

Forschungsmethoden

VORLESUNG WS 2018/2019

FLORIAN KOBYLKA



Rückblick

Ableitung der Hypothesentesttheorie aus dem kritischen Rationalismus

Prüfung der Gleichheitsannahme und ggf. Falsifikation

Kennwerte beim Testen: p , α , β , Teststärke, Effektstärke, α -Niveau

Problematik der Psychologischen Wissenschaft: Replikationskrise

Wenn man 20 mal testet- wie wahrscheinlich ist mindestens ein zufällig signifikantes Ergebnis trotz eigentlichem Nulleffekt?

Nachtrag: Metaanalysen

Metaanalysen

Ergebniszusammenfassung

Einschätzung der wahren Effektgröße

Probleme von Metaanalysen:

1. Heterogene Studienqualität
2. Heterogene Operationalisierung
3. Abhängige Messungen
4. Schubladenproblem

Herangehensweisen:

1. Beschaffung unpublizierter Studien
2. Funnel-plot
3. Fail-Safe-N Methode
4. Weitere Alternativen

Lösungsansätze

- Replizieren
- Wert von Nullbefunden erhöhen
- Effektstärken berechnen und berichten
- Pre-Registration
- Theoriegeleitetes Vorgehen
- Anderes Verfahren: z.B. Bayes-Factor
- Negative Konsequenzen für verzerrende Wissenschaft etablieren

#	Datum	Thema	Inhalts-/Zielnummer(n)
1	26. Apr	Einführung & Logik & empirische Psychologie	1.
2	03. Mai	Selbststudium I (<i>finden, lesen und verstehen</i>)	2. 3. 4.
3	10. Mai	Wissenschaftstheorie	2. 15. 16.
4	17. Mai	Psychometrie & Operationalisierung	1. 3. 6. 10. 11. 15.
5	24. Mai	Entwicklung des Faches	1.
6	31. Mai	Selbststudium II (<i>strukturieren und Abstract formulieren</i>)	2. 15. 16.
7	07. Jun	Versuchs- & Stichprobenplanung	3. 4. 7. 12. 13. 14.
8	14. Jun	Hypothesentest & Replikationskrise	3. 6. 15. 16.
9	21. Jun	Selbststudium III (<i>kritisieren und Ausblick formulieren</i>)	4.
10	28. Jun	Gütekriterien	2. 3. 4. 6. 8. 14. 15.
11	05. Jul	Erhebungstechniken	5.
12	12. Jul	Ethik & Klausurvorbereitung	5.

Gütekriterien

Ziele

Wissen, wozu Gütekriterien gebraucht werden

Gütekriterien kennen und einordnen können

Gütekriterien bewerten und Forschung anhand von Gütekriterien bewerten können

Indikatoren für Gütekriterien kennen und Beeinflussen können, um eigene Forschung zu verbessern

Modulhandbuch 5: Fähigkeit zur fundierten Anwendung psychologischer Untersuchungsmethoden und zur Bewertung von **Objektivität**, **Zuverlässigkeit** und **Gültigkeit** eingesetzter Test-und Messverfahren

Abgrenzung Gütekriterien

Qualitätsbewertung:

- Kriterienkataloge
- Kriterien
- Indikatoren
- Standards

4 allgemeine Standards wissenschaftlicher Forschung

1. Wissenschaftliches Forschungsproblem
2. Wissenschaftlicher Forschungsprozess
3. Wissenschafts- und Forschungsethik
4. Dokumentation des Forschungsprojektes
5. (Anspruch an Wissenschaftlichkeit)

Abgrenzung der Wissenschaft

■ **Tabelle 3.1** Abgrenzung der Wissenschaft von Nicht-, Pseudo- und Parawissenschaft anhand der vier Standards der Wissenschaftlichkeit sowie dem selbsterklärten Anspruch der Wissenschaftlichkeit

	Standard 1: Wissenschaftliches Forschungsproblem	Standard 2: Wissenschaftlicher Forschungsprozess	Standard 3: Wissenschafts- und Forschungsethik	Standard 4: Dokumentation des Forschungsprojektes	Selbsterklärter Anspruch der Wis- senschaftlichkeit
Wissenschaft	+	+	+	+	+
Nicht-Wissenschaft	Teilweise	–	–	–	–
Pseudowissenschaft	Teilweise	–	Teilweise	Teilweise	+
Parawissenschaft	–	+	+	+	+

