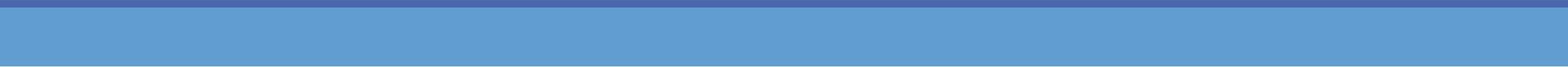


Forschungsmethoden

VORLESUNG SS 2019

FLORIAN KOBYLKA



Rückblick

Gütekriterien

Qualitative Gütekriterien

Quantitative Gütekriterien

Validität

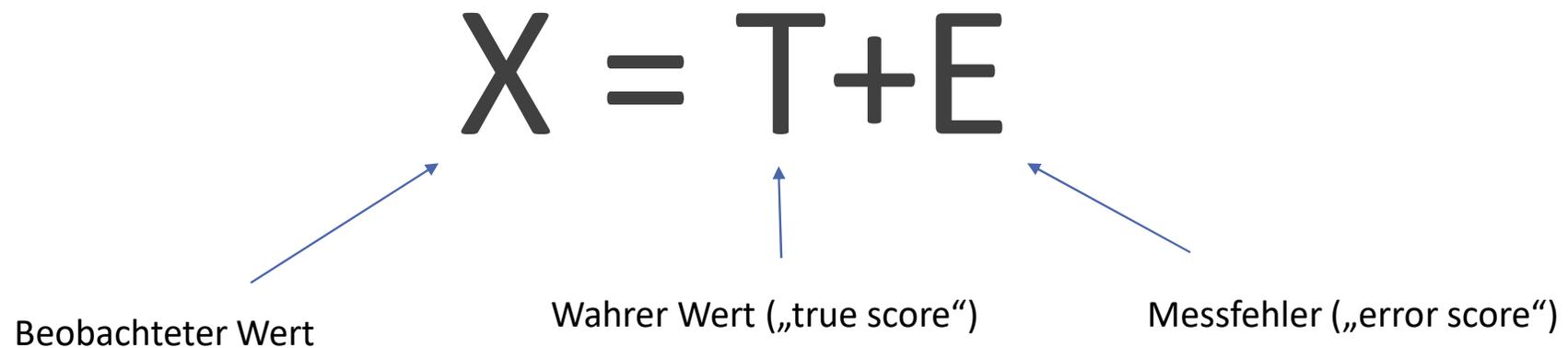
Reliabilität: Gültigkeit der Messung

Nachtrag: Reliabilitätsbestimmung und Objektivität

Klassische Testtheorie

$$X = T + E$$

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

The diagram shows the equation $X = T + E$ in large, bold, black font. Below the equation, three labels are positioned: 'Beobachteter Wert' on the left, 'Wahrer Wert („true score“)' in the center, and 'Messfehler („error score“)' on the right. Three blue arrows point from each label to its corresponding variable in the equation: one from 'Beobachteter Wert' to 'X', one from 'Wahrer Wert („true score“)' to 'T', and one from 'Messfehler („error score“)' to 'E'.

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, schlecht bei heterogenen Tests

