

Forschungsmethoden

VORLESUNG WS 2018/2019

FLORIAN KOBYLKA

Rückblick

Gütekriterien:

Wissen, wozu Gütekriterien gebraucht werden

✓

Gütekriterien kennen und einordnen können:

- Qualitative Gütekriterien
- Validitätskriterien

✓

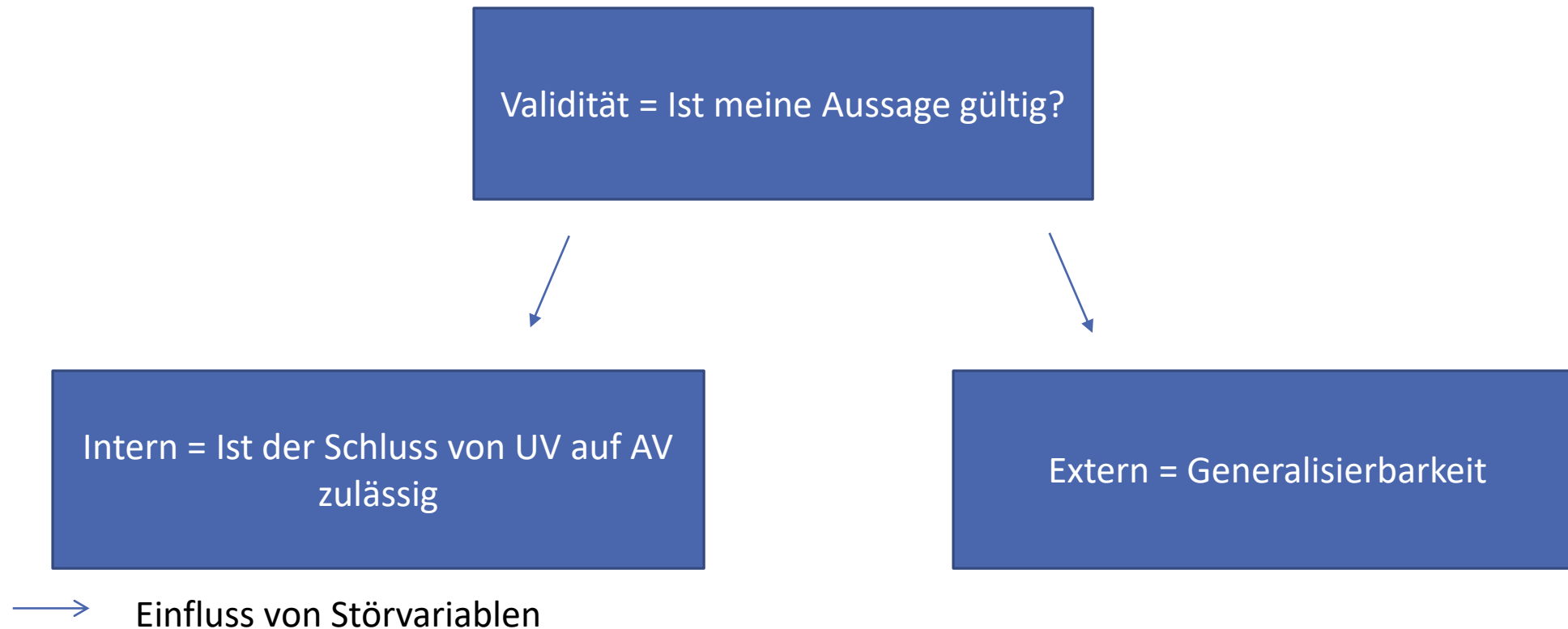
✓

Gütekriterien bewerten und Forschung anhand von Gütekriterien bewerten können (✓)

Indikatoren kennen und Beeinflussen können, um eigene Forschung zu verbessern (✓)

Nachtrag: interne und externe Validität

Interne vs. externe Validität



Heute:

1. Qualitätsmerkmale und deren Erfassung II

Objektivität

Reliabilität

Mixed-Methods Gütekriterien

Termine

#	Datum	Thema	Inhalts-/Zielnummer(n)
1	19. Okt	Einführung & empirische Psychologie	1. 11.
2	26. Okt	Logik und Wissenschaftstheorie I	1. 11.
3	2. Nov	HA 1: Artikel lesen	2. 15. 16.
4	09. Nov	Wissenschaftstheorie II	1. 11.
5	16. Nov	Forschungstraditionen & Scientific Method	1. 11. 15.
6	23. Nov	Entwicklung des Faches I & Psychometrie	1. 3. 7. 10.
7	30. Nov	Entwicklung des Faches II & Hypothesen	3. 4. 6. 12. 15.
8	07. Dez	Versuchs- & Stichprobenplanung	4. 7. 12. 13. 14.
9	14. Dez	Testtheorie	1. 4. 7. 11. 13. 16.
10	21. Dez	Selbststudiumsaufgabe	2. 3. 4. 6. 8. 14. 15.
11	11. Jan	Testtheorie und Replikationskrise	1. 4. 7. 8. 11. 12. 13. 16.
12	18. Jan	Gütekriterien I	5.
13	25. Jan	Gütekriterien II	5.
14	01. Feb	Erhebungstechniken I: Selbstberichtsverfahren & Beobachtung	7. 9. 10.
15	08. Feb	Erhebungstechniken II: Objektive Daten & Psychologische Tests	7. 9. 10.
16	15. Feb	Ethik & Klausurvorbereitung	8.

Gütekriterien II

Objektivität

= intersubjektive Vergleichbarkeit

- Testergebnis soll unabhängig vom Testanwender sein
- Forschungsergebnis soll unabhängig vom Forscher sein

Objektivität

- **Durchführungsobjektivität**
= standardisierte Durchführung

Objektivität

- Durchführungsobjektivität
- **Auswertungsobjektivität**
= standardisierte Auswertung

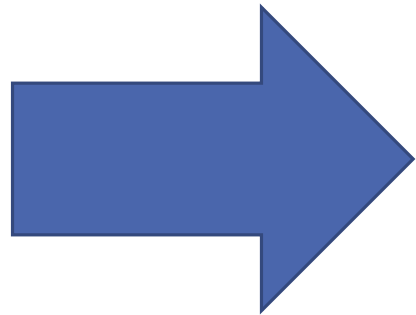
Objektivität

- Durchführungsobjektivität
- Auswertungsobjektivität
- **Interpretationsobjektivität**
= standardisierte Interpretation

Wie kann man die Objektivität maximieren?

- Standardisierung der Testsituation
- Versuchsleiterschulung
- genaue Instruktion
- genaue Richtlinien zur Auswertung und Interpretation
- Normierung

Reliabilität



Beruht auf der Klassischen Testtheorie (KTT)

Klassische Testtheorie

Beispiel 1: Messung des Gewichtes eines Buches mit einer Waage

- Gewicht bleibt gleich/schwankt nur gering, Waage ist reliabel

Beispiel 2: IQ-Messung

- Wert 1. Termin: IQ = 135
- Wert 2. Termin: IQ = 119
 - hat sich die Intelligenz der Person verändert?

Klassische Testtheorie - Annahmen

$$X = T + E$$

Beobachteter Wert Wahrer Wert („true score“) Messfehler („error score“)

The diagram illustrates the classical test theory equation $X = T + E$. The variable X is labeled as 'Beobachteter Wert' (Observed score), T as 'Wahrer Wert („true score“)' (True score), and E as 'Messfehler („error score“)' (Error score). Blue arrows point from each label to its corresponding variable in the equation.

Klassische Testtheorie - Annahmen

- der wahre Wert (T) ist immer gleich
- der Fehler (E) variiert von Messung zu Messung
- der Fehler (E) ist unabhängig vom wahren Wert (T)
- die Fehler (E) zweier Tests hängen nicht zusammen
- die Fehler (E) in einem Test sind unabhängig von den wahren Werten in anderem Test
- Aus der Unabhängigkeit des Fehlers ergibt sich: $\mu(E)=0$
- $X = T + E$ & $\mu(E)=0$ ➔ $\mu(X) = \mu(T)$

Reliabilität

- Messgenauigkeit eines Tests
- Varianzzerlegung: $\sigma^2(\mathbf{x}) = \sigma^2(\tau) + \sigma^2(\varepsilon)$
- Anteil der Varianz der wahren Werte an der Varianz der beobachteten Werte:

$$\text{Rel} = \frac{\sigma^2(\tau)}{\sigma^2(\mathbf{x})} = \frac{\sigma^2(\mathbf{x}) - \sigma^2(\varepsilon)}{\sigma^2(\mathbf{x})}$$

Arten der Reliabilitätsbestimmung

- Retest-Reliabilität
- Paralleltest- Reliabilität
- Testhalbierungs- Reliabilität
- Interne Konsistenz
- Interrater-Reliabilität

Retest-Reliabilität

- Testwiederholung
- gleicher Test zu 2 Zeitpunkten an derselben Stichprobe
- Bestimmung der Korrelation zwischen 2 Messungen:
 - Zeitraum zu kurz vs. zu lang
 - relative vs. absolute Stabilität
- pro: einfach in der Konstruktion
- con: Erinnerungs-, und Übungseffekte & Reaktivität

