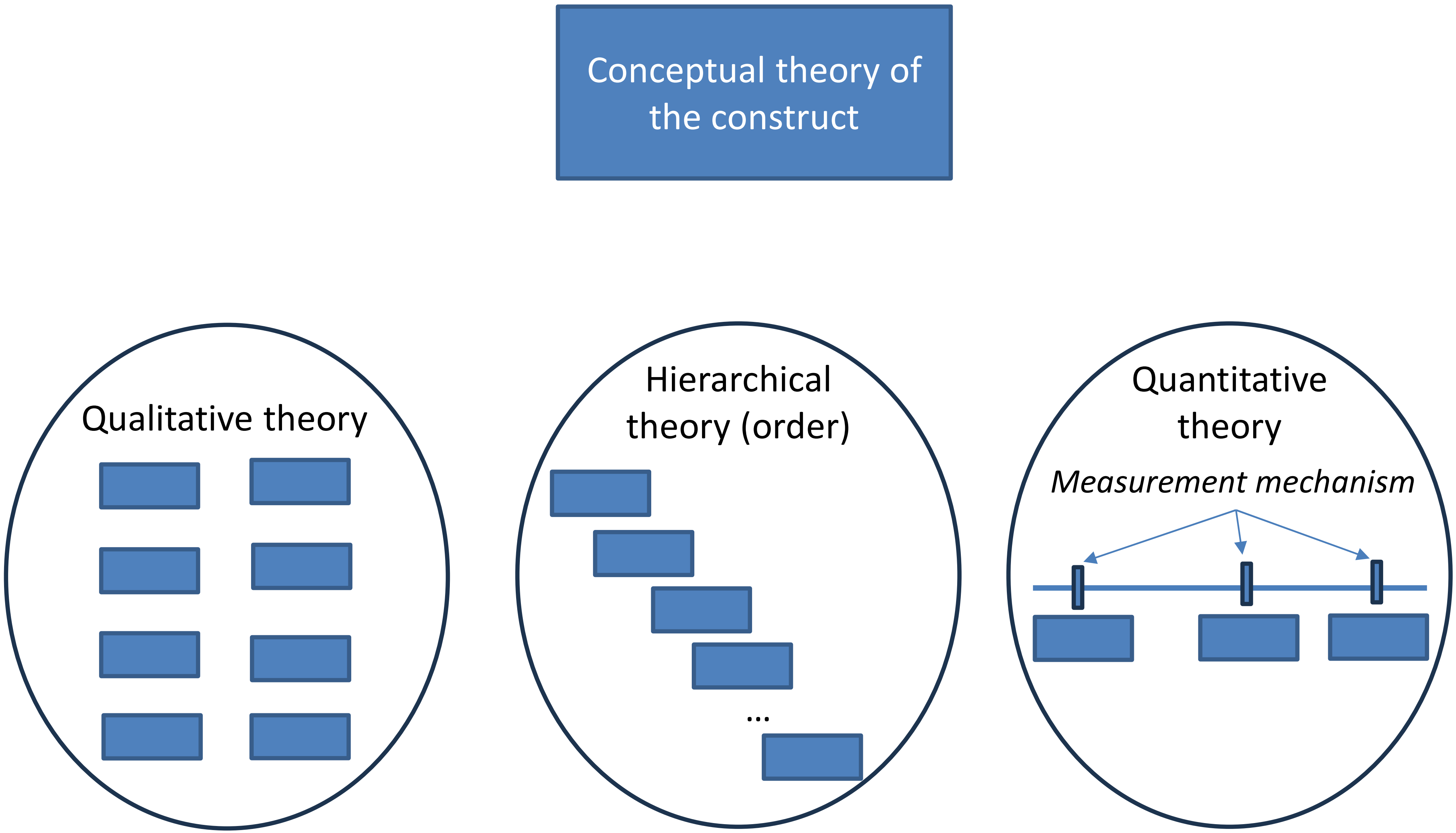


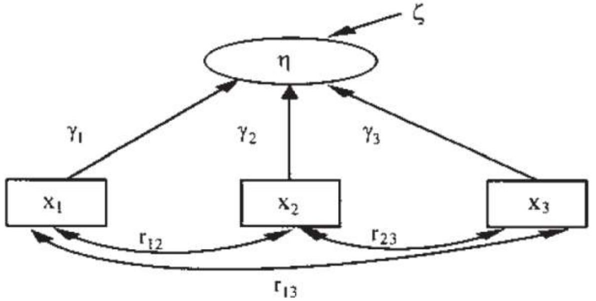
Homburg, C., Schwemmle, M., & Kuehnl, C. (2015). New product design: Concept, measurement, and consequences. *Journal of Marketing*, *79*(3), 41-56.

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

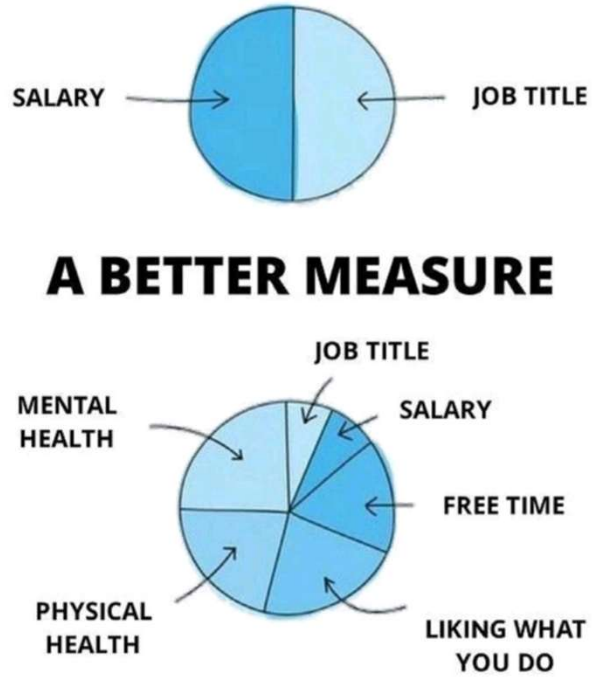
Item No.	The product...	M	SD
A3	...is visually striking	2.26	1.13
A4	...is good looking	2.08	1.07
A7	...looks appealing	2.10	1.07
F3	...is likely to perform well	1.86	.88
F4	...seems to be capable of doing its job	1.76	.85
F5	...seems to be functional	1.78	.86
S4	...would help me in establishing a distinctive 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.


Conceptual theory of the construct
Qualitative theory
Hierarchical theory (order)
...
Quantitative theory
Measurement mechanism



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; but see also Partial Least Squares [PLS])

– measurement error?

Measurement of a latent variable from a realist perspective

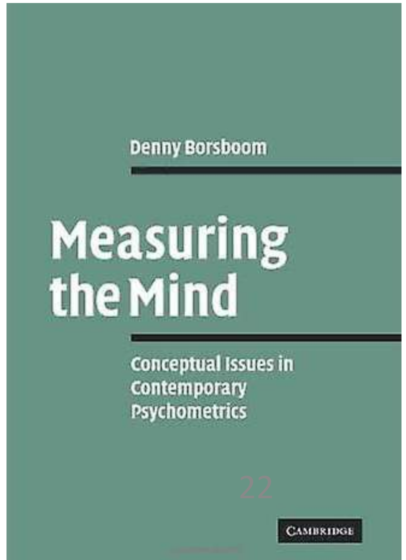
– *Ontological claim: latent variable exists as a quantitative property independent of our measurement*

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

(1) $$\eta \equiv x.$$

- 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, manifest item scores are measurements;
  the concept η is a summary of observables)

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) $$\eta = \gamma_1 x_1 + \gamma_2 x_2 + \dots + \gamma_n x_n,$$

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

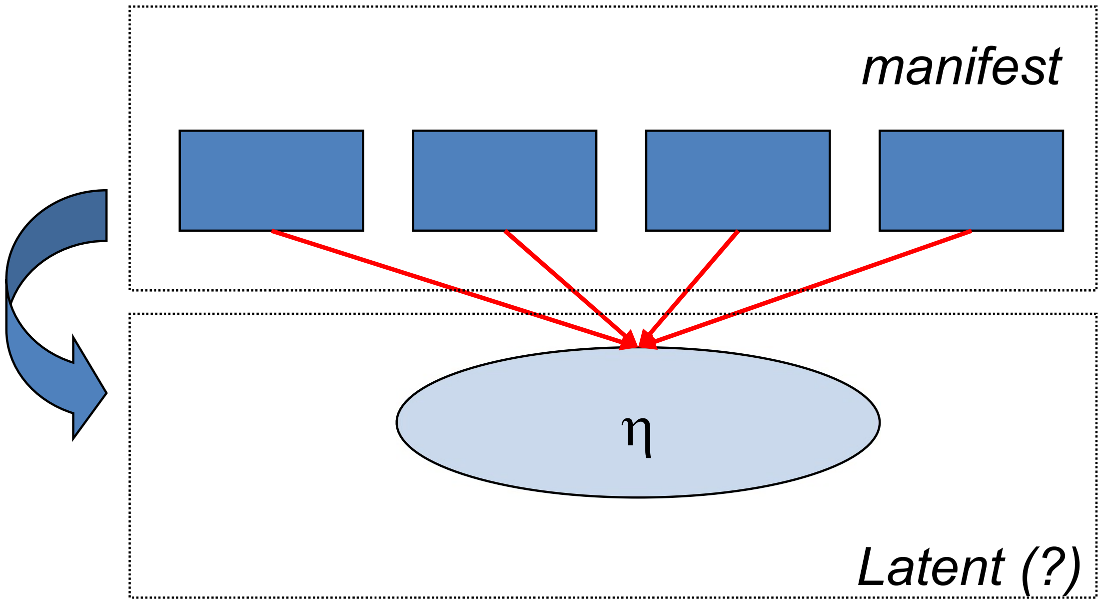
Another formative specification (shown in Figure 1 for n = 3) is provided by Bollen and Lennox (1991, p. 306):

