

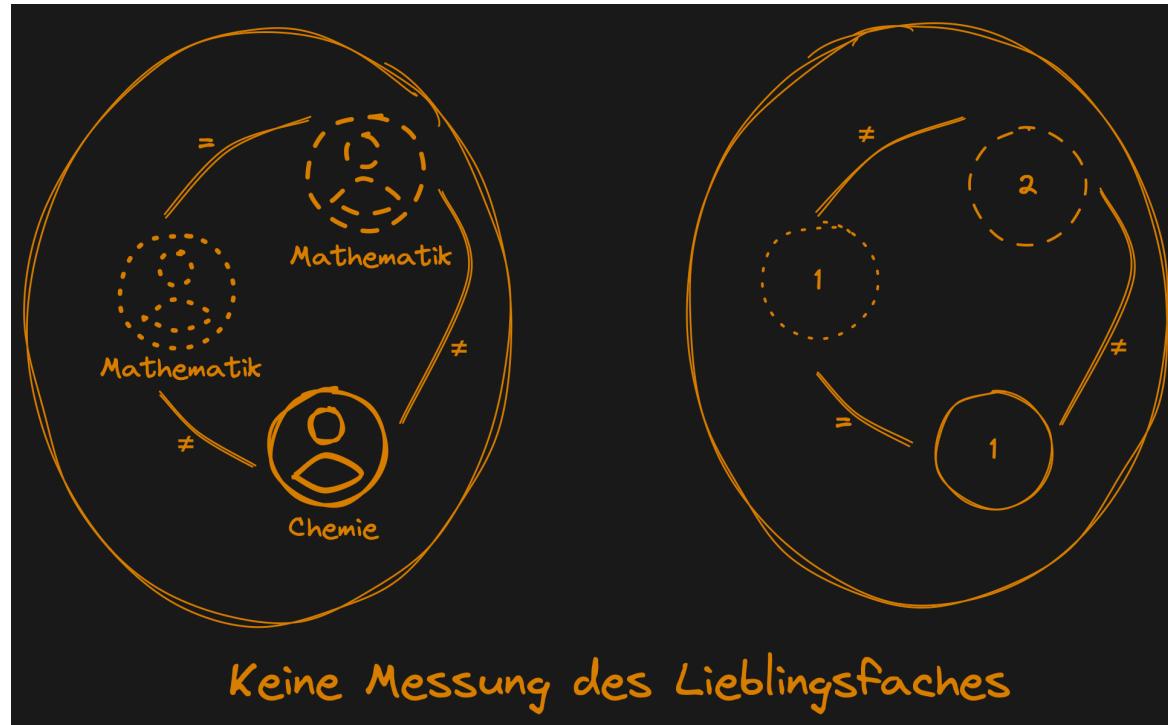
Grundbegriffe der Messtheorie

Samuel Merk

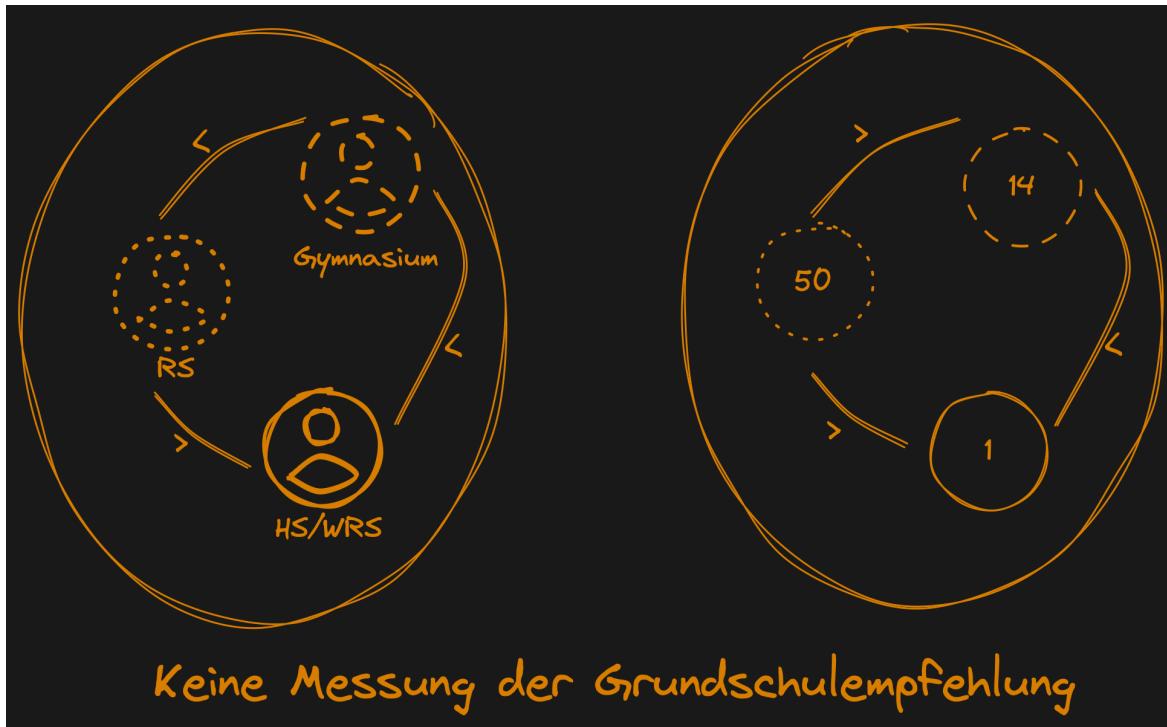
 Was ist eine Messung?