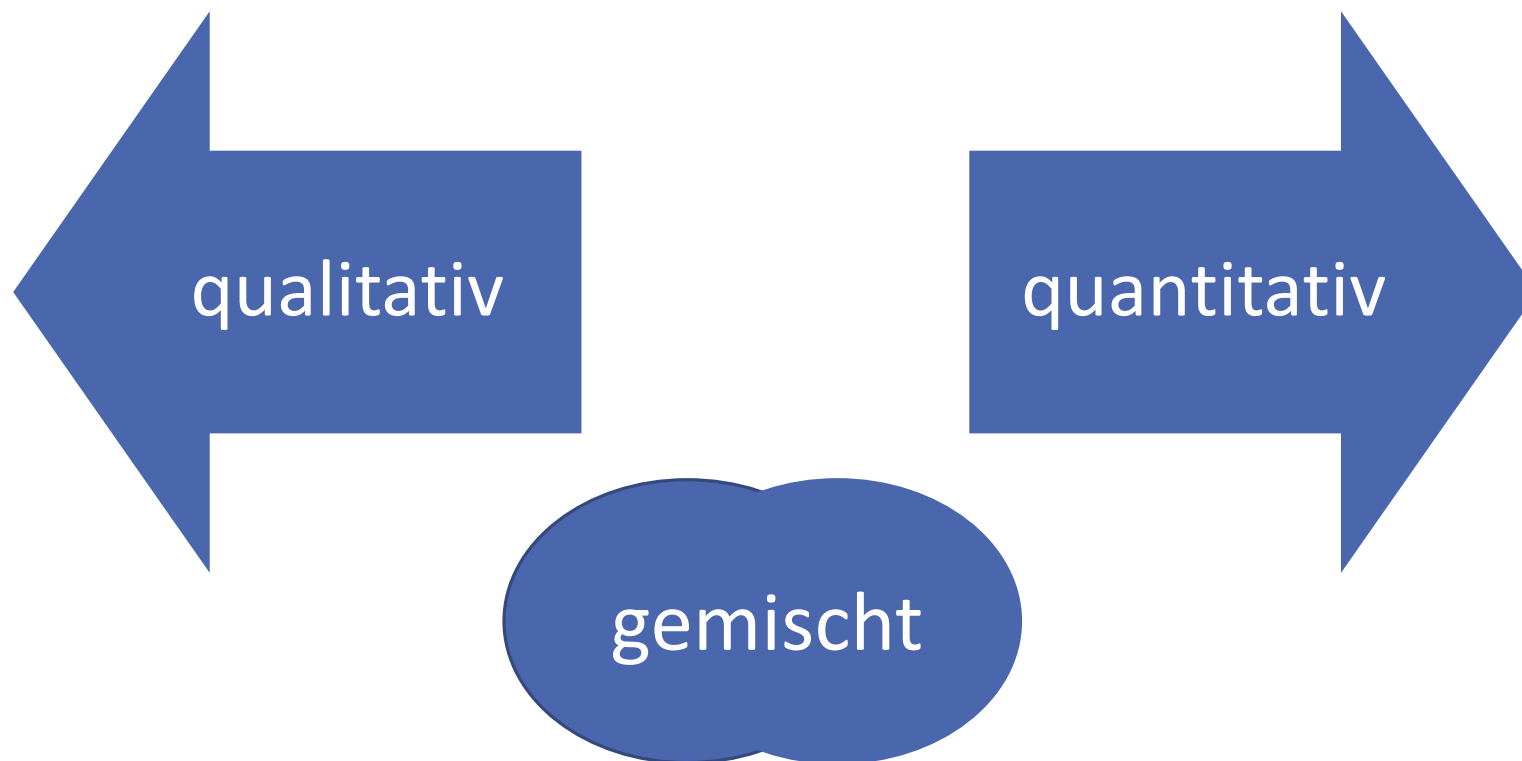
Wissenschaftliche Kriterien

- | | |
|---|-------------------------------------|
| 1. Wissenschaftliches Forschungsproblem: | Inhaltliche Relevanz |
| 2. Wissenschaftlicher Forschungsprozess: | <u>Methodische Strenge</u> |
| 3. Wissenschafts- und Forschungsethik: | Ethische Strenge |
| 4. Dokumentation des Forschungsprojektes: | Präsentationsqualität |
| 5. (Anspruch an Wissenschaftlichkeit): | Selbstreflexion der Qualität |

Warum Gütekriterien?