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

Formative Model

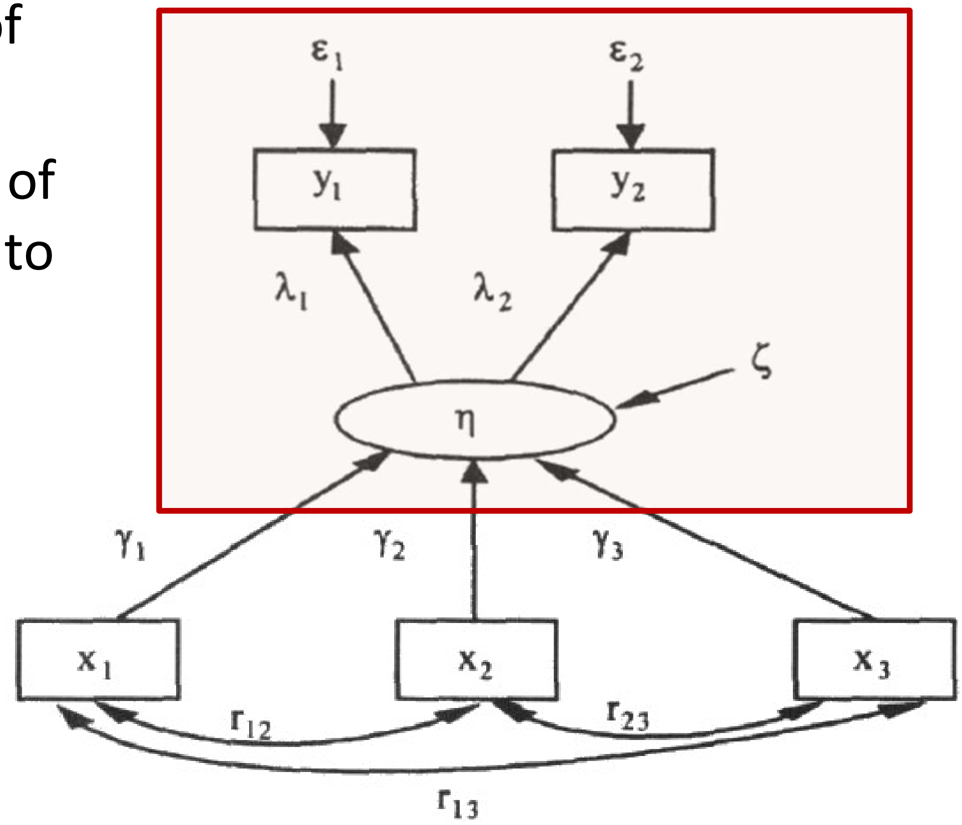
- Landmark article by Diamantopoulos & Winklhofer (2001)
- Scale development (reflective indicators)
- Alternative: index construction
  - Items as formative indicators
  - Indicators have causal effect on latent variable (reversed causality)

Indicators have causal effect on the latent variable



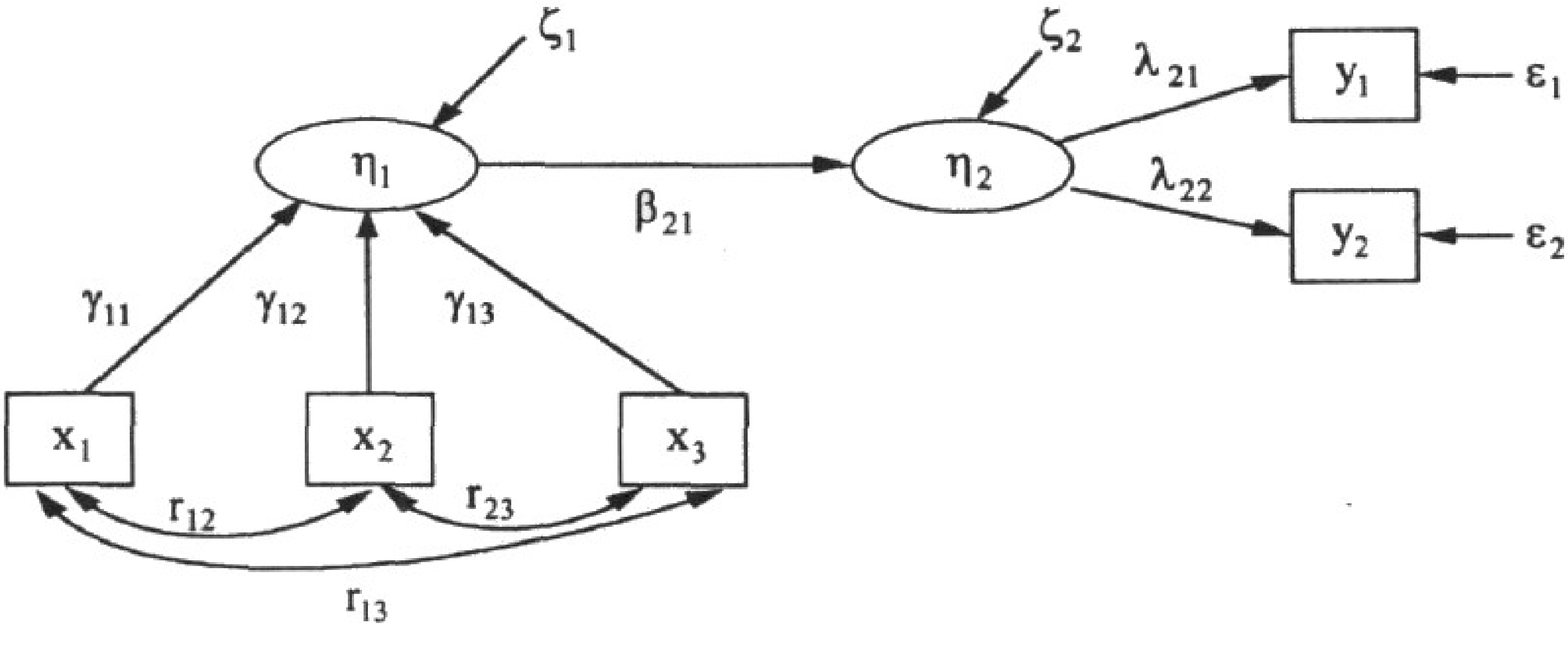
Estimation of Formative Model: MIMIC-Model

- 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

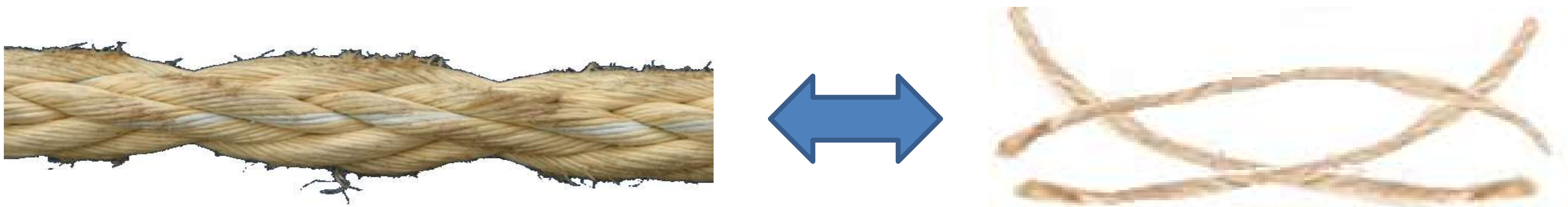


Key question

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

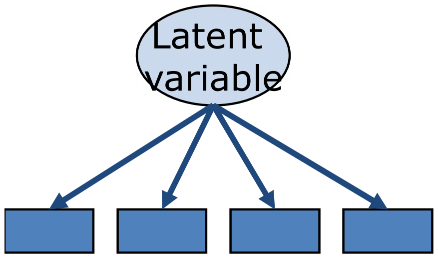
- Indexes summarize multiple variables, or combine them into one composite variable
  - If this is what you want to achieve, then indexes are fine.

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



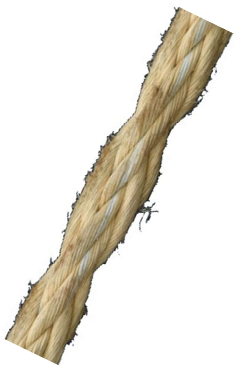
Key differences between an index (formative "measure") and a scale ("reflective measure")

Scale



- Reflective indicator model: items are causally impacted by a latent variable that is supposed to exist independently of the items (critical realism)
- *Caveat: measurement model/theory may not be very sensitive towards ontological claim*

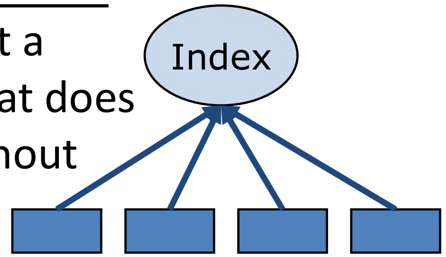
- Items must be correlated
- Combining domains into one overarching summary variable requires domains to be reasonably correlated; summary variable/second order factor accounts for these correlations



- **Purpose: measurement**

Index

- Formative indicator model: items causally impact a summary variable that does not exist as such without forming the index (constructivism)



- Combining components does not require components to be correlated
- **Purpose: Summarizing measurements** (index presupposes measurements)

Key differences between an index (formative "measure") and a scale ("reflective measure")

Scale

- Ultimate criterion: validity
- Attempt can fail (either latent variable does not exist/is not quantitative, scale items are unsuitable, or context of measurement is unsuitable)
- Relationship of latent variable and item empirically estimated (should, in principle, be invariant – objective)

Index

- Index is essentially a rule of how to combine measurements
- It is "correct" by definition
- It may or may not be useful for some purpose
- Problem of measurement (at the level of the components) can be evaded
- Relationship (i.e. impact of item on summary variable) can be defined arbitrarily
- Empirical estimation of path coefficients (from item to index variable) only possible for a given purpose (maximizing prediction or relationship between two indices in Partial least squares, PLS)

Key differences between an index (formative “measure”) and a scale (“reflective measure”)

Scale

- Scale may not contribute to solving a research problem
- Scale may be of limited use (e.g., applicable only in a given context of use)

Index

- Index can fail to predict what it was supposed to predict
- Index can be misleading (when issued as a measurement instrument)
- Index can be accidentally or deliberately deceptive

