

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques.
dev.

item development

domain
identification

item generation

scale development

cognitive
interviews

sampling

scale evaluation
reliability

statistical
models

analysis plans

exploratory models

common factor
models

item response
models

latent class analysis

Literature

Basics (and advanced stuff) of questionnaire development and analysis

Julius Fenn¹

¹University of Freiburg

PhD student at the Institute of Experimental Psychology Freiburg (Cognition, Action, and Sustainability)

20.01.2022

Incredible what is possible with latent variable models

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation

reliability

statistical models

models

analysis plans

exploratory models

common factor models

item response models

latent class analysis

Literature

Combining research goal, design, and analysis

		Research Goal		
		Description	Prediction	Causation
Research Design	Cross-sectional	<i>E.g. proportion or correlation, PCA</i>		
	Longitudinal	<i>E.g., latent growth curve modeling</i>	<i>E.g., forecasting, cross-validation</i>	
	Experimental	<i>E.g., differences between existing groups</i>	<i>E.g., including a covariate</i>	<i>E.g., comparing means across conditions</i>

Adapted from Hamaker, Mulder and Van IJzendoorn (under review, DCN).

Educational research and experimental psychology: mutual research interests

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation

reliability

statistical models

analysis plans

exploratory models

common factor models

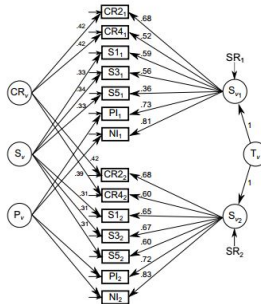
item response models

latent class analysis

Literature

Which cognitive processes underlie differences in cognitive abilities?

Prof. Anna-Lena Schubert (Heidelberg) and others: estimating person parameters by means of cognitive models (e.g. drift-diffusion models) and subsequently applying latent variable models (e.g. latent state-trait models):



Schubert et al. (2016); Frischkorn und Schubert (2018)

Data generation process

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation

reliability

statistical models

analysis plans

exploratory models

common factor models

item response models

latent class analysis

Literature

Data generation process: „When [...] data set has been created, a good model has to be constructed for the data generation process (DGP). This is the stage at which the actual [...] analysis begins... When an [...] model has been constructed for the DGP, it should **only be used for the analysis if it reflects the ongoing in the system of interest properly**“ – Lütkepohl und Krätzig (2004), pp. 1-7

→ a model should make a parsimonious, plausible and substantially meaningful contribution to the data generation process – Hoyle (2012), p. 126

→ in statistics, we assume that we can represent the DGP by discrete and continuous distributions

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation

reliability

statistical models

analysis plans

exploratory models

common factor models

item response models

latent class analysis

Literature

a random variable is defined by assigning numbers to an event of a random process (realisation), for this purpose **distributions** are necessary:

- Random processes can only be described by functional relations $f(X)$, types of representation:

- for discrete random variables \rightarrow probability function
- for continuous random variables \rightarrow density function

\rightarrow distributions are determined by functions and their parameters

A random variable $f(X)$ is a function that assigns a real number to each possible outcome of a random experiment / process.

further information see: [Wikipedia random variables](#); Fahrmeir et al. (2016); Zucchini et al. (2009)

Distributions II: example

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive

interviews

sampling

scale evaluation

reliability

statistical models

analysis plans

exploratory models

common factor models

item response

models

latent class analysis

Literature

Random variables can only take on finitely many values or countably infinitely many values

→ each possible event can be assigned a natural number

Example: rolling 6 or 30 fair dices several times at once and we want to know $A = \{\text{result is 1}\} = \{1\} \rightarrow P(\{1\}) = 1/6$ (binomial distributed):

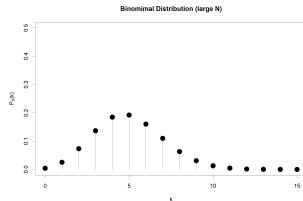
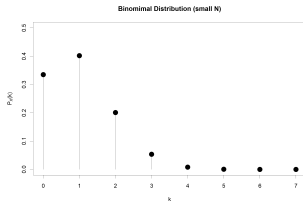


Abbildung: left rolling 6 cubes, right 30 cubes

Is rolling a dice a Gauss distribution?

Distributions III: example 2

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation

reliability

statistical models

analysis plans

exploratory models

common factor models

item response models

latent class analysis


Literature

single items can be binomial distributed (! however the real distribution is never known)

Imagine that participants should answer the following two items on a 6 point Likert Scale:

- Should something be done about climate change for future generations?
- Would you be willing to accept personal restrictions for the fight against climate change?

Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
1	2	3	4	5	6



→ What would you expect?

Take-Home-Message

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation

reliability

statistical models

analysis plans

exploratory models

common factor models

item response models

latent class analysis

Literature

- individual items of a Likert scale result from a binomially distributed process ($Y_1 \sim \text{Bin}(n, p), \dots$)
 - we want to achieve a high variance in the items (all response options are used)
 - variance or covariance (if multiple items) is the most important building block for the statistical models
 - the answering of items should be independent from other people or the ordering of the items
- our statistical models should reflect the ongoing in the system of interest properly / sufficiently and should be theoretically derived

Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful

George Box (1987)

(superficial) model: Cognitive Aspects of Survey Methodology

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation

reliability

statistical models

analysis plans

exploratory models

common factor models

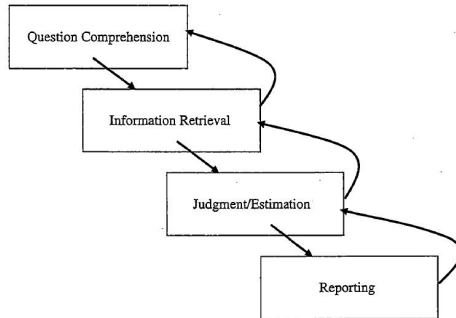
item response models

latent class analysis

Literature

also called „Optimizing-Satisficing-Model“

the CASM model tries to explain how people finally arrive at a response:



→ great template for identifying possible sources of error

Tourangeau und Bradburn (2010); Moosbrugger und Kelava (2012), pp. 57-59

test model: linking observable variables to unobservable variables

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation

reliability

statistical models

analysis plans

exploratory models

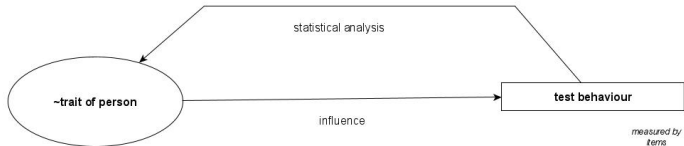
common factor models

item response models

latent class analysis

Literature

whatever the exact processes of question answering are (black box), fundamentally we assume that the answer / reporting depends on a **score on a latent variable**:



→ the central task of test theory is to determine the relationship between test behaviour and the (psychological) characteristic to be assessed

Hambleton und R. Jones (1993); Rost (2004)