- Maßstab für wissenschaftliches Arbeiten
- Bewertung von Verfahren und Messinstrumenten
- Unterscheidung in Gütekriterien für qualitative, quantitative und Mixed-Methods-Forschung

Gütekriterien



Gütekriterien der qualitativen Forschung

- kein einheitlicher Konsens
- Beispiele:
 - Ethische Strenge
 - Triangulation
 - Reflektierte Subjektivität
 - argumentative Validität
 - kommunikative Validität

Vier Gütekriterien für qualitative Forschung nach Lincoln & Guba (1985)

- Vertrauenswürdigkeit
 - Übertragbarkeit
 - Zuverlässigkeit
 - Bestätigbarkeit
- } Glaubwürdigkeit

Vier Gütekriterien für qualitative Forschung nach Lincoln & Guba (1985)

Vertrauenswürdigkeit

der Ergebnisse und Interpretationen anhand der Daten zu erreichen über:

- ausführliche Datenerhebung
- Peer Debriefing
- Reflexion aufgrund von Gegenbeispielen
- Interpretationsüberprüfung
- Kommunikative Validierung

Vier Gütekriterien für qualitative Forschung nach Lincoln & Guba (1985)

Übertragbarkeit

der Schlussfolgerungen auf andere Kontexte zu erreichen über:

- Genaue Kontext- und Personenbeschreibungen zur Einschätzung der Möglichkeit der Übertragbarkeit

Vier Gütekriterien für qualitative Forschung nach Lincoln & Guba (1985)

Zuverlässigkeit

des Forschungsprozesses

- ist er nachvollziehbar gestaltet und durchgeführt?

zu erreichen über:

- Darstellung gegenüber Experten
- Prüfung der mit einer Methode gewonnenen Daten durch Verwendung anderer Methode

Vier Gütekriterien für qualitative Forschung nach Lincoln & Guba (1985)

Bestätigbarkeit

der subjektiv geprägten Forschungsarbeit

- keine Vorbestimmung der Ergebnisse durch Interessen oder Sichtweisen der Forschenden
- Zu erreichen über:
 - Darstellung der Daten und des Forschungsprozesses, zum Beispiel auch durch Dokumente zur Selbstreflexion
 - Expertenaudit, Triangulation
 - Problematisch, da Subjektivität der Forscher auch wichtig ist

Hauptgütekriterien

- Validität
- Reliabilität
- Objektivität

Validität



Gültigkeit der
wissenschaftlichen Aussagen

Validitätsarten

- Inhaltsvalidität
- Konstruktvalidität
- Kriteriumsvalidität

Inhaltsvalidität

Def.: Gültigkeit des Inhalts der Untersuchung

- Sind die Items repräsentativ, relevant und umfassend für das zu messende Merkmal?
- Ist die Forschungsfrage deckungsgleich mit dem untersuchten Gegenstand?

Inhaltsvalidität

Beurteilung:

- Wurde der gesamte Inhaltsbereich des Konstrukts durch die Items erfasst?
- Gibt es irrelevante Items?
- Sind die Items für die Erfassung angemessen?

Konstruktvalidität

Def.: Die Gültigkeit der Interpretation empirischer Daten als Indikatoren theoretischer Konstrukte

- Gibt es empirische Belege dafür, dass ein Test das angestrebte Konstrukt erfasst und nicht ein anderes Konstrukt?

Empirische Bestimmung der Konstruktvalidität

- Gruppenunterschiede
- Korrelationen
- Veränderung durch experimentelle Intervention

Multitrait-Multimethod-Ansatz (Campbell & Fiske, 1959)

- Messung mehrerer Eigenschaften
- Verwendung mehrerer Methoden
- Konvergente Validität vs. diskriminante Validität

Multitrait-Multimethod-Ansatz (Campbell & Fiske, 1959)

Tab. 2.21 Multitrait-Multimethod-Matrix

		Methode 1			Methode 2			Methode 3		
		Trait 1	Trait 2	Trait 3	Trait 1	Trait 2	Trait 3	Trait 1	Trait 2	Trait 3
Methode 1	Trait 1	(Rel.)								
	Trait 2		(Rel.)							
	Trait 3			(Rel.)						
Methode 2	Trait 1				(Rel.)					
	Trait 2					(Rel.)				
	Trait 3						(Rel.)			
Methode 3	Trait 1							(Rel.)		
	Trait 2								(Rel.)	
	Trait 3									(Rel.)

Anmerkungen.

»Reliabilitätsdiagonale«: In der Hauptdiagonalen stehen die Reliabilitäten (Rel.) der Verfahren.

Graue Felder = »Validitätsdiagonalen« (»monotrait-heteromethod«): Ein Merkmal wird mit verschiedenen Methoden gemessen.

Blaue Felder = »Heterotrait-Monomethod-Dreiecke«: Verschiedene Merkmale werden mit der gleichen Methode erfasst.