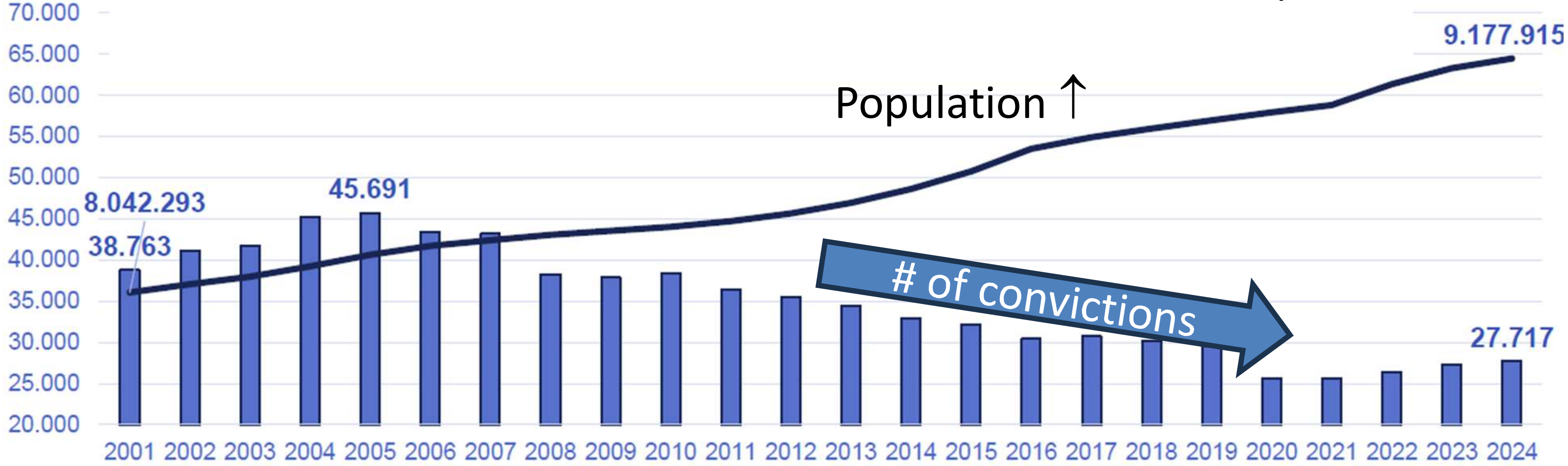
Misleading index

Crime statistics Austria

- Number of convicted criminals went down
- While size of population increased
- *Good news?!*

Index

- Index can be accidentally or deliberately deceptive



dema!nstitut #Demokratie für alle

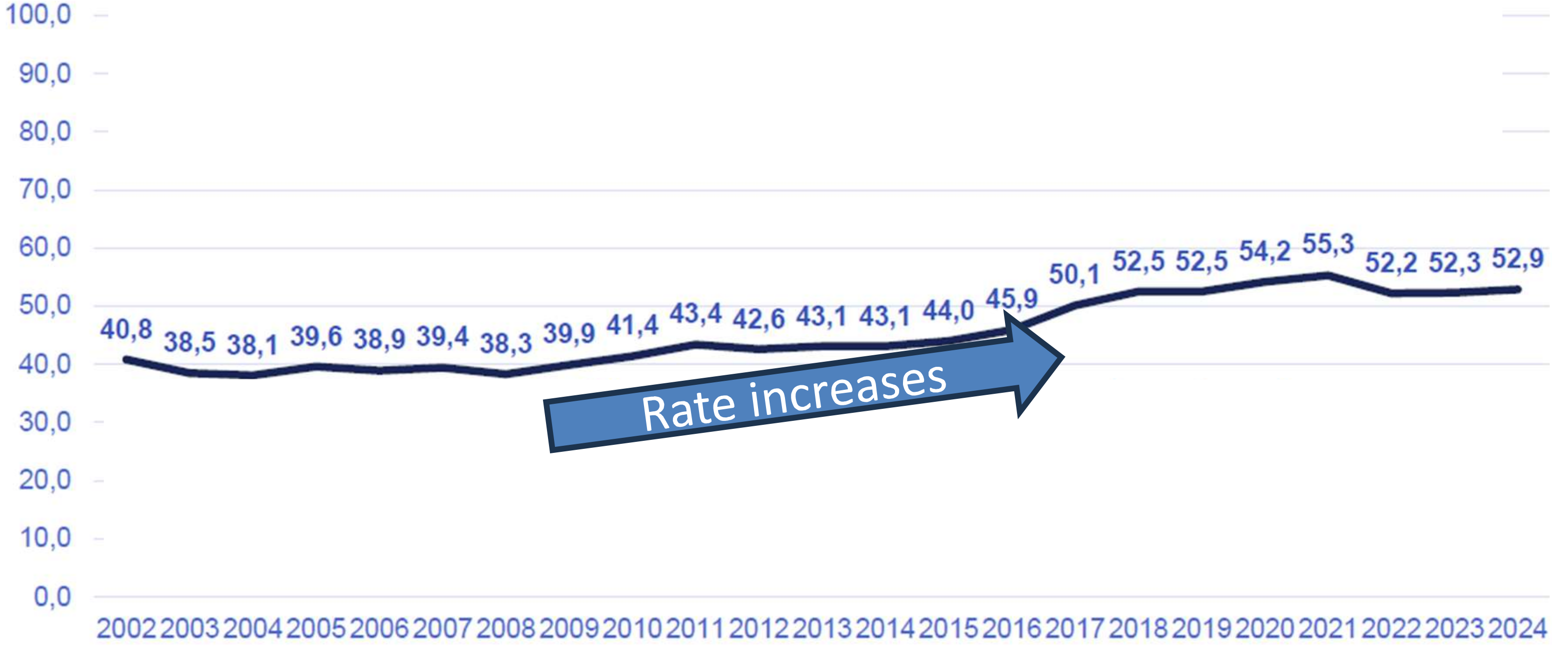
https://www.dema-institut.at/oesterreich-zaehlt-zu-den-sichersten-laendern-der-welt/

Misleading index

Index

Crime statistics Austria

- Since 2018, the *number of suspects* increased
- So, more crimes, fewer convictions?
- Suggests lower rate of cases solved
- Do we have a problem?



https://www.dema-institut.at/oesterreich-zaehlt-zu-den-sichersten-laendern-der-welt/

Misleading index

Index

Crime statistics Austria

- Do we have a problem?
- Not necessarily for the *number of suspects* is, since 2018, an index formed by:
  - Identified suspects
  - Number of criminal acts – *repeated offenses count multiple times*
  - Number of different crimes committed (at the same time) – *one act may contribute multiple times to the index*
  - Number of victims – *serial burglaries (cars, basements) count multiple times*
- Number of conviction cannot even come close to number of suspects as defined above
- In Vienna, only three offenders account for 9% of the suspects; three juvenile offenders even account for 32% of all young suspects

https://www.dema-institut.at/oesterreich-zaehlt-zu-den-sichersten-laendern-der-welt/