Paralleltest-Reliabilität

- zwei inhaltlich äquivalente Tests zu zwei Zeitpunkten an derselben Stichprobe gemessen
- Korrelation der beiden Tests
- pro: reduziert Erinnerungseffekt und Betrug
- con: aufwendig in Konstruktion

Testhalbierungs-Reliabilität

- einmalige Testdurchführung an einer Stichprobe
- Aufteilung des Tests im Nachhinein in zwei möglichst gleiche Hälften
- Bestimmung der Korrelation zwischen den beiden Testhälften
- pro: einfach berechnet
- con: inhaltlich nicht immer sinnvoll

Interne Konsistenz

- Jedes Item wird als theoretischer Paralleltest betrachtet
- Schätzung der durchschnittlichen Korrelation zwischen den Items unter Berücksichtigung der Testlänge

pro: bei einmaliger Erfassung möglich, präziser als Testhalbierung

con: nur sinnvoll möglich bei homogenen Tests

Sonderfall: Interrater-Reliabilität

Beurteilungen von Experten werden gemessen:

- Intrarater-Reliabilität
- Interrater-Reliabilität: Maß für die Objektivität

Einflussfaktoren auf die Reliabilität

- Homogenität /Heterogenität der Items
- Objektivität der Messung
- Varianz des Merkmals
- Messfehler

Wie hoch sollte die Reliabilität sein?

- möglichst hoch
- abhängig vom Merkmal
- Vergleich mit ähnlichen Verfahren
- Abhängig von Einsatzbedingungen
- Kosten-Nutzen-Abwägung

Wie kann man die Reliabilität verbessern?

- Homogenität erhöhen (wenn sinnvoll)
- Testverlängerung
- Messfehler reduzieren
 - durch klare Instruktion
 - durch klare Formulierung der Items
 - durch standardisierte Testbedingungen

Kritik an der klassischen Testtheorie

Vorzüge:

- Wenige Grundannahmen
- Ableitung der Reliabilität

Kritikpunkte:

- Axiome empirisch nicht prüfbar
- Nichtbeachtung des Skalenniveaus
- Zweifel an Unkorreliertheit von wahrem Wert und Messfehler
- Nur für stabile Eigenschaften

Nebengütekriterien

- Skalierung
- Normierung
- Testökonomie
- Nützlichkeit
- Zumutbarkeit
- Unverfälschbarkeit
- Fairness
- Akzeptanz

Skalierung

- Korrespondenz von Testergebnis und empirischen Merkmalsunterschiede zwischen den Merkmalsträgern
- Skalenniveauabhängig

Normierung

- Referenzerhebung, um die Einordnung und Interpretation der Werte einer Person zu vergleichen
- Umfangreiche Erhebung
- pro: schnelle Einschätzung
- con: nur für Konstante Merkmale

Testökonomie

- Test soll einen möglichst geringen finanziellen und zeitlichen Aufwand einnehmen
- abhängig vom Erkenntnisgewinn

Nützlichkeit

- praktische Relevanz
- mehr Nutzen als Schaden von Forschung/ Maßnahmen /Entscheidungen

Zumutbarkeit

- Testperson soll psychisch, körperlich und zeitlich nicht unnötig belastet werden
- abhängig von Verwendungszweck

Unverfälschbarkeit

Verfahrenskonstruktion derart, dass Testergebnisse nicht manipuliert werden können

- 1) Motivationaler Ansatz: Nutzen für Teilnehmer
- 2) Transparenz: vollständige Aufklärung erst nach dem Test

Fairness

keine systematisch Benachteiligung oder Bevorzugung einzelner Personengruppen

Akzeptanz

Subjektive Meinungen, Bewertungen oder gesellschaftspolitische Überzeugungen für oder gegen ein Verfahren