What is a Latent Variable?: depiction

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques.
dev.

item development

domain
identification

item generation

scale development

cognitive
interviews

sampling

scale evaluation
reliability

statistical
models

analysis plans
exploratory models

common factor
models

item response
models

latent class analysis

Literature

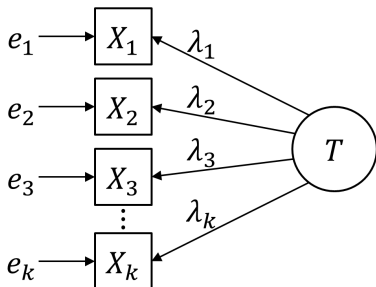


Abbildung: congeneric measurement model

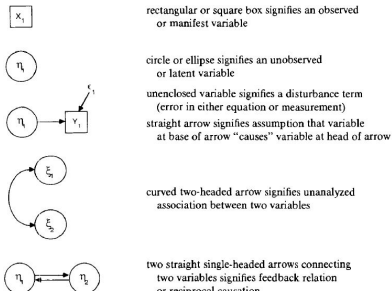


Abbildung: primary symbols used in path analysis

other possible modes of representation: Matrix notation, system of equations

Bollen (1989); Hoyle (2012)

What is a Latent Variable?: technical aspects

technical aspects:

- definition: *Latent variables are random variables whose realized values are hidden*
 → models which are dealing with latent variables are called **Latent Variable Models** (this kind of models can be used for descriptive analysis, predictive, inferential or causal questions, [What is the question?](#))

An possible operationalisation of a latent variable is a linear measurement model with the equation $Y = \lambda * \eta + \epsilon$.

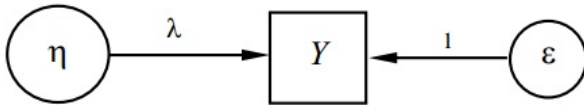


Abbildung: simple linear measurement model – Macho (2016), p. 92

this corresponds to the fundamental equation of Classical Test Theory:
 $Y = T + \epsilon$

What is a Latent Variable?: theoretical aspects

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques.

dev.

item development

domain

identification

item generation

scale development

cognitive

interviews

sampling

scale evaluation

reliability

statistical

models

analysis plans

exploratory models

common factor

models

item response

models

latent class analysis

Literature

great dissertation: Schurig (2017): „Latente Variablenmodelle in der empirischen Bildungsforschung - Die Schärfe und Struktur der Schatten an der Wand“

theoretical aspects:

- the theoretical status of latent variables has not been clarified
 - regarded this variables as representations of real entities or as useful inventions?
 - Advantage: use of latent variables allow more generalisable reasoning than manifest variables
- latent variables need a substantive scientific foundation, whereby the bridging problem between observed and latent variable must be solved by theoretical assumptions and statistical modelling

→ even more surprising the statistical models „get rid“ of the latent variable by assuming local independence: $Pr(y_j | \eta_j) = \prod_{i=1}^n Pr(y_{ij} | \eta_j)$

Schurig (2017); Skrondal und Rabe-Hesketh (2007)

Take-Home-Message

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation

reliability

statistical models

analysis plans

exploratory models

common factor models

item response models

latent class analysis

Literature

- we assume that some kind of undefined latent variable „causes“ the test behaviour
 - since the exact mechanisms are unclear, I would never speak of causality
- each measurement consists of a true value (latent variable) and an error (unique variance)
- in the context of questionnaire development, the central aim is to reproduce the correlation structure between items with a statistical model

What is measurement?

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques.
dev.

item development

domain
identification

item generation

scale development

cognitive
interviews

sampling

scale evaluation
reliability

statistical
models

analysis plans

exploratory models

common factor
models

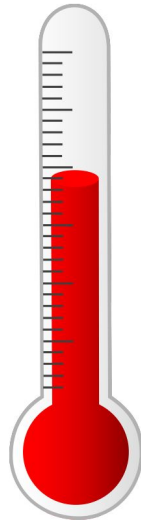
item response
models

latent class analysis

Literature

How do you measure
temperature?

- ▶ By looking at a thermometer
- ▶ For this to make sense, we need to assume that:
 - ▶ Temperature **causes** the level shown in a thermometer
 - ▶ The thermometer features relatively little **measurement error**



slide from presentation: [master psychology courses Structural Equation Modeling](#); in more detail see Markus und Borsboom (2013), p. 19–43

What is measurement? II

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques.
dev.

item development

domain
identification

item generation

scale development

cognitive
interviews

sampling

scale evaluation

reliability

statistical
models

analysis plans

exploratory models

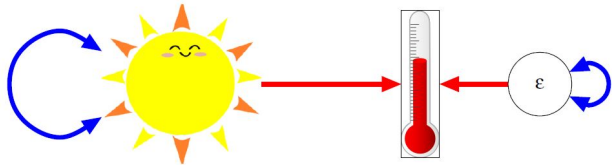
common factor
models

item response
models

latent class analysis

Literature

We can summarize a causal hypothesis in a **path diagram**:



- ▶ Circular nodes (or suns): latent (unobserved) variables
- ▶ Square nodes: observed variables
 - ▶ Also termed *indicators* of a latent variable
- ▶ **Unidirectional links: causal effects**
- ▶ **Bidirectional links: (co)variances**

our assumptions lead to a path diagram

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation

reliability

statistical

models

analysis plans

exploratory models

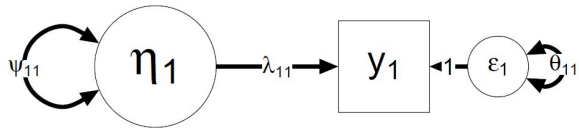
common factor models

item response models

latent class analysis

Literature

Assume all observed and latent variables are **normally distributed** and all causal effects are **linear**. Without loss of information, we can center data and assume all means are 0 (we don't use means until week 3). Now, the path diagram encodes a **causal equation**:



$$y_{i1} = \lambda_{11}\eta_{i1} + \varepsilon_{i1}$$

$$\eta_1 \sim N(0, \sqrt{\psi_{11}})$$

$$\varepsilon_1 \sim N(0, \sqrt{\theta_{11}})$$

λ_{11} is called a **factor loading**, ε_{i1} the **residual variance** and ψ_{11} the **factor variance**.

→ remark: reliability is defined as: $\frac{\lambda_{11}^2 \psi_{11}}{\lambda_{11}^2 \psi_{11} + \theta_{11}}$

Variance splitting

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation reliability

statistical models

analysis plans

exploratory models

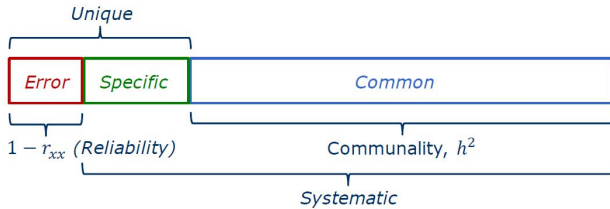
common factor models

item response models

latent class analysis

Literature

Within the framework of factor-analytical models (common factor model), the variance of indicators is divided into common and unique variance:



The variance shared between the indicators is the commonality; the remaining variance is the unique variance, which is divided into indicator-specific method variance (specific) and measurement error variance (error).