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

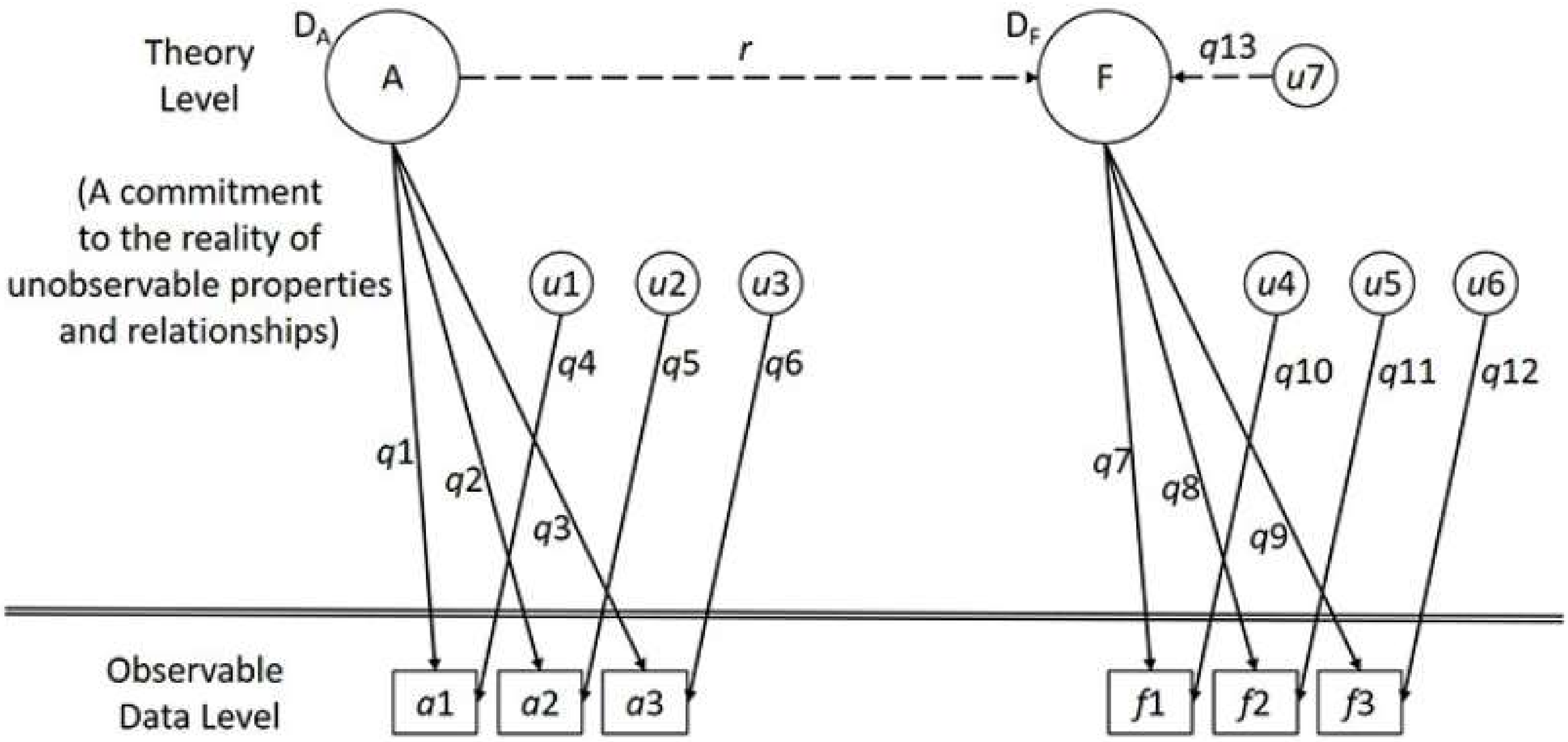


Figure 1. **(a)**

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

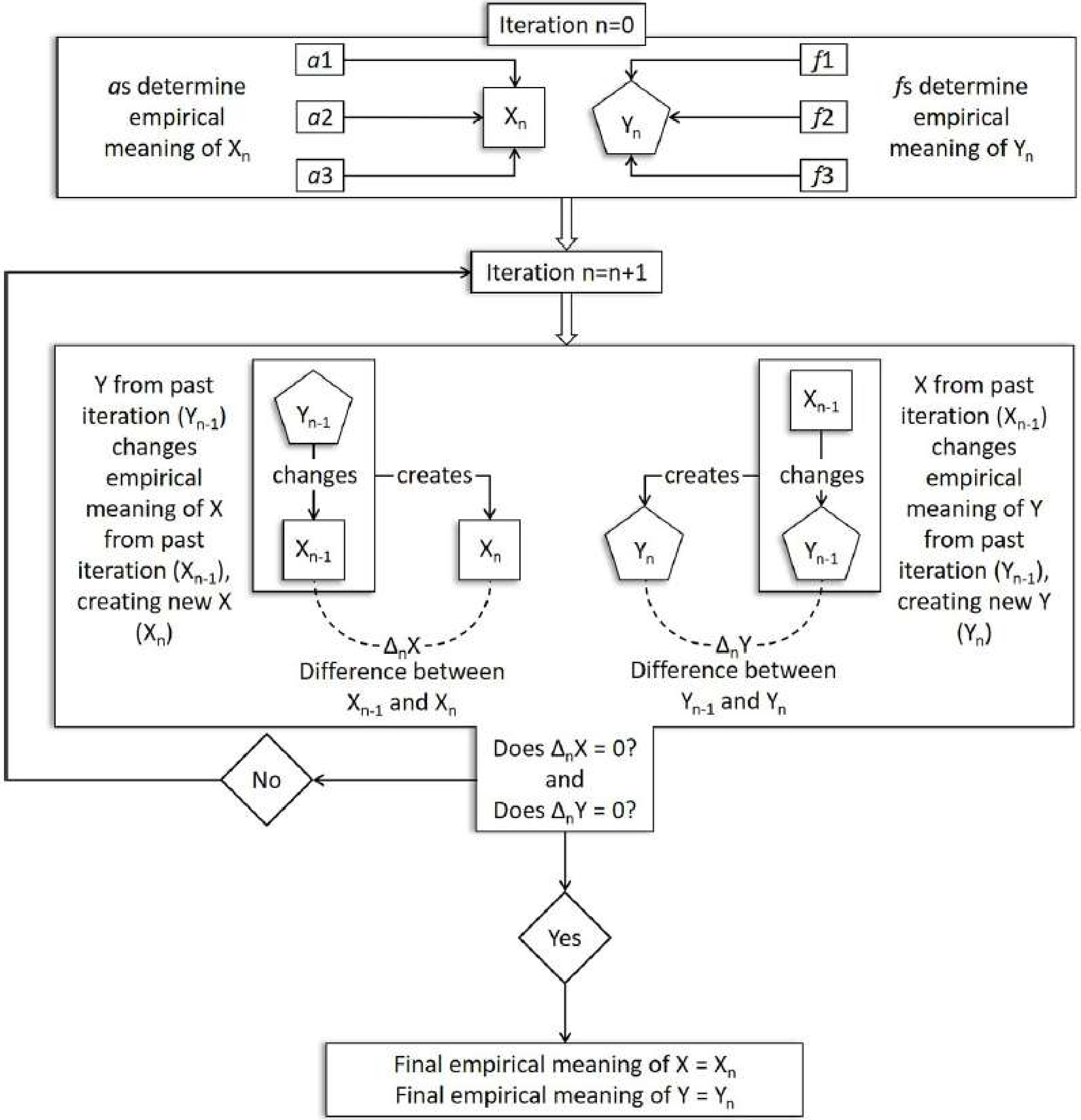


Figure 2.
How PLS gives its composites empirical meanings

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

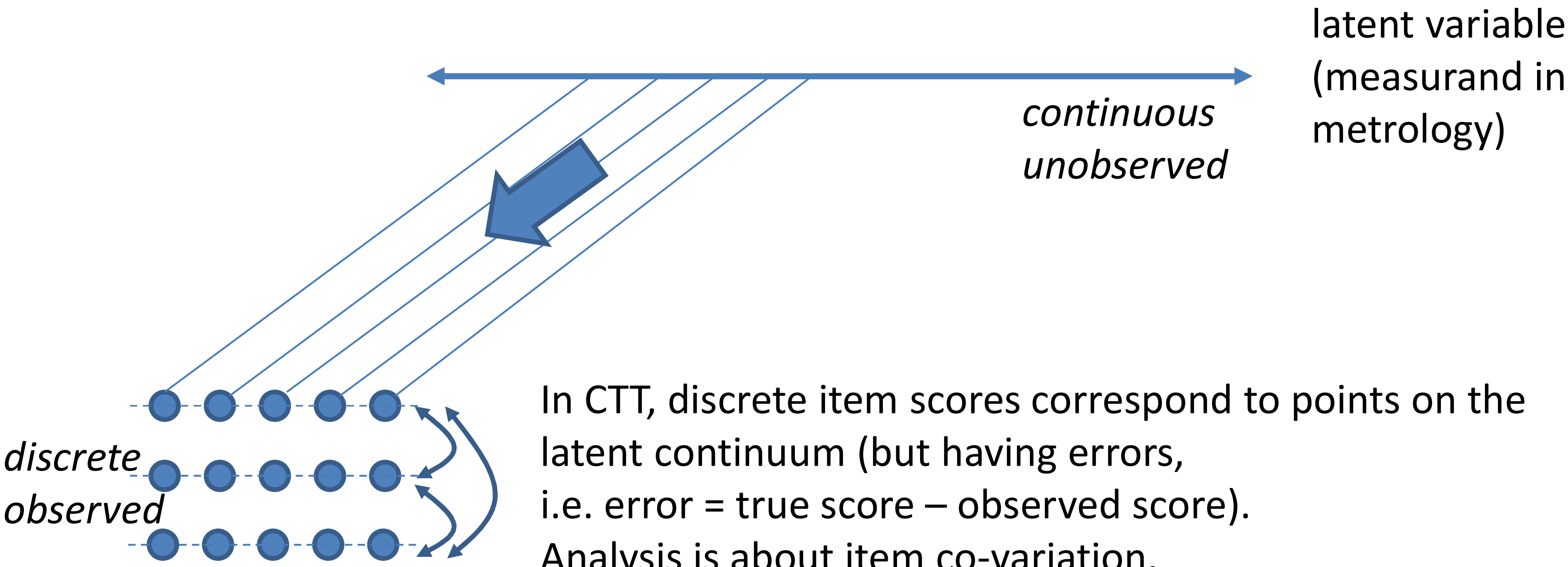
A miracle of measurement or accidental constructivism? How PLS subverts the realist search for truth

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

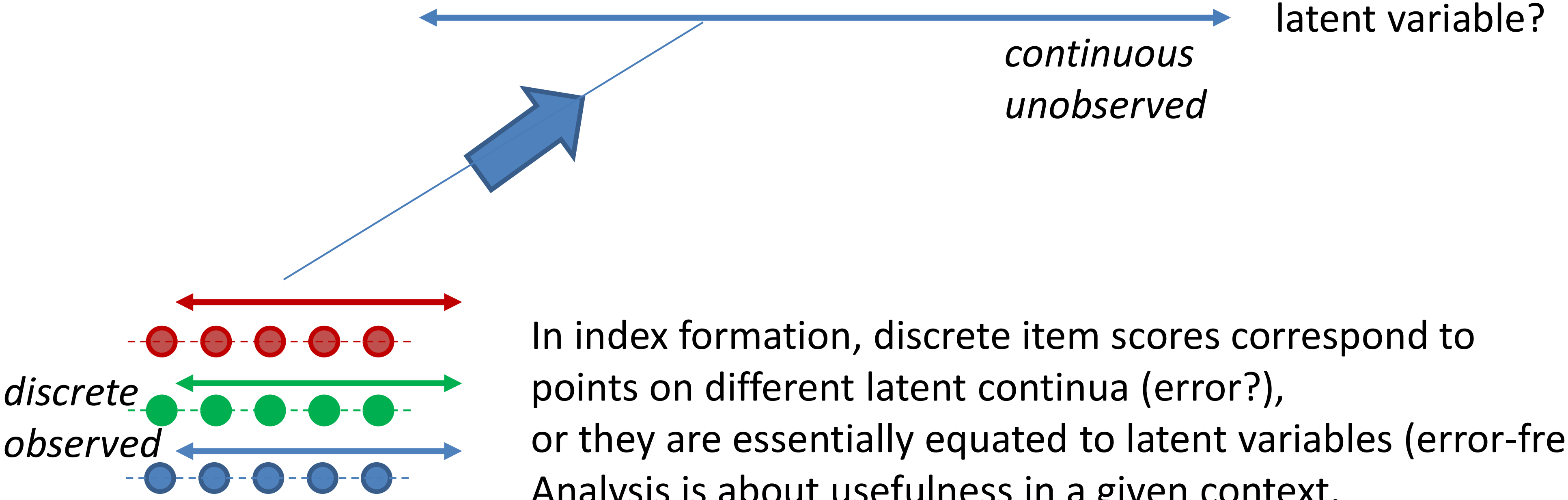
- PLS then examines the change in the scores produced by the new and old X (Δ_nX) and the new and old Y (Δ_nY) and asks "do the new X and Y variables have the same empirical meanings as the old X and Y scores?"
- If the answer is no, PLS goes through a second iteration (n = 2). It calculates the relationships 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.

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.

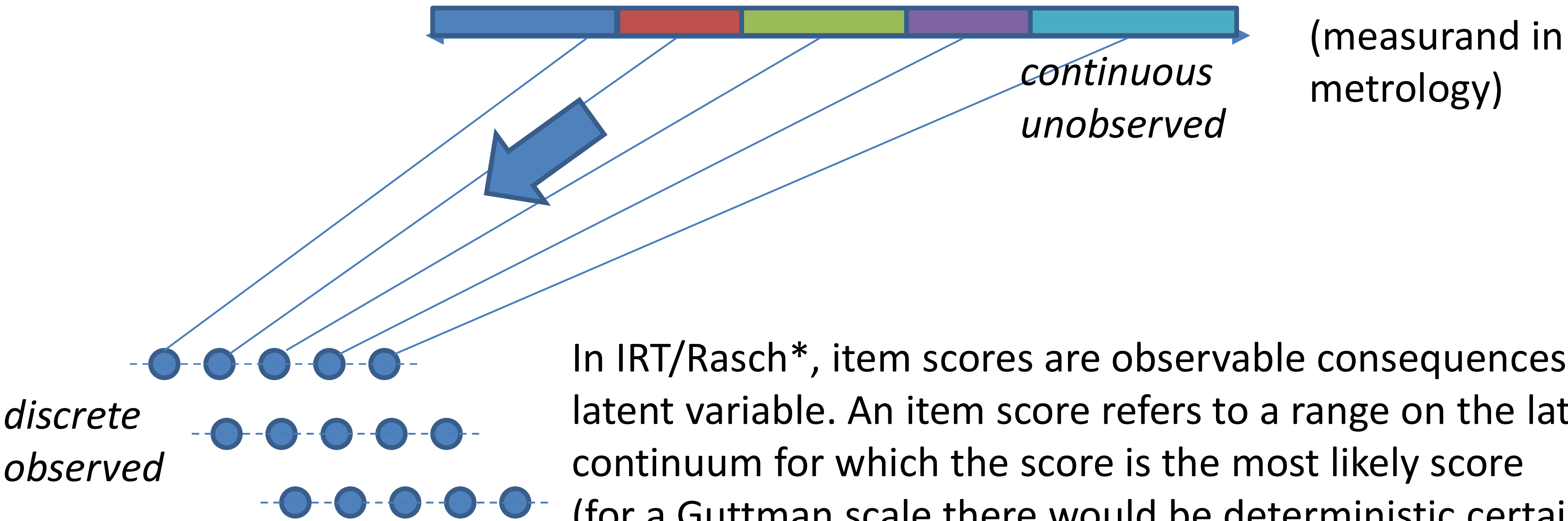
Elements in Measurement II: Index



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.