Alle weißen Felder unter der Reliabilitätsdiagonalen = »Heterotrait-Heteromethod-Dreiecke«: Korrelation zwischen verschiedenen Merkmalen, die zudem mit unterschiedlichen Methoden gemessen wurden.

Die Felder über der Hauptdiagonalen bleiben leer.

Bedrohungen der Konstruktvalidität

- Konzept nicht spezifiziert
- Konfundierung
- Mono-Operationalisierung
- Generalisierung von Treatmentausprägungen
- Reaktivität von Selbstauskünften
- Reaktivität der Experimentalsituation
- Versuchsleitereffekt
- Kontroll- und Treatmentgruppeninteraktionen

Kriteriumsvalidität

- Testergebnis soll mit bestimmten Verhaltensweisen, Leistungen etc. außerhalb der Testsituation zusammenhängen
- Güte des Schlusses vom Testergebnis auf das Kriterium

Übereinstimmungsvalidität → Testwerte und Kriterien zeitlich parallel erfasst

prognostische Validität → Kriterium später als Testwerte erfasst

Wahl einer geeigneten Validierungsmethode

- Wird das Merkmal umfassend erfasst und ist der Test repräsentativ für das Merkmal? → Inhaltsvalidität
- Misst der Test das Konstrukt, das er erfassen soll, gibt es empirische Belege? → Konstruktvalidität
- Kann man auf Basis der Testergebnisse ein Kriterium vorhersagen? → Kriteriumsvalidität

Andere Validitätsarten

- inkrementelle Validität
- Augenscheinvalidität
- interne Validität
- externe Validität
- statistische Validität
- ökologische Validität

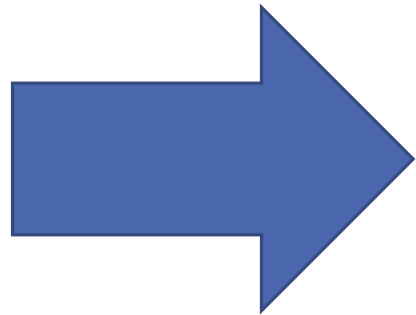
Einflussgrößen auf die Validität

- Konfundierung mit irrelevantem Merkmal
- Reliabilität
- Objektivität
- gemeinsame Methodenvarianz
- Validität von Kriterium/Konstrukt
- Stichprobenmerkmale

Interpretation des Validitätskoeffizienten

- liegt zwischen 0 und 1
- Berücksichtigung von:
 - Richtwerten aus Metaanalysen
 - Ausmaß, in dem ein Test die Qualität von Entscheidungen verbessern kann

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

$$X = T + E$$

Beobachteter Wert

Wahrer Wert („true score“)

Messfehler („error score“)

Klassische Testtheorie

- 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: $\mu(E)=0$
- $X = T + E$ & $\mu(E)=0$ ➡ $\mu(X) = \mu(T)$

Kritik an der klassischen Testtheorie

Vorzüge:

- Wenige Grundannahmen
- Ableitung der Reliabilität

Kritikpunkte:

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

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, geeignet für relativ stabile Merkmale
- con: Absolute Veränderung, Erinnerungs-, und Übungseffekte & Reaktivität, VPN-Aufwand (2x)

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, VPN Aufwand(2x)

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, wenig Aufwand für VPN
- 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, VPN freundlich

con: nur sinnvoll möglich bei homogenen Tests

Einflussfaktoren der 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 Testverlängerung
- Messfehler reduzieren
 - durch klare Instruktion
 - durch klare Formulierung der Items
 - durch standardisierte Testbedingungen

Sonderfall: Interrater-Reliabilität

Beurteilungen von Experten werden gemessen:

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

Objektivität

= intersubjektive Vergleichbarkeit

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

Durchführungsobjektivität

= standardisierte Durchführung

Auswertungsobjektivität

= standardisierte Auswertung

Interpretationsobjektivität

= standardisierte Interpretation

Wie kann man die Objektivität maximieren?