Kline (2015), p. 190

Take-Home-Message

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation reliability

statistical models

analysis plans

exploratory models

common factor models

item response models

latent class analysis

Literature

- each measurement consists of a true value (latent variable) and an error (unique variance), which is divided in error and specific variance
 - the specific variance can be taken into account via so-called *multitrait-multimethod models*
 - the acceptable size of the measurement error depends on the consequences of a wrong decision (no statement is possible at the individual level! - Schmidt-Atzert und Amelang (2018), pp. 36-62)
- in the context of questionnaire development, the central aim is to reproduce the correlation structure between items with a statistical model → precondition of communality: high (but not too high) linear correlations between the items

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation

reliability

statistical models

analysis plans

exploratory models

common factor models

item response models

latent class analysis

Literature

concrete measurement, than test administration should be carried out taking into account three central quality criteria of tests, which build on each other (no reliable measurement possible without objectivity, etc.):

objectivity → reliability → validity

Recommendation for reading: Rost (2004); Schmidt-Atzert und Amelang (2018); Moosbrugger und Kelava (2012); Moosbrugger und Kelava (2020)

great article: Glaesmer et al. (2015): „Kriterienkatalog zur Beurteilung psychodiagnostischer Selbstbeurteilungsinstrumente–Empfehlung des Deutschen Kollegiums für Psychosomatische Medizin (DKPM)“

overview of test theoretical terms in German: [TU Dresden wissenschaftliches Arbeiten](#)

Quality criterion: objectivity

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation reliability

statistical models

analysis plans

exploratory models

common factor models

item response models

latent class analysis

Literature

- **objectivity:** Test score is objective if it is independent of any influences outside the tested person (e.g. situational conditions, experimenter - all exogenous variables¹)
 - Implementation objectivity (*Durchführungsobjektivität*): standardisation of implementation conditions (writing a test manual, training test leader, standardisation of all other conditions).
 - Objectivity of evaluation (*Auswertungsobjektivität*): the interpretation of the test result is not dependent on the person who evaluates the test (measurable by interrater reliability, such as Kendall's coefficient of concordance).
 - Objectivity of interpretation (*Interpretationsobjektivität*): different test users come to the same conclusions with identical test scores.

¹variables whose covariance structure is not explained by the statistical model, this includes error terms, unobserved influencing variables and exogenous latent constructs (Macho 2016, pp. 62-63)

Quality criterion: reliability

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation

reliability

statistical models

analysis plans

exploratory models

common factor models

item response models

latent class analysis

Literature

- **reliability:** the extent to which a test measures what it measures.

The focus here is on measurement accuracy. Reliability is shown theoretically by the fact that repeated measurements under the same conditions produce the same measurement results (the central contribution to the development of reliability measurement is made by the classical test theory, which establishes a theory of measurement error²).

- Reliability can be estimated by different methods, often internal consistency - Cronbach's Alpha (measure of how items in a scale correlate with one another) is estimated

²The classical test theory assumes that the test performance of a person on the question i is composed of $x_i = \tau_i + \epsilon_i$. Here τ_i corresponds to the person's true score on question i , which is composed of an item response x_i and the error ϵ_i , where the error is unbiased (Moosbrugger und Kelava 2012, pp. 104-105)

Quality criterion: validity

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques.
dev.

item development

domain
identification

item generation

scale development

cognitive
interviews

sampling

scale evaluation

reliability

statistical
models

analysis plans

exploratory models

common factor
models

item response
models

latent class analysis

Literature

- **validity:** A test is considered valid if it actually measures the characteristic it is supposed to measure and not some other characteristic. The measurement of validity is done in two steps:
 - (1) via structure-searching (such as exploratory factor analysis) and structure-testing (such as confirmatory factor analysis) procedures, **construct validity** is determined (see in detail e.g. Messick 1994). This indicates the extent to which conclusions can be drawn from test results for example about psychological personality traits.
 - (2) the agreement of the results of the individual constructs should be high with constructs that measure the same or similar characteristics (convergent validity) and the agreement with results from constructs that measure other characteristics should be low (discriminant validity). This can be analysed via correlations

validation: an argument of accumulating evidence

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation

reliability

statistical models

analysis plans

exploratory models

common factor models

item response models

latent class analysis

Literature

argument-based approach to validation: To validate an interpretation or use of measurements is to evaluate the rationale, or argument, for the proposed conclusions and decisions ... Ultimately, the need for validation derives from the scientific and social requirement that public claims and decisions be justified

- interpretive argument: specifies the proposed interpretations and uses of assessment results by laying out a network of inferences and assumptions leading from the observed performances to the conclusions and decisions based on the assessment scores
- validity argument: provides an evaluation of the interpretive argument's coherence and the plausibility of its inferences and assumptions

Kane (2012); Markus und Borsboom (2013)

further quality criteria of scales / single indicators I

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques.
dev.

item development

domain
identification

item generation

scale development

cognitive

interviews

sampling

scale evaluation

reliability

statistical
models

analysis plans

exploratory models

common factor
models

item response
models

latent class analysis

Literature

A variety of further quality criteria of indicators were developed by the „Key National Indicators Initiative“ (2005):

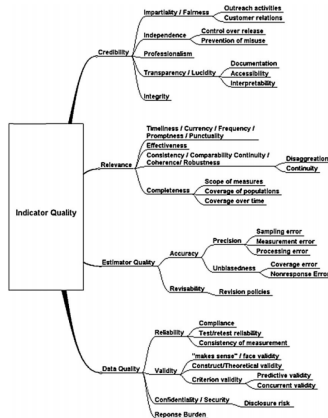


Abbildung: Quality criteria of indicators according to the „Key National Indicators Initiative“; Source: Groves und Lyberg 2010, 856

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation

reliability

statistical models

analysis plans

exploratory models

common factor models

item response models

latent class analysis

Literature

→ „Key National Indicators Initiative“ borrowing from the concept of total survey error (*total survey error*), which provides a conceptual framework to describe possible sources of survey error (see Groves und Lyberg [2010](#); for introduction see Faulbaum [2014](#)). Further issues of indicator quality (keyword test economy, usefulness, reasonableness, standards,..) in overview books on test theories (e.g. Moosbrugger und Kelava [2012](#))

Take-Home-Message

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques.
dev.