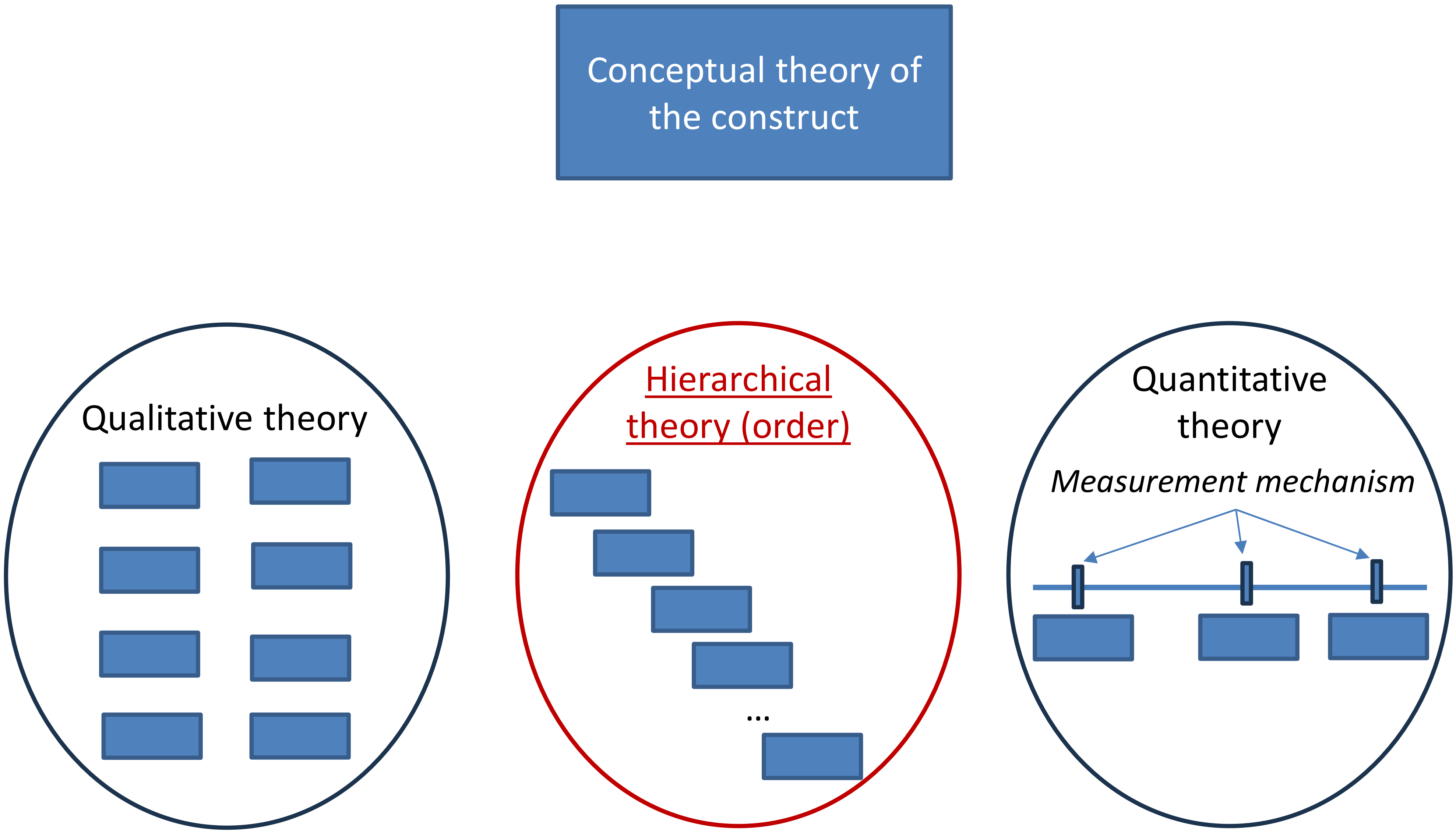
Elements in Measurement III: IRT/RMT

(Item response theory; Rasch measurement theory)



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.*


Conceptual theory of the construct
Qualitative theory
Hierarchical theory (order)
...
Quantitative theory
Measurement mechanism

Guttman model

- Two dichotomous items A and B
  - Yes/mastered/agreed 1
  - No/failed/disagreed 0
  - Item B harder than item A

	Item B – 0	Item B - 1
Item A - 0	Person < A	*[Guttman error]*
Item A - 1	A < Person < B	B < Person

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!

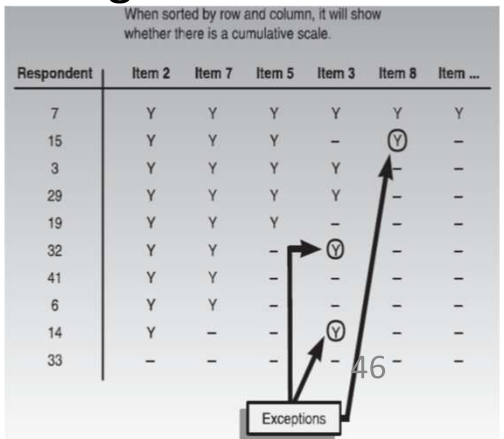
But there is no way to determine the distance between successive integer sum scores

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	–	–	–	–	–	–

Thurstone: Paired comparison

- Two dichotomous items A and B
  - Item B harder than item A   Item A   Item B
  - Which item is harder (has more of the property)?
  - In a Guttman scale, every person would say item B; except for some “Guttman errors”
  - However, the closer items A and B are located to each other, the more likely “errors” occur
  - Thurstone: errors allow for estimation of distance between items A and B based on normally distributed perception errors