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

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

Rohwerte und standardisierte Werte

- Rohwerte
- Normwerte
- z-Standardisierung
- T-Werte
- Z-Werte
- Stanine Werte
- IQ

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

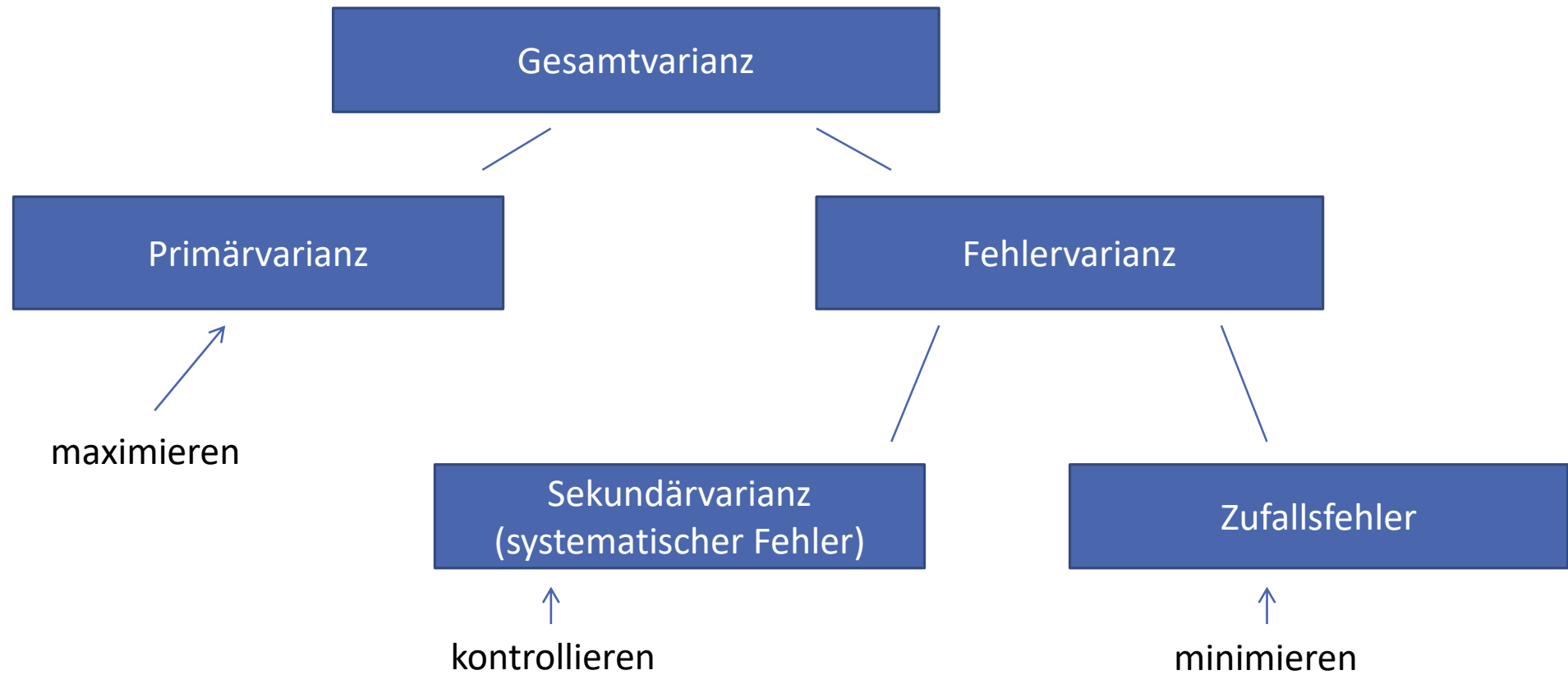
Fazit

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

Übung: Welche der folgenden Aussagen trifft zu?

- 1) Die Inhaltsvalidität wird empirisch bestimmt.
- 2) Beim MTMM-Ansatz stellt der Heterotrait-Monomethod-Koeffizient die Reliabilität dar.
- 3) Bei der prognostischen Validität werden Testwerte und Kriterien zeitlich parallel erfasst.
- 4) Die Konstruktvalidität sind empirische Belege dafür, dass ein Test das angestrebte Konstrukt erfasst.
- 5) Bei der MTMM-Matrix sollten die Heterotrait-Heteromethod-Koeffizienten am höchsten ausfallen.
- 6) Konstruktvalidität und Inhaltsvalidität sind im Kern identisch

Übung:

- Was sind die Gütekriterien für qualitative Forschung?
- Welche verschiedenen Formen der Validität gibt es? Wie kann man diese bestimmen?
- Wie funktioniert der MTMM-Ansatz? Welche Validität wird damit bestimmt?
- Was ist interne Validität, was ist externe Validität?
- Was ist die zentrale Annahme der Klassischen Testtheorie?
- Welche verschiedenen Formen der Reliabilität gibt es und was macht sie aus?
- Wie kann man die Reliabilität eines nicht wiederholbaren Tests erfassen?
- Wie kann man die Reliabilität verbessern?
- Was ist das Max-kon-Min-Prinzip?
- Welche Möglichkeiten gibt es, um mit Störvariablen umzugehen?

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