item development

domain
identification

item generation

scale development

cognitive
interviews

sampling

scale evaluation
reliability

statistical
models

analysis plans

exploratory models

common factor
models

item response
models

latent class analysis

Literature

it is not trivial to generate items, visible in the **concept of generalisability of tests**: For the scaling of tests there are theoretically (in)finately many plausible statistical models and question possibilities (e.g. G. Fischer und Molenaar [1995](#), pp. 6-7; Robitzsch et al. [2011](#); Shadish, Cook und Campbell [2002](#)):

$$Var(\hat{d})_{\text{Total}} = Var(\hat{d})_{\text{Persons}} + Var(\hat{d})_{\text{Items}} + Var(\hat{d})_{\text{Models}} + Var(\hat{d})_{\text{Occasions}}$$

Sources of variability of results:

- persons
- selected items (there is a potential universe of possible items for the query of individual knowledge areas)
- statistical models
- time point of measurement

see: [Generalizability Theory: Overview](#)

guidelines questionnaire development

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation

reliability

statistical models

analysis plans

exploratory models

common factor models

item response models

latent class analysis

Literature

The following considerations are based on guidelines for questionnaire development, especially Rattray und M. C. Jones [2005](#); Artino et al. [2014](#); **Boateng et al. 2018**

There are three phases to creating a rigorous scale-item development, scale development, and scale evaluation:

- item development
- scale development
- scale evaluation

item development

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation

reliability

statistical models

analysis plans

exploratory models

common factor models

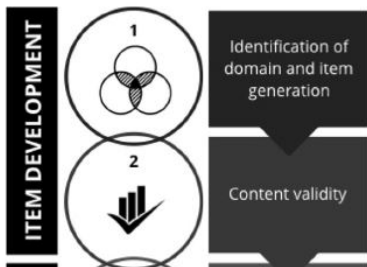
item response models

latent class analysis

Literature

aims:

- To specify the boundaries of the domain and facilitate item generation
- To identify appropriate questions that fit the identified domain
- To evaluate each of the items regarding content relevance, representativeness, and technical quality by experts and target population



Major modes of searching

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques.
dev.

item development

domain
identification

item generation

scale development

cognitive
interviews

sampling

scale evaluation

reliability

statistical
models

analysis plans

exploratory models

common factor
models

item response
models

latent class analysis

Literature

different modes should be used concurrently and used search strategies should be made explicit

Footnote Chasing

Cooper 1985

References in review papers written by others

References in books by others

References in nonreview papers from journals you subscribe to

References in nonreview papers you browsed through at the library

Topical bibliographies compiled by others

Mann 1993

Searches through published bibliographies (including sets of footnotes in relevant subject documents)

Related Records searches

Consultation

Cooper 1985

Communication with people who typically share information with you

Informal conversations at conferences or with students

Formal requests of scholars you knew were active in the field (e.g., solicitation letters)

Comments from readers/reviewers of past work

General requests to government agencies

Mann 1993

Searches through people sources (whether by verbal contact, E-mail, electronic bulletin board, letters, etc.)

Searches in Subject Indexes

Cooper 1985

Computer search of abstract data bases (e.g., *ERIC*, *Psychological Abstracts*)

Manual search of abstract data bases

Mann 1993

Controlled-vocabulary searches in manual or printed sources

Key word searches in manual or printed sources

Computer searches—which can be done by subject heading, classification number, key word. . .

Browsing

Cooper 1985

Browsing through library shelves

Mann 1993

Systematic browsing

Citation Searches

Cooper 1985

Manual search of a citation index

Computer search of a citation index (e.g., *SSCI*)

Mann 1993

Citation searches in printed sources

Computer searches by citation

main concern of literature retrieval

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation

reliability

statistical models

analysis plans

exploratory models

common factor models

item response models

latent class analysis

Literature

return all of the studies relevant to the review ('recall') without retrieving irrelevant studies ('precision')

Recall:

- expresses (as a percentage) the ratio of relevant documents retrieved to all those in a collection that should be retrieved
- fiction: if we could identify all existing relevant documents in order to count them, we could retrieve them all, and so recall would always be 100 percent

Precision:

- expresses (as a percentage) the ratio of documents retrieved and judged relevant to all those actually retrieved
- measures how many irrelevant documents-false positives-one must go through to find the true positives or hits

→ trade-off: high recall leads to degrading precision

Grames et al. (2019), 1646; Cooper, Hedges und Valentine (2009), 56ff.

define single items

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation

reliability

statistical models

analysis plans

exploratory models

common factor models

item response models

latent class analysis

Literature

Possible process of item / indicator development proceeds in three steps:

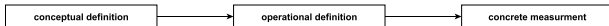


Abbildung: inspired by Planty 2010, pp . 20-38

- **conceptual definition:** describes in the abstract the qualities and concepts that the indicator will purport to describe
- **operational definition:** developing a concrete, operational (or working) definition for the indicator and identifying the precise measures that will be used to represent the concept described in the conceptual definition
- **concrete measurment:** deciding how the operational concept should be measured (definition of a question)

Planty (2010)

item generation: deductive, inductive, external method

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques.
dev.

item development

domain
identification

item generation

scale development

cognitive
interviews

sampling

scale evaluation

reliability

statistical
models

analysis plans

exploratory models

common factor
models

item response
models

latent class analysis

Literature

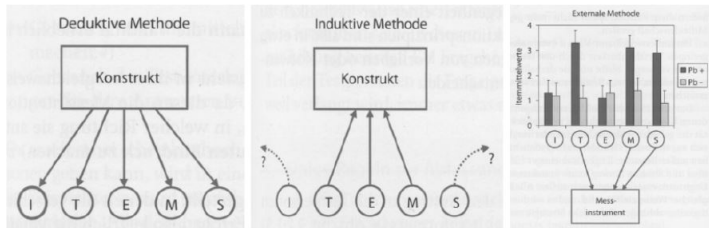


Abbildung: see Schmidt-Atzert und Amelang 2018, p. 111

Schmidt-Atzert und Amelang (2018), pp. 97-112

deductive method: Which items are relevant? CIPO

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques.
dev.

item development

domain
identification

item generation

scale development

cognitive
interviews

sampling

scale evaluation

reliability

statistical
models

analysis plans

exploratory models

common factor
models

item response
models

latent class analysis

Literature

Possible process of indicator development proceeds in three steps:

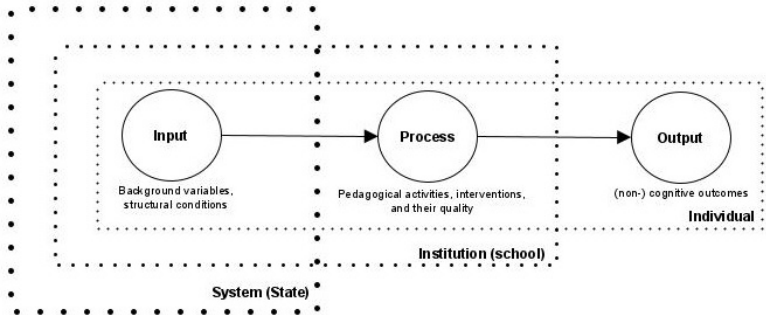


Abbildung: adjusted „CIPO“ model

Adjustment according to: N. Fischer et al. (2011); Klieme und Vieluf (2013); Tippelt und Schmidt-Hertha (2018)