Messung als Homomorphismus

Eine Messung („measurement“) meint in der quantitativen Sozialforschung eine **Zuordnung von Objekten zu Zahlen**, sofern diese Zuordnung eine **homomorphe (strukturerhaltende) Abbildung** eines empirischen Relativs in ein numerisches Relativ ist (Orth, 1983).



Messung als Homomorphismus



Weiterführende Literatur

- Eid, M., Gollwitzer, M., & Schmitt, M. (2013). Statistik und Forschungsmethoden: Lehrbuch. Mit Online-Materialien. Beltz.
Steyer, R., & Eid, M. (2013). Messen und Testen. Springer-Verlag.



Typologien von Variablen

Typologie I: Skalenniveaus

Das Skalenniveau einer Variable beschreibt, welche Relationen im numerischen Relativ sinnvoll sind (Döring & Bortz, 2016).

Skalenniveau	Sinnvolle Relationen	Beispiel
Nominalskala	Gleichheit	Lieblingsfach
Ordinalskala	Gleichheit + Ordnung	Grundschulempfehlung
Intervallskala	Gleichheit + Ordnung + Abstand	tägl. Internetnutzung

Typologie II: Inferenzniveau

Das Inferenzniveau einer Variable beschreibt, in welchem Ausmaß bei einer Messung geschlussfolgert/abstrahiert wird ([Lotz, Gabriel, & Lipowsky, 2013](#)).

Niedriginferente Variablen

- Sind mehr oder weniger direkt beobachtbar
- Werden durch “Kodierungen” generiert
- *Beispiel: “Anzahl der Fehltage einer Schülerin”*

Hochinferente Variablen

- Sind nicht direkt beobachtbar
- Werden durch “Ratings” generiert
- *Beispiel: “Kognitive Aktivierung”*

Typologie III: Theo. Kausalzusammenhang

Als **abhängige Variable (AV)** wird eine (kausal) beeinflusste Variable bezeichnet, während die **unabhängige Variable (UV)** die beeinflussende Variable darstellt. Eine **Mediatorvariable (MeV)** ist zugleich AV und UV. Beeinflusst eine Variable eine Wirkung bezeichnet man sie als **moderierende Variable (MoV)** (Eid, Gollwitzer, & Schmitt, 2013).



Weiterführende Literatur

Pearl, J., Glymour, M., & Jewell, N. P. (2016). Causal inference in statistics: A primer. Wiley.