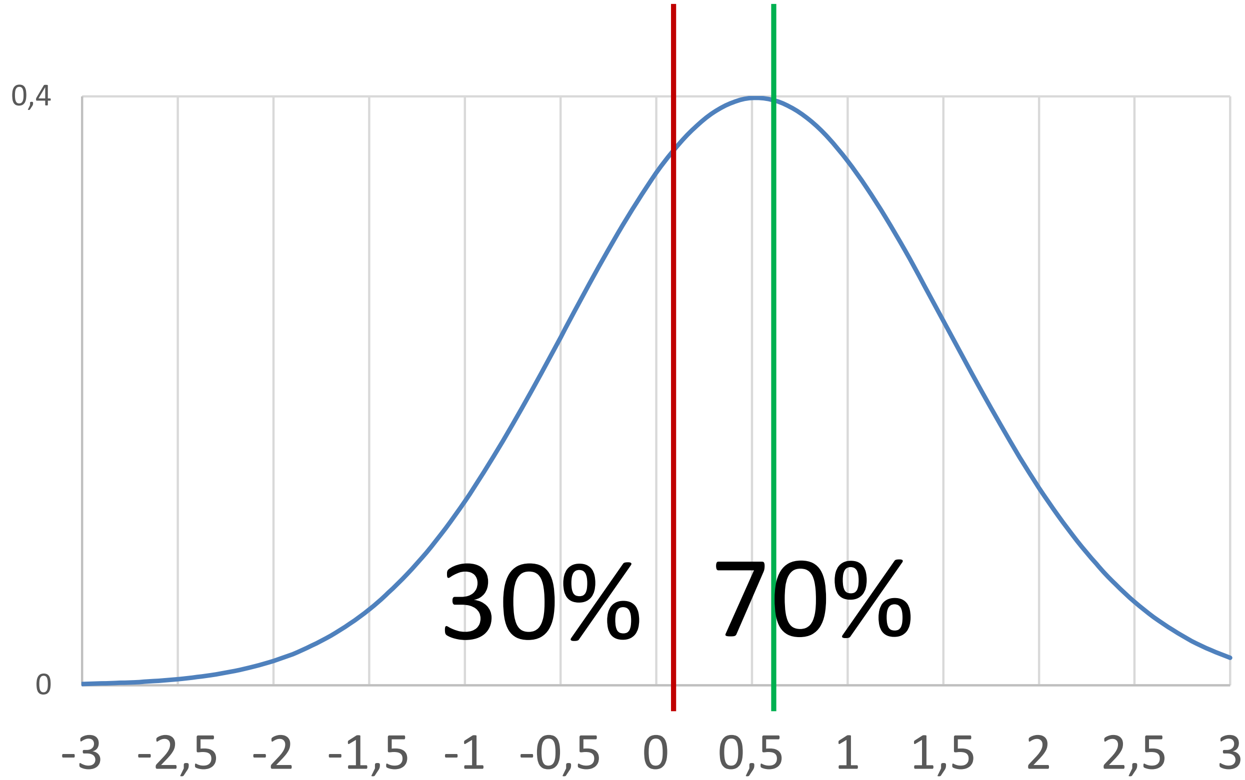
Thurstone: Paired comparison

Thurstone: errors allow for estimation of distance between items A and B

- 70 persons perceive item A < item B
- 30 persons perceive item A > item B

- $z_{30\%}$=-0.524

0.524: distance between A and B



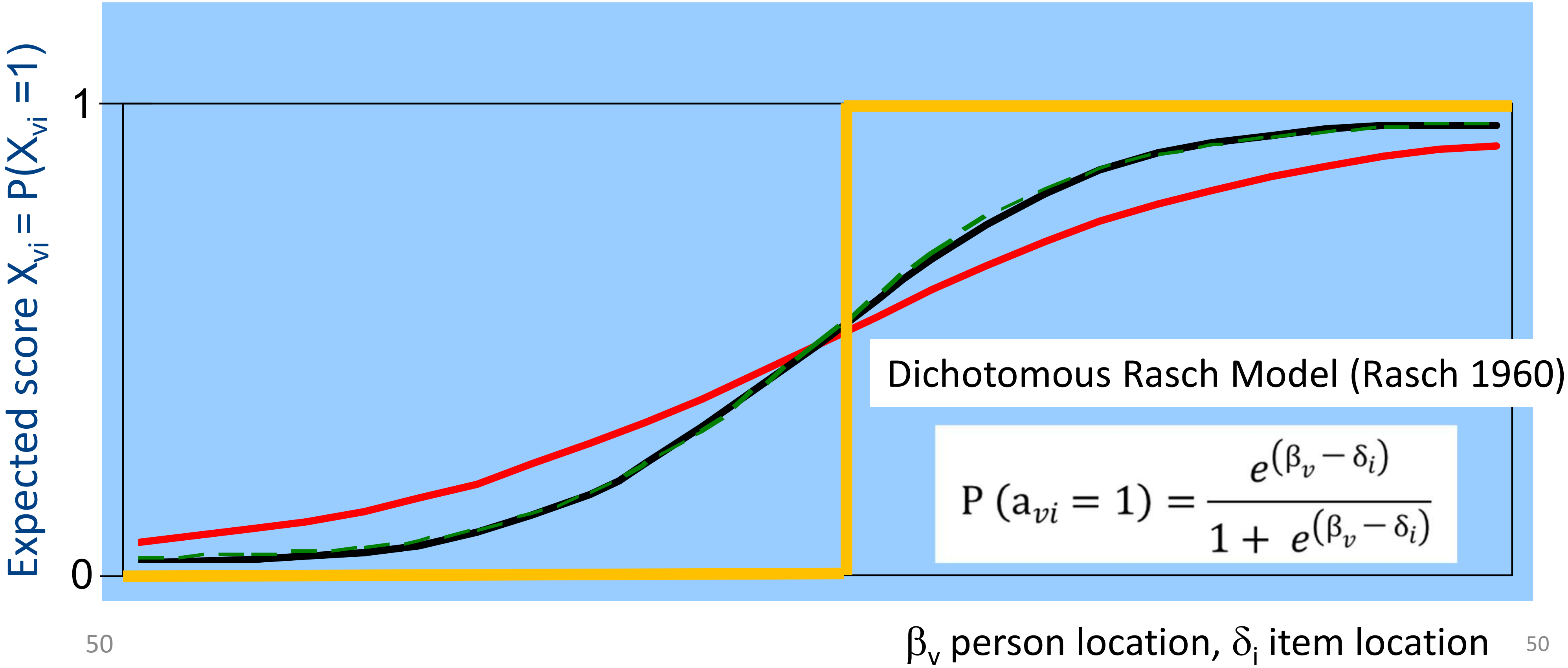
Guttmann + Thurstone -> Rasch

- Guttman does not account for error
- Thurstone‘s paired comparisons disregard person location
- Rasch: estimates item locations and person locations (with parameter separation)

	Item B – 0	**Item B - 1**
Item A - 0	40	***30***
Item A - 1	***70***	60

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



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

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 ; *<u>but different person fit!</u>*)

item 1 item 2 item 3 item 4

less —————————————————————— 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, <u>Y Y N N = 0.19</u>, 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