Recommendation for reading: Keller (2014); Ditton (2000); primary source: Stufflebeam (1971)

Which items are relevant? CIPO II [german]

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation

reliability

statistical models

analysis plans

exploratory models

common factor models

item response models

latent class analysis

Literature

Komponenten	Schüler	Schule	Staat
Input	Alter, Geschlecht, Migrationshintergrund; Vorwissen; Einstellungen	soziales Einzugsgebiet; Ressourcen; Schulform	Demographie, gesellschaftliche Debatten; Investitionen
Prozess	Lerntempo, Motivation	organisatorische Gestaltung; Elternkooperation; Lehrerkoooperation	Regulierungsinstrumente, wie Maß der Schulautonomie; Standardsetzung
Output	Lernergebnisse	Anzahl Abbrecher auf Schulebene; durchschnittliche Schulleistung; Schulklima	Bildungsstand, Wirtschaftswachstum, Wohlstand der Gesellschaft

Tabelle: Einzelne Komponenten des „CIPO“-Modells in Bezug zu den Akteursebenen

Variablen entnommen aus: Klieme und Vieluf (2013), 234; Tippelt und Schmidt-Hertha (2018), 383

Why do we need to collect so many items and additional variables? [partly german]

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques.
dev.

item development

domain
identification

item generation

scale development

cognitive
interviews

sampling

scale evaluation
reliability

statistical
models

analysis plans

exploratory models

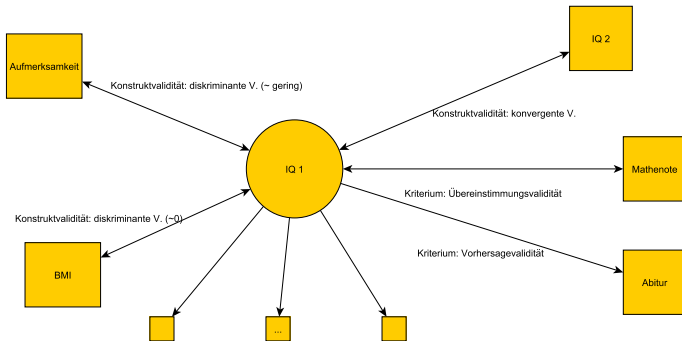
common factor
models

item response
models

latent class analysis

Literature

Definition **construct validity**: indicates the extent to which a test or survey procedure measures a characteristic of interest in a way that is consistent with existing construct definitions and theories



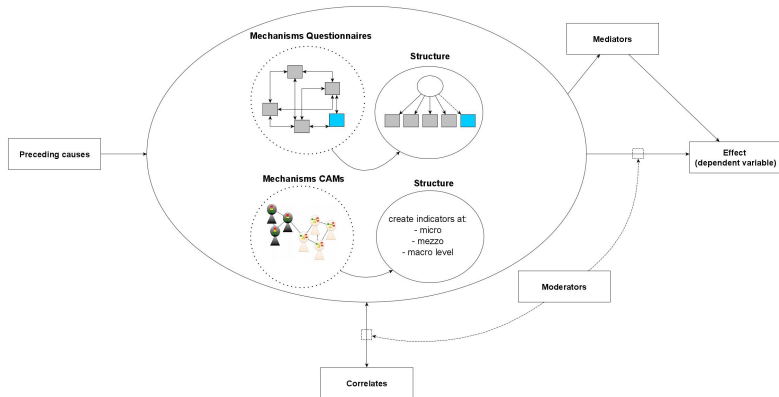
see also: [blog post](#)

e.g. Moosbrugger und Kelava (2020); Cronbach und Meehl (1955)

construct validity: gradually becoming more complex

confronted by a complex system, we need to „complicate ourselves“ and bring together knowledge from different disciplines

Complex System



scale development

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques.
dev.

item development

domain
identification

item generation

scale development

cognitive
interviews

sampling

scale evaluation
reliability

statistical
models

analysis plans

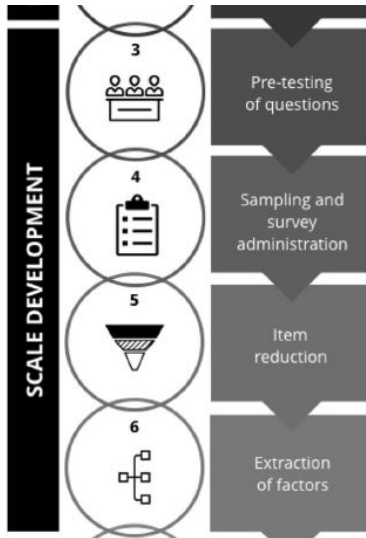
exploratory models

common factor
models

item response
models

latent class analysis

Literature



aims:

- assess the extent to which questions reflect the domain of interest
- collect data with minimum measurement errors and to ensure the availability of sufficient data
- determine the optimal number of factors or domains that fit a set of items

→ Removal of individual items by means of statistical models

Cognitive interviews

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques.
dev.

item development

domain
identification

item generation

scale development

cognitive
interviews

sampling

scale evaluation

reliability

statistical
models

analysis plans

exploratory models

common factor
models

item response
models

latent class analysis

Literature

Cognitive interviews are a qualitative method that examines the question-answer process of the test subjects

→ test the comprehensibility of questions and possible difficulty in answering them. Theoretically, cognitive interviews are based on a heuristic question-answer model (see slide 9). This consists of four stages:

- 1 Understanding the question, this requires the respondent to syntactically relate words and grammar and semantically decode their meaning content
- 2 recalling information requires recalling information from long-term memory to working memory
- 3 evaluation and assessment of a question can be made directly in the case of an already reflected topic or only an impression can be given
- 4 when giving an answer, the assessment made must be reflected in the appropriate answer format of the questionnaire

e.g. Willis (2004); Miller et al. (2014)

data collection: random sample

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation

reliability

statistical models

analysis plans

exploratory models

common factor models

item response models

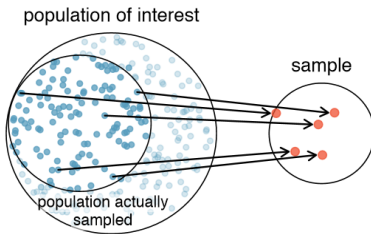
latent class analysis

Literature

- **simple random sample:** random selection of statistical units; each object has the same probability of being included in the sample.

- Advantage: systematic errors are avoided; can be representative

Example of an biased random sample - Causes e.g. *non-response bias*, *incomplete frame population*:



Graphics taken from: [OpenIntro Statistics \(Diez et al 2019\)](#)

data collection: framework

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation

reliability

statistical models

analysis plans

exploratory models

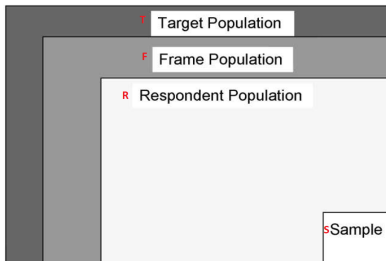
common factor models

item response models

latent class analysis

Literature

Population from which one draws the sample determines the universality of the results and conclusions !