Gütekriterien einer Messung

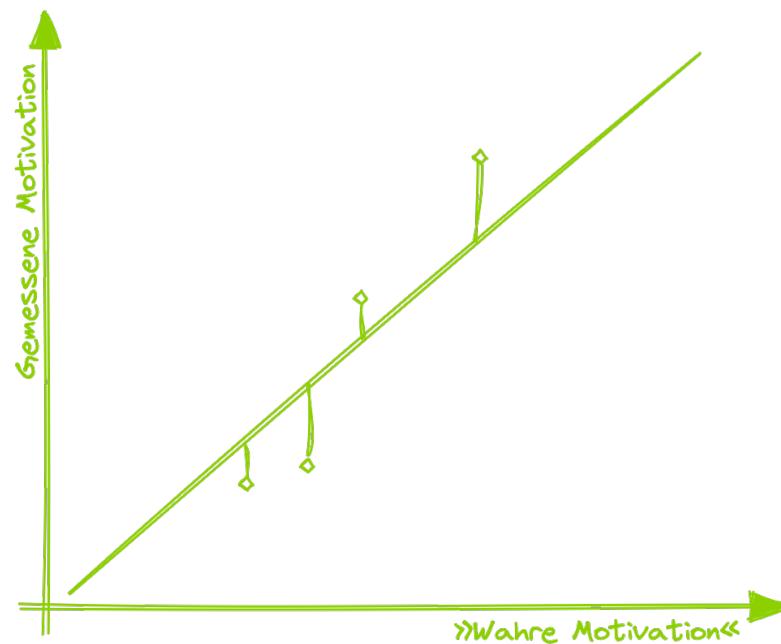
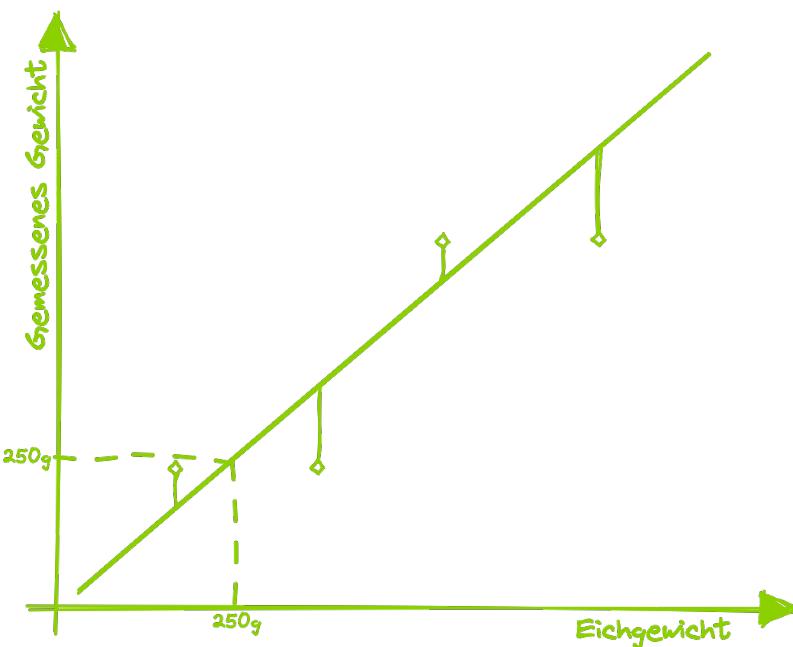
Objektivität

Zur Beschreibung der Objektivität einer Messung oder eines Tests wird typischerweise zwischen der **Durchführungsobjektivität**, der **Auswertungsobjektivität** und der **Interpretationsobjektivität** differenziert (Döring & Bortz, 2016), also der Unabhängigkeit des Ergebnisses der Messung/des Test von der durchführenden, auswertenden bzw. interpretierenden Person.