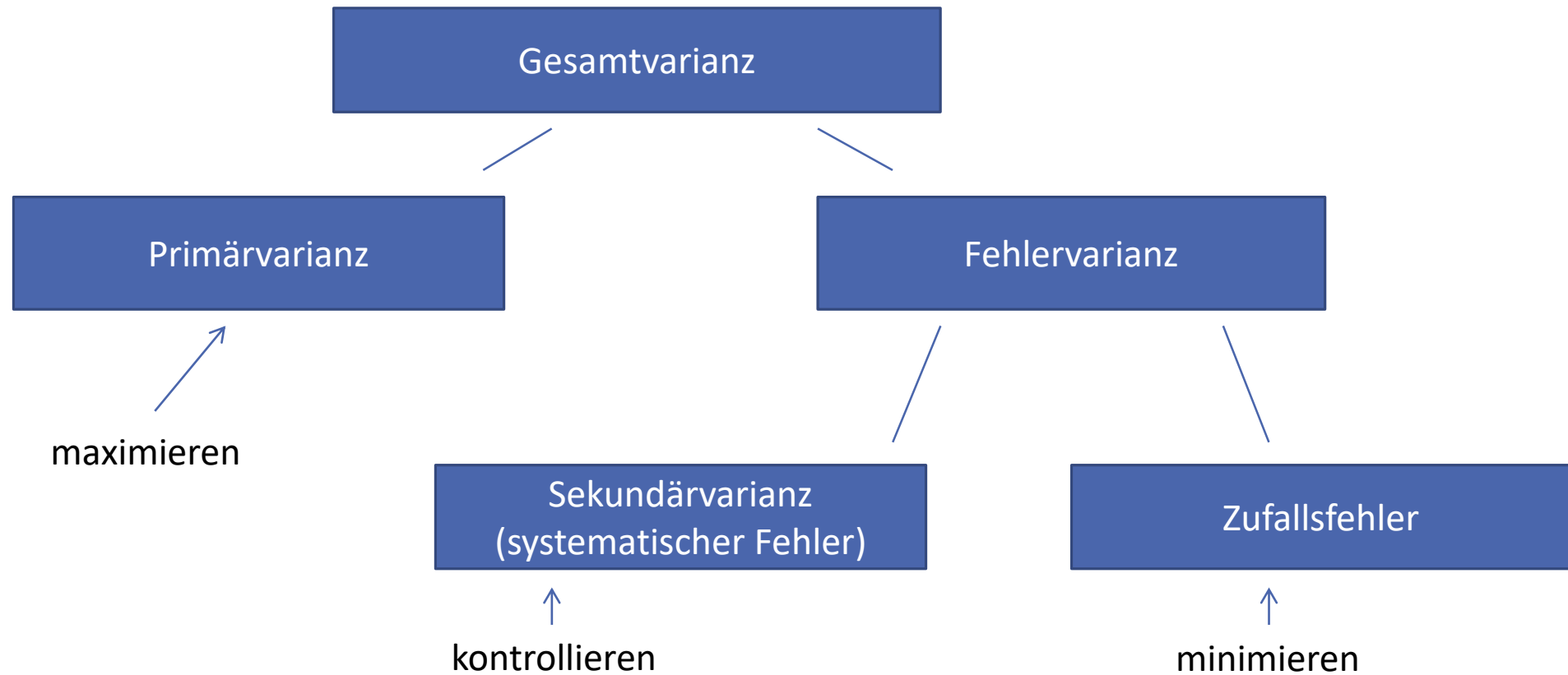
Fazit

Gütekriterien sollten im Forschungsprozess kontinuierlich reflektiert werden,

wobei nie alle Kriterien gleichzeitig perfekt erfüllt sein können.

Sie dienen der eigenen potentiellen Qualitätsverbesserung und dem wissenschaftlichen Diskurs.

Max-kon-min-Prinzip (nach Kerlinger, 1973)



Mixed-Methods Kriterien

- sowohl qualitative, als auch quantitative Kriterien
- extra Arbeitsschritte:
 1. Qualität des Mixed-Designs
 2. Interpretationsqualität
- Kriterien:
 1. Inferenzqualität
 2. Inferenzübertragbarkeit
 - Synthetisierbarkeit
 - Nützlichkeit

Vielen Dank für die
Aufmerksamkeit!

Literatur

- Bortz, J., & Döring, N. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. 5. Auflage. Heidelberg: Springer-Verlag → Kap. 3
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion*. 3. Auflage. München: Pearson Studium → Kap. 2, Kap. 4.3
- Hussy, W., Schreier, M., & Echterhoff, G. (2010). *Forschungsmethoden in Psychologie und Sozialwissenschaften-für Bachelor*. Berlin, Heidelberg: Springer-Verlag. → Kap. 8.1
- Moosbrugger, H., & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion*. 2. Auflage. Berlin, Heidelberg: Springer-Verlag. → Kap. 2, 2.3, 2.7
- Schmidt-Atzert, L., & Amelang, M. (2012). *Psychologische Diagnostik*. 5. Auflage. Berlin, Heidelberg: Springer-Verlag. → Kap. 2.1.2, 2.3 , 2.3.4
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105.
- Cronbach, L.J., & Meehl, P.E. (1955): Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Krohne, H.W. & Hock, M. (2007). *Psychologische Diagnostik. Grundlagen und Anwendungsfelder*. Stuttgart: Kohlhammer → Kap. 3.5
- Kerlinger, F.N. (1973). *Foundations of behavioral research* (2nd ed.). New York: Holt, Rinehart & Winston