→ graph can be rewritten as $T \subseteq F \subseteq R \subseteq S$

- presence of non-responses (*non-response*): $S \subset R$
- List of statistical objects to be drawn (*frame population*) not fully covering population to be selected (*target population*): $F \subset T$

scale evaluation

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques.
dev.

item development

domain
identification

item generation

scale development

cognitive
interviews

sampling

scale evaluation

reliability

statistical
models

analysis plans

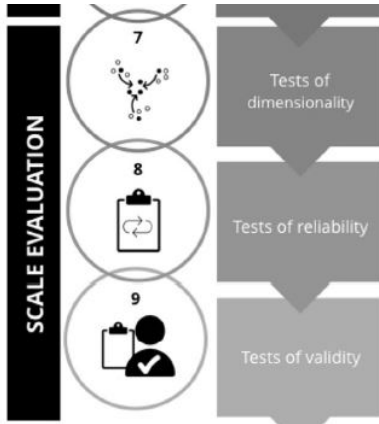
exploratory models

common factor
models

item response
models

latent class analysis

Literature



aims:

- testing if latent constructs are as hypothesized (see construct validity slide 36)
- establishing if responses are consistent / reliable (when repeated)
- **ensuring you measure the latent dimension you intended** (see validation as a process 24)

reliability: recommendation

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation

reliability

statistical models

analysis plans

exploratory models

common factor models

item response models

latent class analysis

Literature

When interpreting Cronbach's alpha, it should be noted that this measure represents a „lower bound to the reliability“, as well as „persistently and incorrectly taken to be a measure of the internal structure of the test, and hence as evidence that the items in the test "measure the same thing." However, alpha does not provide the researcher with this sort of information". (Sijtsma 2009, 107; see also Macho 2016, 87). In the article, Sijtsma (2009) shows that it is necessary to explicitly check the dimensionality of items representing an assumed construct and to test whether a τ -equivalent measurement model exists for the items. The reasons for this are that Cronbach's alpha is neither a measure of internal consistency (correlation of items) nor a measure of unidimensionality (also called homogeneity under the assumption of a general factor) (Sijtsma 2009; Revelle und Zinbarg 2009). In addition, the items should be tested for normal distribution according to Trizano-Hermosilla and Alvarado (2016), since non-normally distributed items lead to a negative bias in Cronbach's alpha regardless of the sample size (Trizano-Hermosilla und Alvarado 2016). Overall, instead of Cronbach's alpha, it is suggested to use other measures such as the „greatest lower bound“ (glb) or McDonald's ω_t (see e.g. Sijtsma 2009; Revelle und Zinbarg 2009). However, the use of the „correct“ reliability measure depends on the data basis (Revelle und Zinbarg 2009; Trizano-Hermosilla und Alvarado 2016).

Latent Variable Models

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation

reliability

statistical models

analysis plans

exploratory models

common factor models

item response models

latent class analysis

Literature

definition **latent variable model**: a statistical model that relates a set of observable variables (so-called manifest variables) to a set of latent variables

Table 1. *Traditional latent variable models*

Observed variable(s)	Latent variable(s)	
	Continuous	Categorical
Continuous	Common factor model Structural equation model Linear mixed model Covariate measurement error model	Latent profile model
Categorical	Latent trait model/IRT	Latent class model

→ we assume that the responses on the indicators or manifest variables are the result of an individual's position on the latent variable(s), and that the manifest variables have nothing in common after controlling for the latent variable (local independence)

Skrondal und Rabe-Hesketh (2007); Skrondal und Rabe-Hesketh (2004)

Latent Variable Models II: local independence

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation

reliability

statistical models

analysis plans

exploratory models

common factor models

item response models

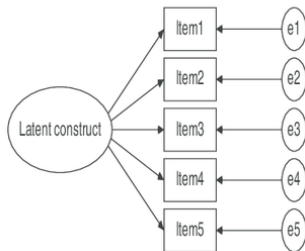
latent class analysis

Literature

local independence: latent variable explains why the observed items are related to one another (see example in [Wikipedia](#))

reflective measurement model:

Reflective measurement model



→ even more surprising the statistical models „get rid“ of the latent variable by assuming local independence: $Pr(y_j | \eta_j) = \prod_{i=1}^n Pr(y_{ij} | \eta_j)$

Urbach, Ahlemann et al. (2010)

Data Science Pipeline

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques.
dev.

item development

domain
identification

item generation

scale development

cognitive
interviews

sampling

scale evaluation

reliability

statistical
models

analysis plans

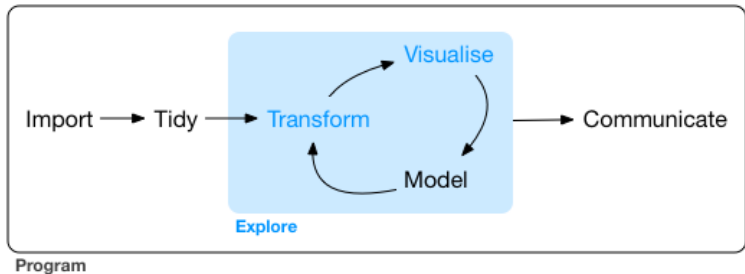
exploratory models

common factor
models

item response
models

latent class analysis

Literature

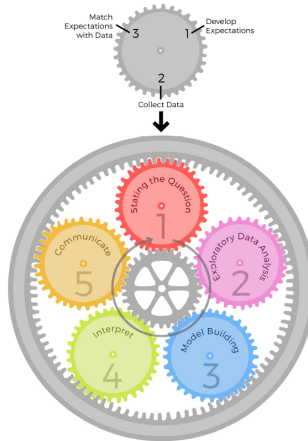


→ collaborate with others: [recommendations for data sharing](#)

R for Data Science

analysis plans

A data analysis plan is „a detailed document outlining procedures for conducting an analysis on data“



Epicycles of Analysis

overview: exploratory models

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain

identification

item generation

scale development

cognitive

interviews

sampling

scale evaluation

reliability

statistical models

analysis plans

exploratory models

common factor models

item response

models

latent class analysis

Literature

compared to structure-testing (such as confirmatory factor analysis, CFA) procedures, structure-searching (such as exploratory factor analysis, EFA) not assume any a-priori hypothesis, but the boundaries between the approaches are becoming increasingly blurred:

- mathematically, a one-dimensional CFA is identical to an EFA
- there are hybrids between these two procedures, so called *Exploratory Structural Equation Modeling* (Marsh et al. 2014)

→ grouping is based on patterns of variation (correlation)

variable-centred vs. person-centred:

In addition to methods that summarise correlation between variables, it is also useful to identify subgroups in data to test the robustness of the statistical models using for example *Latent class and latent transition analysis* (Collins und Lanza 2010) or *cluster analysis* (Hastie, Tibshirani und Friedman 2009; Kassambara 2017); innovative idea is for example the <https://github.com/m-Py/anticlust>

→ grouping is based on the distance (proximity)

overview: confirmatory factor analysis (CFA)

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation

reliability

statistical models

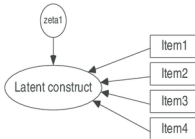
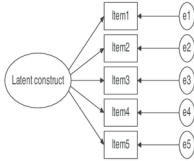
analysis plans

exploratory models

common factor models

item response models

latent class analysis

Description	Formative measurement model	Reflective measurement model
Epistemic relationship		
Criteria	<p>(1) Direction of causality</p> <p>Flow from measures to latent construct</p> <p>Measures define latent construct</p> <p>Measures should not be interchangeably</p> <p>Measures need not have the similar content or common theme</p> <p>Dropping measures should not change the conceptual meaning of the latent construct</p> <p>Not necessary for measures to correlate with each other</p> <p>(3) Correlation among the measures</p> <p>(4) Nomological net of the measures may differ</p>	<p>Flow from latent construct to measures</p> <p>Construct defines measures</p> <p>Measures should be interchangeably</p> <p>Measures should have the similar content or common theme</p> <p>Dropping measures should not change the conceptual meaning of the latent construct</p> <p>Measures are expected to correlate with each other</p> <p>Nomological net of the measures should not differ</p>

Source: Urbach and Ahlemann (2010)

fundamental principle of confirmatory factor analysis

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation

reliability

statistical models

analysis plans

exploratory models

common factor models

item response models

latent class analysis

Literature

The central aim of this procedures is to find parameters for the implied covariance matrix of the model $\Sigma(\theta)$ that most closely match the observed data Σ

$$\Sigma = \Sigma(\theta)$$

Deviation of the covariance matrices is the error

General structure of CFA models

$$\mathbf{Y} = \mathbf{\Lambda} \boldsymbol{\eta} + \boldsymbol{\varepsilon}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{bmatrix} = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1K} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{m1} & \lambda_{m2} & \cdots & \lambda_{mK} \end{bmatrix} \cdot \begin{bmatrix} \eta_1 \\ \vdots \\ \eta_K \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}$$

Model-implied variance-covariance matrix

$$\Sigma(\theta) = \mathbf{\Lambda} \boldsymbol{\Psi} \mathbf{\Lambda}' + \boldsymbol{\Theta}_\varepsilon$$

overview: item response theory (IRT)

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation

reliability

statistical models

analysis plans

exploratory models

common factor models

item response models

latent class analysis

Literature

central idea of the IRT: The probability of an item response is determinable by the person ability / attitudes / ... on the measured latent trait, as well as one or more parameters describing the items (G. Fischer und Molenaar 1995, p. 4)

In general, there are three possible applications:

- taxonomy of binary item response models (e.g. Rasch Model, also called One Parameter Logistic Model)
- taxonomy of polytomous item response models (e.g. Graded Response Model) - G. Fischer und Molenaar 1995
- „explanatory“ item response models (for example to detect ordering effects or differential item functioning, ...) - De Boeck und Wilson 2004

overview: latent class analysis (LCA)

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques.
dev.

item development

domain
identification

item generation

scale development

cognitive

interviews

sampling

scale evaluation

reliability

statistical
models

analysis plans

exploratory models

common factor
models

item response

models

latent class analysis

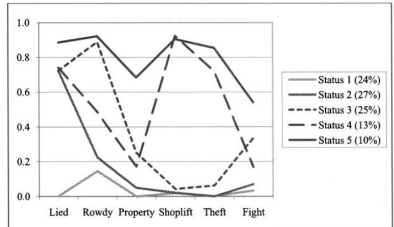
central idea of LCA: to detect latent (or unobserved) heterogeneity in samples

Table 1.3 Four-Latent-Class Model of Past-Year Delinquency (Add Health Public-Use Data, Wave I; $N = 2,087$)

Assigned label	Latent Class			
	1 Non-/Mild Delinquents	2 Verbal Antagonists	3 Shoplifters	4 General Delinquents
Probability of membership	.49	.26	.18	.06
Conditional probability of a Yes* response				
Lied to parents	.33	.81[†]	.78	.89
Publicly loud/rowdy/unruly	.20	.82	.62	1.00
Damaged property	.01	.25	.25	.89
Stolen something from store	.03	.02	.92	.88
Stolen something worth < \$50	.00	.03	.73	.88
Taken part in group fight	.04	.31	.24	.64

*Recoded from original response categories.

[†]Conditional probabilities > .5 in bold to facilitate interpretation.



Collins und Lanza (2010)

Literature

Motivation

DGP

CASM

test model

Latent Variables

measurement

test theory

total survey error

Generalisability

template ques. dev.