Reliabilität

Reliabilität ist Ausmaß an Messfehlerfreiheit



Validität

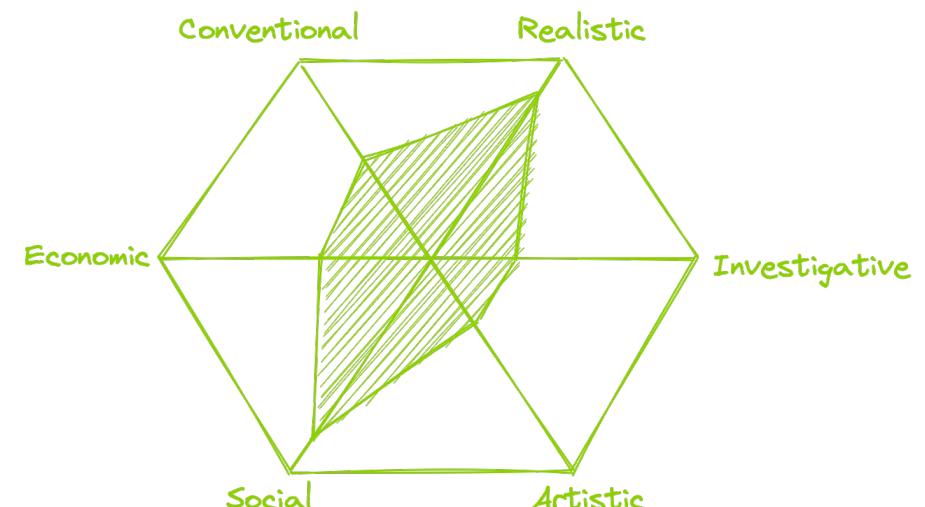
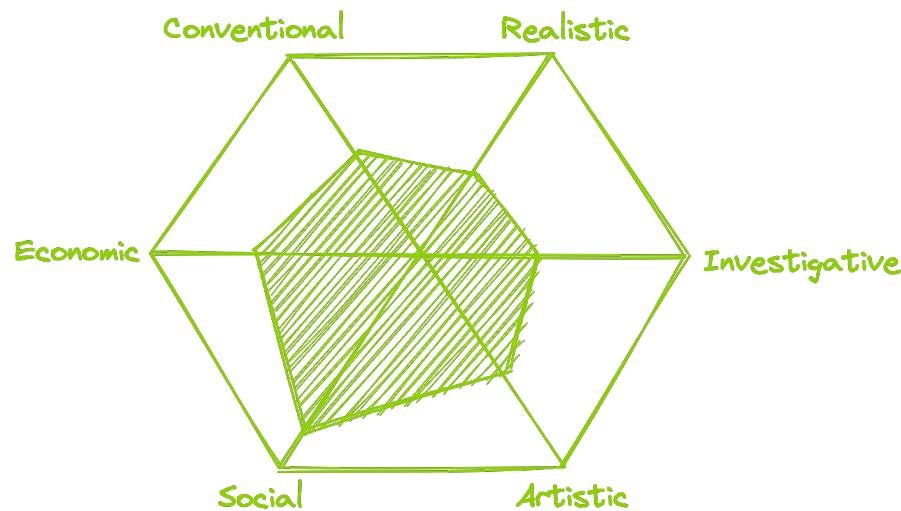
Kriteriumsvalidität

“Ein Test weist Kriteriumsvalidität auf, wenn vom Verhalten der Testperson innerhalb der Testsituation erfolgreich auf ein »Kriterium«, nämlich auf ein Verhalten außerhalb der Testsituation, geschlossen werden kann” ([Moosbrugger & Kelava, 2020](#)).

Validität

Konstruktvalidität

Das Ausmaß der theoretischen und empirischen Belege für die Angemessenheit der Interpretation von Testwerten (Cronbach & Meehl, 1955).





Erhebungsverfahren

Beobachtung

Wissenschaftliche Beobachtung ist die systematische und **regelgeleitete Registrierung des Auftretens** und der Ausprägung von ausgewählten, psychologisch relevanten Merkmalen oder Ereignissen. Sie folgt einem zuvor festgelegten Beobachtungsplan, der festlegt, **1)** was beobachtet werden soll, **2)** welche Aspekte weniger oder nicht relevant sind, **3)** welchen Interpretationsspielraum der Beobachtende bei der Beobachtung hat, **4)** wann, wie lange und wo die Beobachtung erfolgt und **5)** auf welche Weise das Beobachtete registriert und protokolliert wird (**Hussy, Schreier, & Echterhoff, 2013**).

Zentrale Vorteile

- Potentielle Vermeidung von Reaktanz
- Teilw. höhere Reliabilität

Zentrale Nachteile

- Oft nicht ökonomisch
- Kognitive und psychische Variablen oft nur schwierig zugänglich

Selbstauskünfte

Interview, Fragebögen und Tests stellen Beispiele für Selbstauskünfte dar.

Zentrale Vorteile

- Kognitive und psychische Variablen zugänglich
- Teilweise sehr ökonomisch

Zentrale Nachteile

- Reaktanz oft höher
- Entwicklung von Fragebögen und Tests oft sehr aufwändig

Physiologische Messung

Beispiele für physiologische Messungen im bildungswissenschaftlichen Bereich stellen die Messung der Hautleitfähigkeit und der Blickbewegungen dar.