- 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

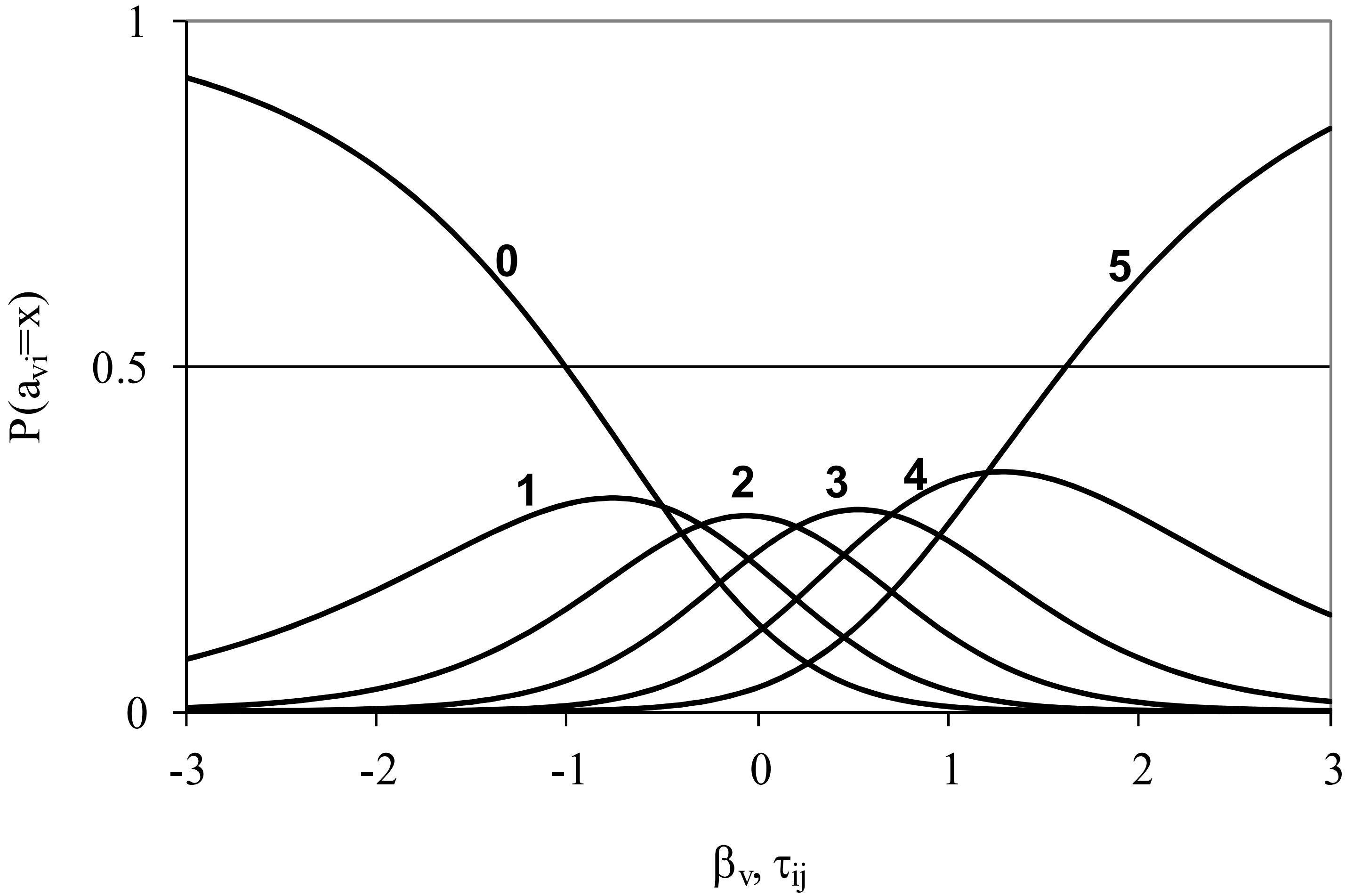
Polytomous Rasch model

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

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

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

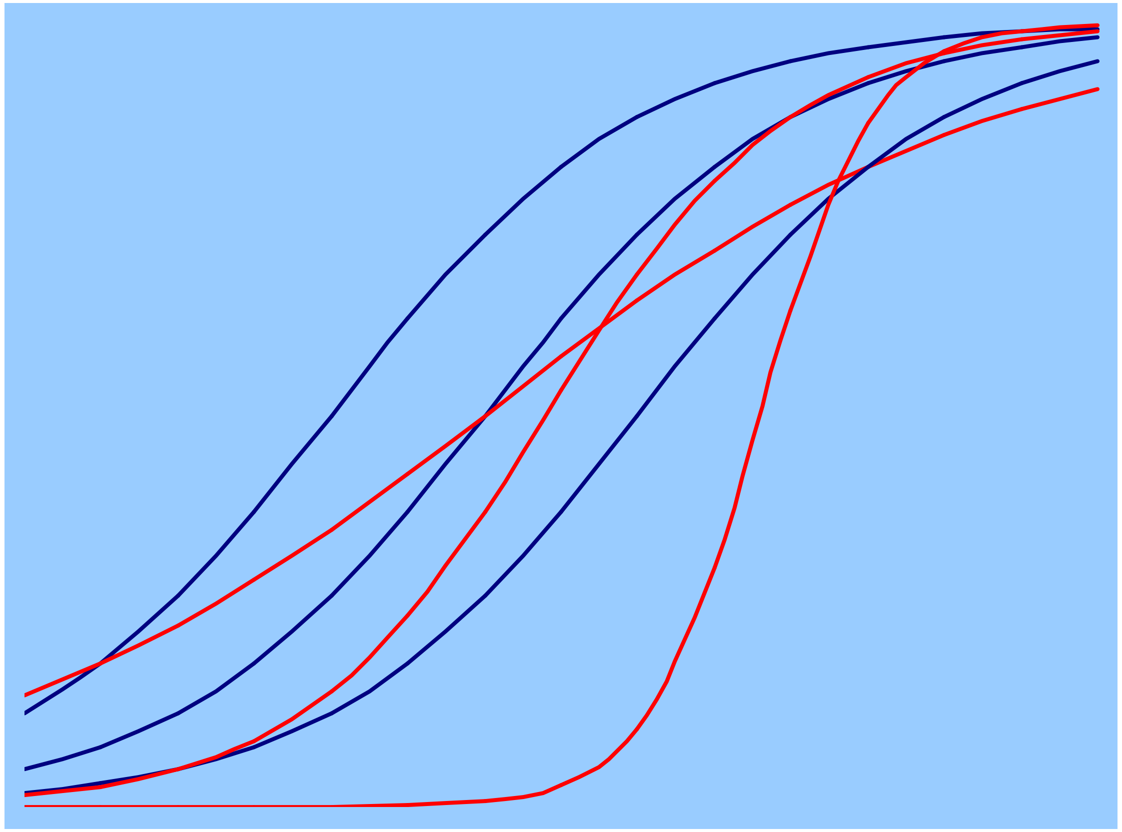
Polytomous Rasch Model



55

Different IRT-Models

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

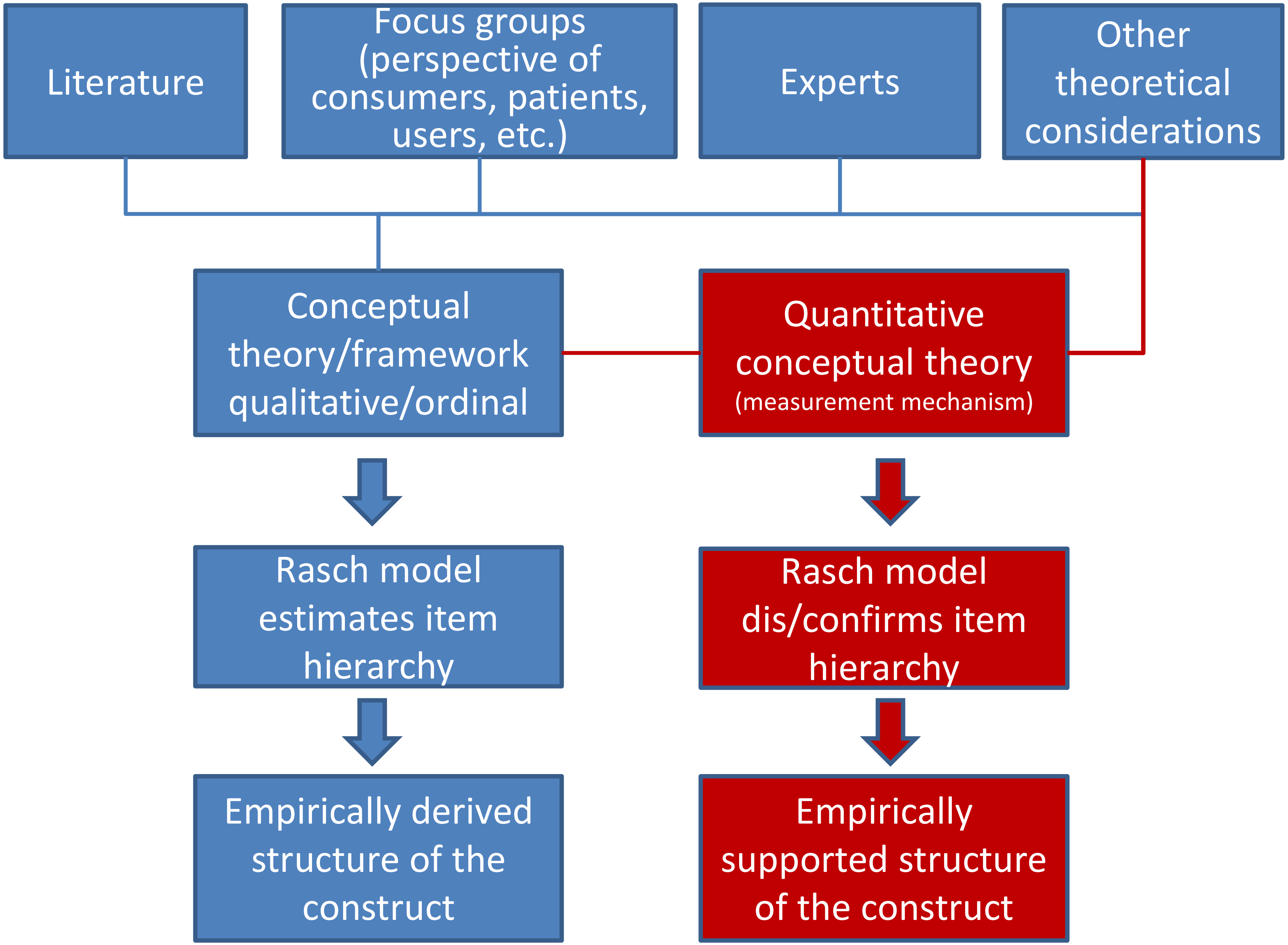


$$P(a_{vi} = 1) = \frac{e^{\beta_v - \delta_i}}{1 + e^{\beta_v - \delta_i}}$$

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

56

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]

Example: Lexile Framework

Lexile framework

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

Example: Measurement of perceived risk

- Scale to measure perceived risks when using various tobacco and nicotine-containing products (such as combustible cigarettes, e-cigarettes, smokeless tobacco, etc.)
- Multiple dimensions (health risk, addiction risk, social risk, practical risk)
- Comparability
  - Across user groups
  - Across products
  - Across nationalities

Example: Measurement of perceived health risk

- 18-item long version
- 9-item short version

Item location	Health issue	Risk continuum
-0.49	having lung cancer	Frequently endorsed
-0.29	an earlier death	→ Higher risk
-0.09	aging faster	
-0.01	having mouth or throat cancer	
-0.01	having a cough early in the morning	
0.04	having other types of cancer	
0.15	losing some sense of taste	
0.26	being physically unfit	→ Lower risk
0.53	having sores of the mouth or throat	Less frequently endorsed

Perceived risk

- Various risk definitions (depending on the field of enquiry) all of which include the same two components:
  - outcome (typically negative) of some intensity/magnitude
  - has a probability of occurrence

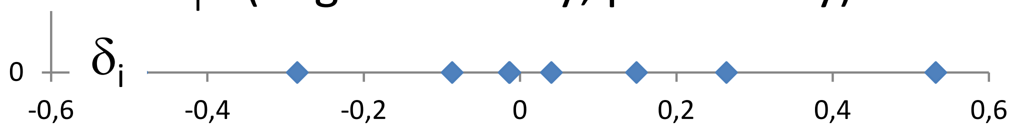
(negative) utility of a health issue

probability of a health issue (because of smoking CC)

TNP-independent

TNP-dependent

δ_i=f(negative utility; probability)



Risk as a function of

- the sum of probability and negative utility (additive model; Kaplan and Garrick, 1981)
- or probability times negative utility (multiplicative model; Daniel et al., 2009)

Kaplan, S. and B.J. Garrick, *On the quantitative definition of risk.* Risk analysis, 1981. **1**(1): p. 11-27.
Daniel, S., Daniel, P., David, M., & Mark, T. (2009). Probability and Consequence. *Mechanical Engineering*, *131*(03), 38-41.

Measurement of *probability*
(smoking CC)

Probability

*This section asks about your views on **cigarette smoking** and health problems. Please imagine a cigarette smoker, who smokes one pack of cigarettes a day, starting at age 18.*

Please select the answer that best reflects your view, keeping in mind that there are no right or wrong answers. Please answer all questions (emphasis added)

*What do you think are **the chances that this cigarette smoker gets the following health problems sometime during their lifetime…?***

<short form items>

Response scale

1. No chance, almost no chance (less than 1 in 10 chance)
2. Very slight possibility (1 in 10 chance)
3. Slight possibility (2 in 10 chance)
4. Some possibility (3 in 10 chance)
5. Fair possibility (4 in 10 chance)
6. Fairly good possibility (5 in 10 chance)
7. Good possibility (6 in 10 chance)
8. Probable (7 in 10 chance)
9. Very probable (8 in 10 chance)
10. Almost sure (9 in 10 chance)
11. Certain, practically certain (more than 9 in 10 chance)

Juster, F. T. (1966). Consumer buying intentions and purchase probability: An experiment in survey design. *Journal of the American Statistical Association*, *61*(315), 658-696.

Measurement of *negative utility*

Negative utility

This section asks you *how bad it would be to get various health problems during your lifetime*. It **does not matter how or why you would get the health problem**. All of the health problems listed are certainly bad to get. Please select the answer that best reflects your view, keeping in mind that there are no right or wrong answers. Please answer all questions.

(emphasis added)

How bad would it be if you got each of the following, sometime during your lifetime …