item development

domain identification

item generation

scale development

cognitive interviews

sampling

scale evaluation

reliability

statistical models

analysis plans

exploratory models

common factor models

item response models

latent class analysis

1. Arlt, Anthony et al. (2018). Developing questionnaires for educational research. *BMJ: Guide No. 87*. In: *Statistical master* 10.6, S. 402-426.
2. Bartholomew, David (2003). *Structural Varieties, Models and Misconceptions*. Springer.
3. Bowling, Gailford D. et al. (2002). *Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer*. In: *Frontiers in public health* 5, S. 3-16.
4. Bollen, Kenneth (1989). *Structural Equations with Latent Variables*. John Wiley.
5. Collins, Linda M. and Stephen T. Lewis (2002). Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences. John Wiley & Sons.
6. Cooper, Henry, Larry Hedges and Jeffrey Vechstein (2009). *The handbook of research synthesis and meta-analysis, Second Edition*. Russell Sage Foundation.
7. Cronbach, Lee and Paul Meehl (1955). Coefficient validity in psychological tests. In: *Psychological bulletin* 52.4, S. 262-802.
8. De Boeck, Paul and Mark Wilson (2006). *Explanatory item response models*. Springer.
9. Dillen, Herman (2002). Qualitätskontrolle und Qualitätsicherung in Schule und Universität. Ein Überblick aus Stand der empirischen Forschung. In: *Zeitschrift für Pädagogik* 48.
10. Fabrigar, Leung et al. (2004). Statistik: Der Weg zur Datenanalyse. Springer-Verlag.
11. Fackhaus, Frank (2014). "Total survey error". In: *Handbuch Methoden der empirischen Sozialforschung*. Springer, S. 439-455.
12. Fischer, Gerhard and Ina Miesner (2001). *Item Models, Foundations, Recent Developments, and Applications*. Springer.
13. Fischer, Norde et al. (2011). Geometrische Entwicklung. Qualität. Bildungen: Längsschnittdatensätze der Studie zur Entwicklung von Geisteswissenschaften (BEE) Seite 1000.
14. Frühwirth, Gidon and Anna-Lena Schulten (2018). Cognitive Models in Intelligent Research: Advantages and Recommendations for Their Application. In: *Journal of Intelligent* 6.33, S. 3-22.
15. Giesecke, Heide et al. (2018). "Kriterienkatalog zur Bewertung psychodiagnostischer Selbstbeurteilungsinstrumente-Empfehlung des Deutschen Kollegiums für Psychometrische Methoden (DKPM)". In: *PPM/PPM/psychometrie Psychometrie - Methodische Psychologie* 45.07, S. 266-296.
16. Giesecke, Heide et al. (2018). "An automated approach to identifying search terms for systematic reviews using keyword co-occurrence networks". In: *Methods in Ecology and Evolution* 19.10, S. 1849-1858.
17. Giesecke, Robert M. and Lars Lyberg (2010). "Total survey error: Past, present, and future". In: *Public opinion quarterly* 74.3, S. 489-676.
18. Hambleton, Ronald and Russell Jones (1991). *An Rasch measurement model on: Comparison of classical test theory and item response theory and their applications to test development*. In: *Educational measurement issues and practices* 12.3, S. 39-47.
19. Hartzel, Thomas, Robert Thibaut and Jerome Freudenau (2008). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.
20. Hays, Rob (2011). *Handbook of structural equation modeling*. Guilford Press.
21. Jöreskog, Kenneth and David Sörbom (2003). *Theory Construction and Model-Building Skills: A Practical Guide for Social Scientists*. Guilford Publications.
22. Kane, Michael (2012). "Matching more interpretations and uses". In: *Language Testing* 29.3, S. 3-17.
23. Kattmann, Ulrike (2017). *Practical guide to cluster analysis in R: Unsupervised machine learning*. John Wiley.
24. Kuhn, Peter (2011). *Statistische Verfahren der Bildungsforschung: Wie das Bildungswesen die Statistik im Blickfeld bekommt*. Springer-Verlag.
25. Kuhn, Peter and Jörg Voder (2016). *Statistische Methoden in der Bildungsforschung*. Beltz Juventa, S. 229-248.
26. Kuhn, Peter (2011). *Principles and practice of structural equation modeling*. Guilford publications.
27. Leinens, Kathi, Inge Ritz and Hans Ellrich (2011). *International handbook of survey methodology*. Routledge.
28. Little, Robert and Markus Fackhaus (2014). *Applied time series econometrics*. Cambridge University Press.
29. Marks, English (2014). *Algebraic Test Theory*. Universität de Paderborn.
30. Markov, Kathi and Denny Borsboom (2013). *Frontiers of test validity theory: Dimension, variation, and meaning*. Routledge.
31. Mark, Robert W. et al. (2014). *Exploratory structural equation modeling: the integration of the best features of exploratory and confirmatory factor analysis*. In: *Annual review of clinical psychology* 13, S. 59-103.
32. Meade, Norde (2018). *Validity of psychological assessment: Validation of inference from person responses and performance on scientific inquiry into score meaning*. Techn. Rep. Conference on Contemporary Psychological Assessment.
33. Miller, Robert et al. (2014). *Cognitive interviewing methodology*. John Wiley & Sons.
34. Montgomerie, Neil and Augustin Kellou (2012). *Testtheorie und Fragebogenkonstruktion*. Springer.
35. — (2012). *Testtheorie und Fragebogenkonstruktion*. Springer.
36. Peng, Roger and Elizabeth Mainel (2014). *The Art of Data Science: A guide for anyone who works with Data*. Stylized consulting LLC.
37. Penny, Mike (2018). *Understanding education indicators: A practical primer for research and policy*. Teachers College Press.
38. Rayner, Andrew and Marjorie C. Jones (2006). "Essential elements of questionnaire design and development". In: *Journal of Clinical Nursing* 16.3, S. 340-361.
39. Revelle, William and Mark Eisinger (2018). "Confidence alpha, beta, omega, and the gfc: Comments on Nijman". In: *Psychometrika* 74.1, S. 103-104.
40. Reubens, Alexander et al. (2011). "Die Bedeutung der Itemsanalyse und der Modellhaft für die Langzeitstudien-Erfassung von Kompetenzen". In: *Zeitschrift für Entwicklungspsychologie und pädagogische Psychologie*.
41. Ren, Jürgen (2016). *Leitfaden Testtheorie Testkonstruktion (2., vollständig überarbeitete und erweiterte Auflage)*.
42. Schmidt, Hans, Günter and Manfred Amelang (2010). *Psychologische Diagnostik* 8. Auflage. Springer-Verlag.
43. Schmitt, Anna-Lena et al. (2018). "Test Characteristics of Diffusion Model Parameters". In: *Journal of Intelligence* 6.7, S. 3-22.
44. Schmitt, Michael (2017). *Latent Variablenmodelle in der empirischen Bildungsforschung: Die Schule und Struktur der Schulentz in der Welt*. Oldenbourg. Technische Universität Darmstadt.
45. Schmitt, William W., Thomas D. Cook and David E. Kousser (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Harcourt College Learning.
46. Shmida, Yael (2019). "On the use, misuse, and the very limited usefulness of Cook's alpha". In: *Psychometrika* 74.1, S. 107-120.
47. Simpson, Scott (2010). "Conducting a meta-analysis: What to consider when choosing statistics for a study". In: *The Canadian Journal of Hospital Pharmacy* 65.4, S. 310-317.
48. Staudt, Andrea and Sophia Kuhn-Held (2014). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. De Gruyter.
49. — (2017). *Latent variable modeling a survey*. In: *International Journal of Statistics* 10.4, S. 710-730.
50. Staudt, Andrea (2011). "The relevance of the CFF evaluation model for Educational Accountability". Paper presented at the Annual Meeting of the American Association of School Administrators.
51. Torgth, Rudolf and Bernhard Schmidt-Helja (2014). *Handbuch Bildungsforschung*. Springer.
52. Tourangeau, Roger and Norman Bradburn (2014). "The Psychology of Survey Response". In: *Handbook of Survey Research*. Emerald Group, S. 529-558.
53. Tourangeau, Robert, Dale and Jack Aakre (2018). *Best alternatives to Cook's alpha validity in validity conditions: arguments and empirical measurements*. In: *Frontiers in psychology* 7.
54. Ullrich, Peter, Frederik Bollen et al. (2010). *Structural equation modeling in information systems research using partial least squares*. In: *Journal of Information technology theory and applications* 12.3, S. 9-33.
55. Wills, Gordon (2014). *Cognitive interviewing: A tool for improving questionnaire design*. Sage Publications.
56. Zumbach, Walter et al. (2009). *Statistik für Schüler und Lehrer: Eine Einführung für Wissenschaftler und Sozialwissenschaftler*. Springer-Verlag.