Zentrale Vorteile

- Teilweise nur schwer verfälschbar
- Unbewusste Prozesse werden zugänglich

Zentrale Nachteile

- Meist nicht verdeckt durchführbar
- Oft Laborbedingungen erforderlich

Dokumentenanalyse

Bei der Dokumentenanalyse werden bereits existierende schriftliche Artefakte untersucht (z.B. Schriftliche Leitbilder, Klassentagebücher, Schriftwechsel, Homepages). Diese können sowohl in hoch- wie niedrig strukturierten Prozessen gesammelt, transformiert und analysiert werden (qualitative und quantitative Dokumentenanalyse).

Zentrale Vorteile

- Keine Genese von Daten notwendig
- Teilweise sehr ökonomisch

Zentrale Nachteile

- Beschränkte Auswahl an Variablen
- Artefakte müssen präexistieren

👍 Dos and Don'ts bei der Fragebogenerstellung

Fragebogendesign: Begriffe

- Item = Einzelfrage
- Skala = Mehrere Items, die Operationalisierungen einer Variablen darstellen
- Rating-Item = Item, das intervallskaliert ist
- Likert-Item = Item, deren Stamm eine Aussage darstellt und deren Antwortmöglichkeit die abgestufte Zustimmung zu dieser Aussage enthält
- Offenes Item = Item, bei dem die Antwortmöglichkeiten nicht vorgegeben sind
- Mittelkategorie = Antwortmöglichkeit, die äquidistant zu den beiden Polen ist
- Unipolares Item = Item, das geringe bis starke Ausprägungen eines einzigen Merkmals umfasst (z.B. »nie« bis »immer«)
- Bipolares Item = Item, das von einem Pol über eine neutrale Kategorie bis zu einem anderen Pol formuliert ist (z.B. »lehne ab« über »neutral« bis »stimme zu«).

Fragebogendesign: Faustregeln

- Verwenden Sie nach Möglichkeit unbedingt etablierte (also von Wissenschaftler*innen validierte) Items und Skalen. Dies wird umso wichtiger je höherinferent Ihre Variable ist.
- Formulieren Sie Items so, dass Sie ein möglichst hohes Skalenniveau erhalten (statt »Rauchen Sie?«, »Wie viele Zigaretten rauchen Sie durchschnittlich pro Tag?«)
- Vermeiden Sie mehrdimensionale Interpretationsmöglichkeiten (typisch bei und/oder Verknüpfungen z.B. »Ich fahre sehr gerne und sehr schnell Auto«)
- Vermeiden Sie (doppelte) Verneinungen (z.B. »Ich bin nicht oft unglücklich«)
- Implementieren Sie Qualitätschecks (z.B. sog. Items mit sicherer Antwort im gleichen Antwortformat wie die vorherigen Items »Ich bin noch nie im Leben zu spät gekommen«)
- Planen Sie für jedes Item eine Auswertungsstrategie **bevor** Sie die Daten erheben.

Literatur

- Cronbach, L. J., & Meehl, P. E. (1955). *Construct validity in psychological tests.* *Psychological Bulletin*, 52(4), 281.
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5., vollst.). Berlin, Heidelberg: Springer.
- Eid, M., Gollwitzer, M., & Schmitt, M. (2013). *Statistik und Forschungsmethoden: Lehrbuch. Mit Online-Materialien* (Third). Beltz.
- Hussy, W., Schreier, M., & Echterhoff, G. (2013). *Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Lotz, M., Gabriel, K., & Lipowsky, F. (2013). *Niedrig und hoch inferente Verfahren der Unterrichtsbeobachtung. Analysen zu deren gegenseitiger Validierung*.
- Moosbrugger, H., & Kelava, A. (Eds.). (2020). *Testtheorie und Fragebogenkonstruktion*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Orth, B. (1983). Grundlagen des Messens. In H. Feger & J. Bredenkamkp (Eds.), *Enzyklopädie der Psychologie: Themenbereich B, Serie I Forschungsmethoden der Psychologie, Bd. 3: Messen und Testen* (pp. 136–180). Göttingen: Hogrefe.