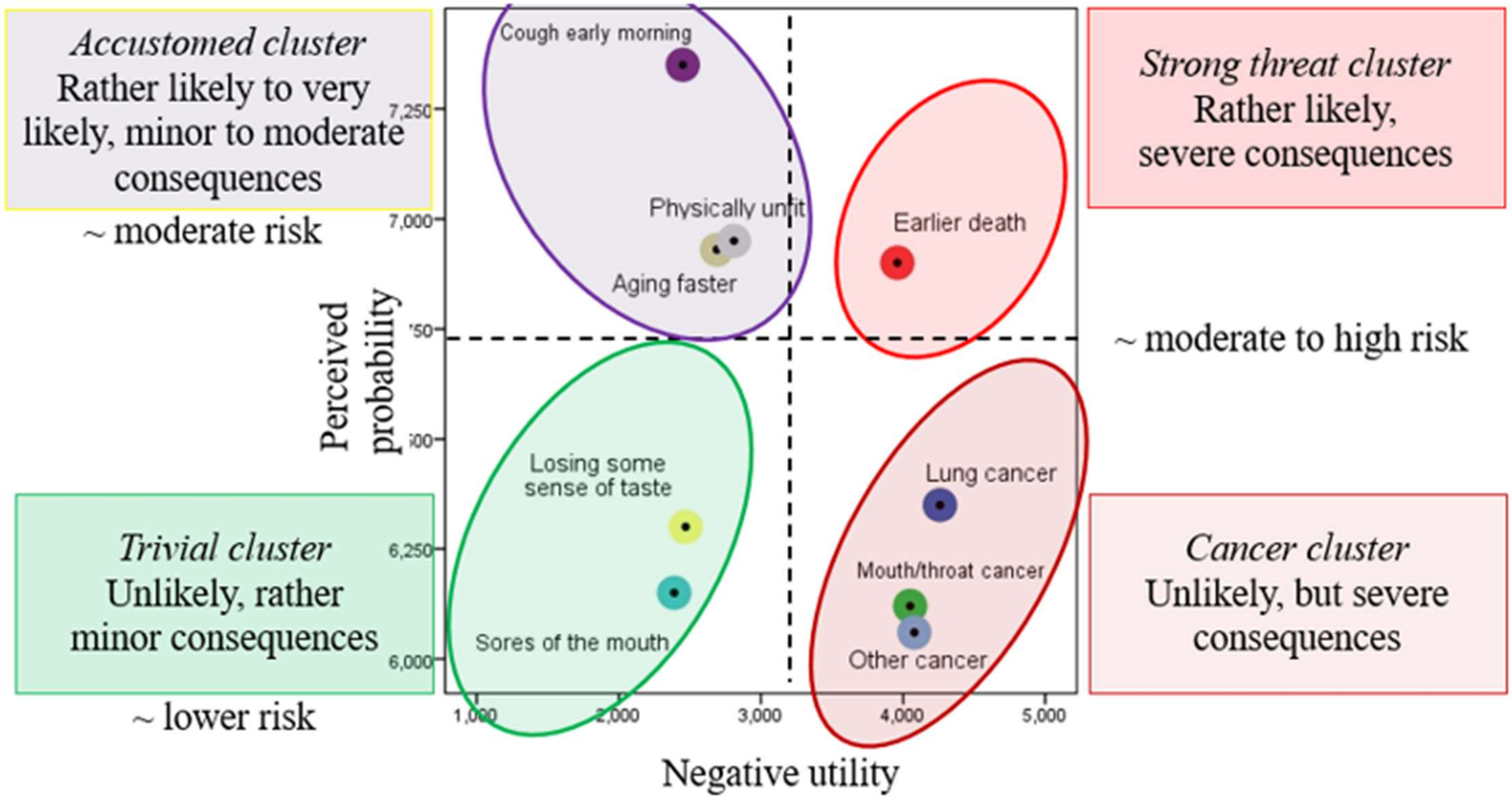
<short form items>

Response scale

1. Slightly bad
2. Moderately bad
3. Very bad
4. Extremely bad
5. The worst imaginable

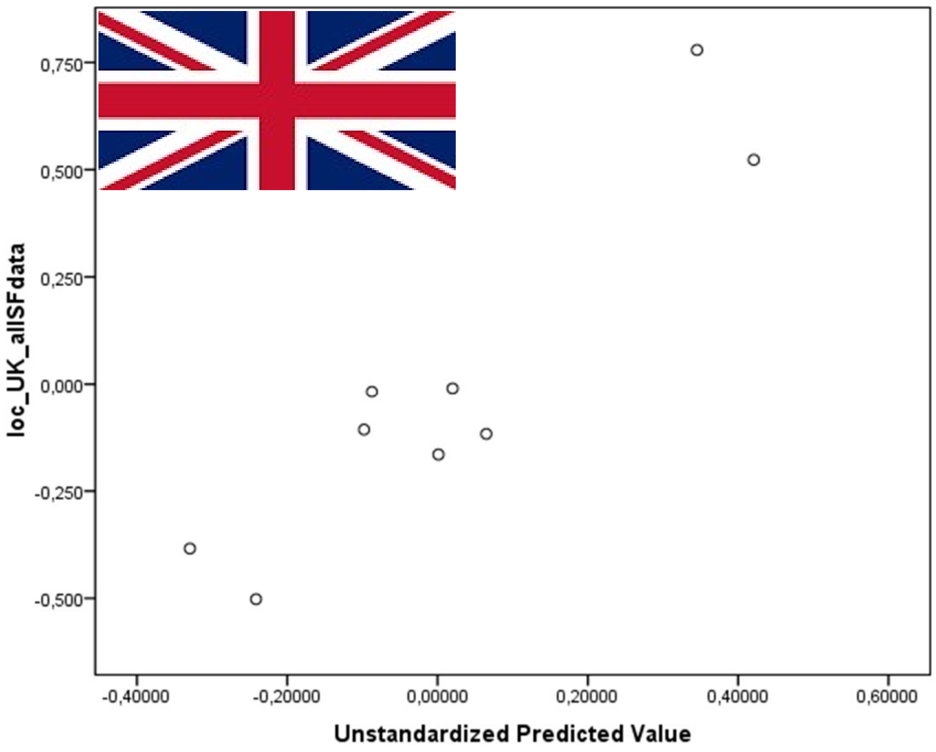
- Scale development and validation study
  - US
  - Comparable measurement in Italy and Japan
  - One common set of item locations
- Subsequent study
  - UK
  - Slight deviations but original item locations remain broadly valid
- *Original item locations US scale development*
- *UK-specific item locations*

Probability and negative utility

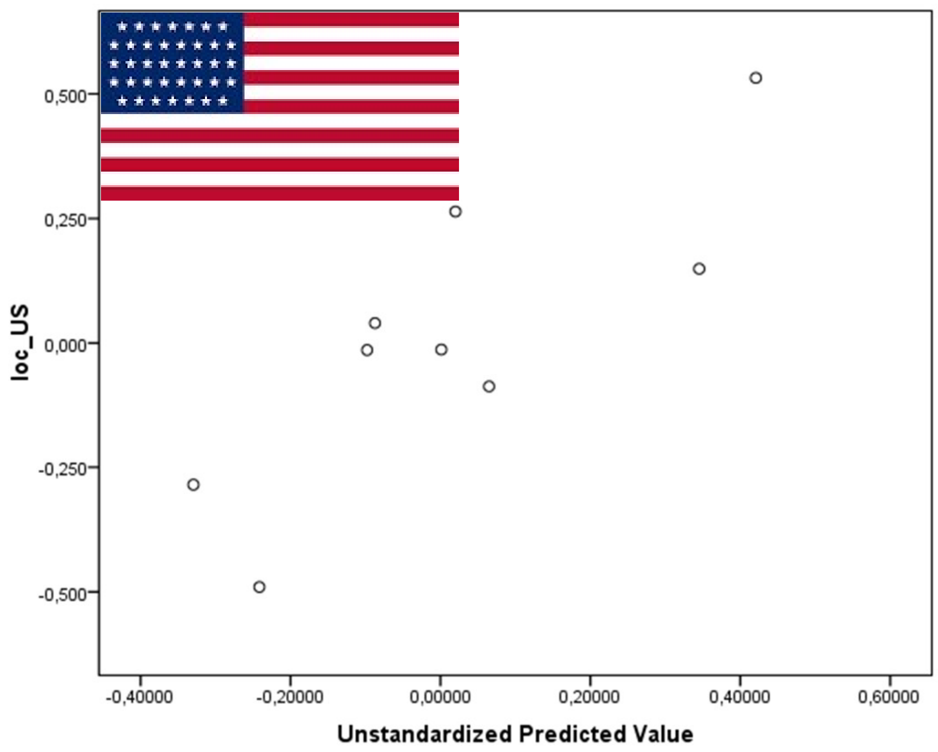


Predicted item locations	Model (predictors)	Regression analysis r	r^2	p (F)
US scale development calibration	[1] Probability × Negative utility (multiplicative model)	0.79	0.62	0.01
	[2] Probability + Negative utility (additive model)	**0.83**	**0.69**	**0.03**
UK calibration all available data	[3] Probability × Negative utility (multiplicative model)	0.81	0.65	0.01
	[4] Probability + Negative utility (additive model)	**0.93**	**0.87**	**<0.01**

Predicted item locations	Predictor	Regression coefficient b	Standardized regression coefficient β	Collinearity (tolerance)	ANOVA
US scale development calibration	Probability (rescaled 0-to-100)	−3.67	−0.53	0.85	$F(2,6)=6.58$ $p=0.031$ $r^2=0.69$
	Negative utility	−0.32	−0.88	0.85	
	Intercept	2.90			
UK calibration all available data	Probability (rescaled 0-to-100)	−7.05	−0.75	0.85	$F(2,6)=20.41$ $p=0.002$ $r^2=0.87$
	Negative utility	−0.48	−0.92	0.85	
	Intercept	5.33			



0,750
0,500
0,250
0,000
-0,250
-0,500
loc_UK_allSFdata
-0,40000
-0,20000
0,00000
0,20000
0,40000
0,60000
Unstandardized Predicted Value

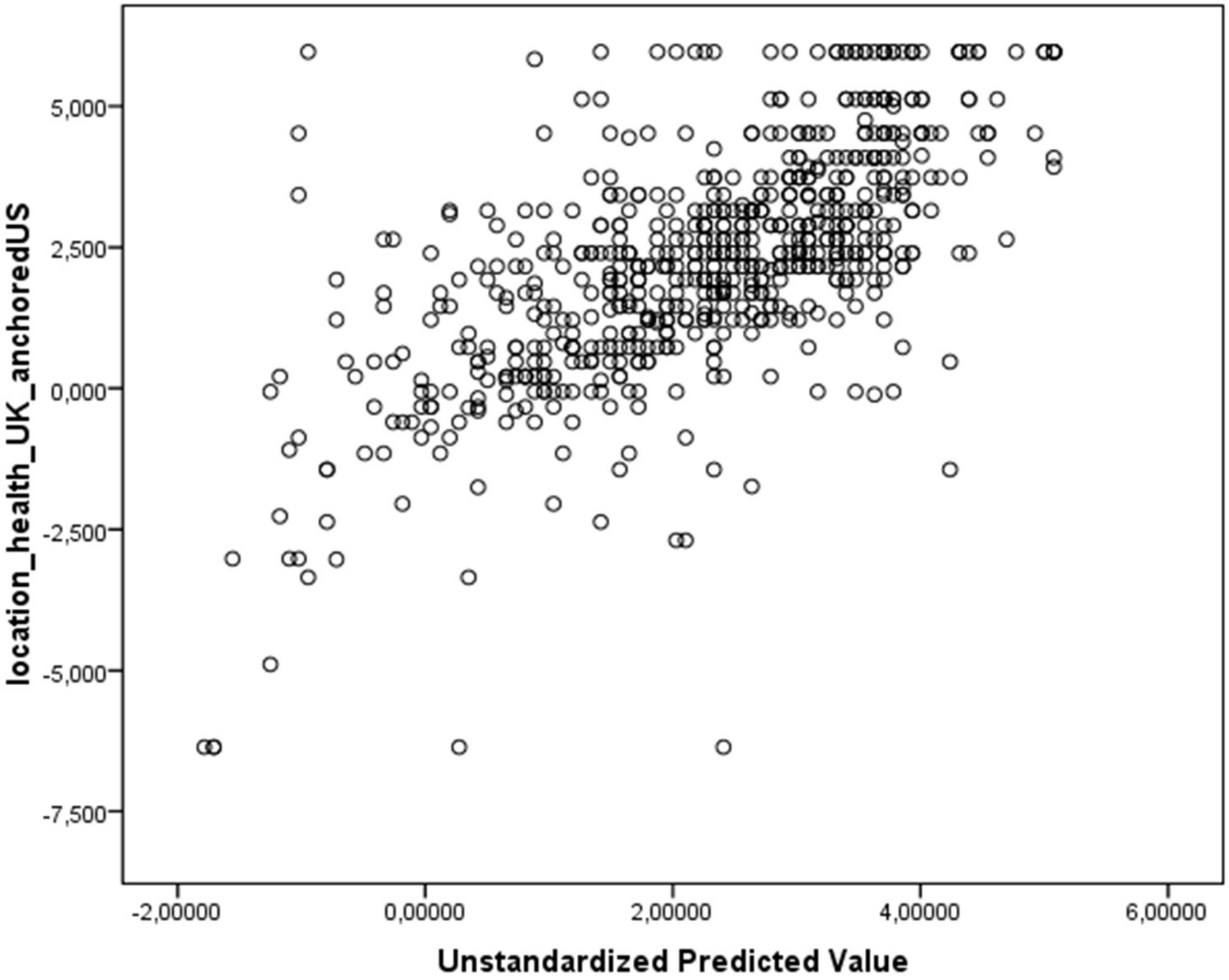


0,500
0,250
0,000
-0,250
-0,500
loc_US
-0,40000
-0,20000
0,00000
0,20000
0,40000
0,60000
Unstandardized Predicted Value

Individual (person) measurements

Predicted participant locations	Model (predictors)	Regression analysis r	r^2	p (F)
Exclusive use of cigarettes	[1] Probability × Negative utility (multiplicative model)	0.65	0.43	<0.01
	[3] Probability + Negative utility (additive model)	**0.70**	**0.49**	**<0.01**

Predicted participant locations	Predictor	Regression coefficient b	Standardized regression coefficient β	Collinearity (tolerance)	ANOVA
Exclusive use of cigarettes	Probability	7.06	0.63	0.78	F(2,404)=197.48 p<0.01 r^2=0.49
	Negative utility	0.36	0.14	0.78	
	Intercept	−2.81			



location_health_UK_anchoredUS
5,000
2,500
0,000
-2,500
-5,000
-7,500
-2,00000
0,00000
2,00000
4,00000
6,00000
Unstandardized Predicted Value

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.