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 Objektivität

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

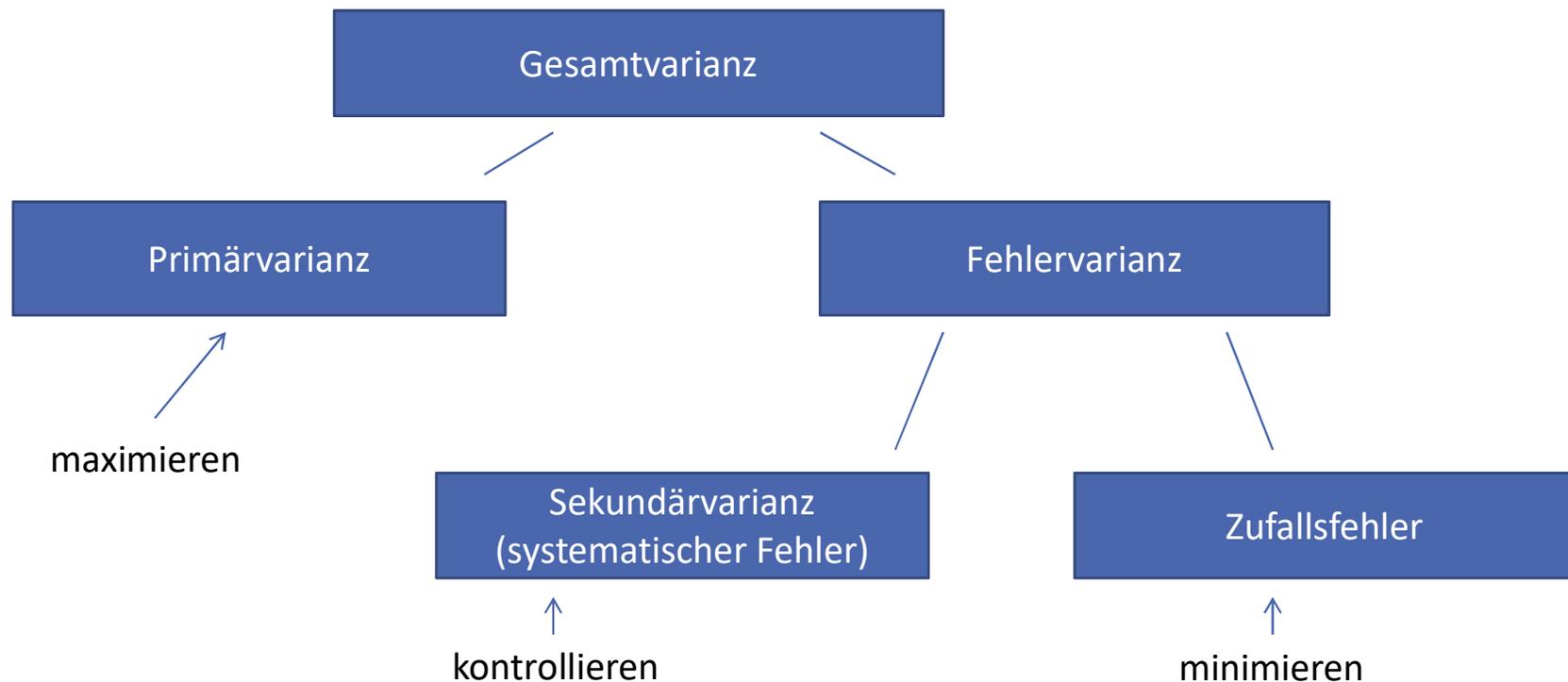
Fazit

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?

Termine

#	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 & Klausurvorbereitung	5.
12	12. Jul	Ethik	8.

Klausurvorbereitung

Organisatorisches

- 24 Fragen
- 5 Antwortmöglichkeiten, davon immer nur eine richtig
- Empfehlung: Forschungsmethoden-Teil am Stück durchführen
- Fragen: FAQ auf unserer Website

Inhalte

- Durchführung der Klausur in Ilias
- Vorlesungsinhalte
- prüfungsrelevant sind sowohl die Inhalte der Folien als auch die mündlich vermittelten Inhalte
- keine Inhalte von FoMe auf dem Klausurzettel

Erhebungstechniken

Ziel

Welche Möglichkeiten zur Datenerhebung gibt es?

Was sind Vor- und Nachteile der jeweiligen Methode?

Modulhandluch:

7. Grundlegendes **Wissen** um Aufbau und Struktur von psychologischen **Instrumenten, Untersuchungsanordnungen** und **Vorgehensweisen**

9. **Datenerhebung** und **Verfahren**

10. **Registrierung** und **Messung**

Ordnungsmöglichkeiten

- Einzelerhebung vs Gruppenerhebung
- Analyseeinheit
- Reaktive vs non-reaktive Verfahren
- Transparente vs intransparente Verfahren
- Teilnahme vs Nicht-Teilnahme des Forschenden

Übersicht: Erhebungsinstrumente

1. Selbstberichtsverfahren
 - Interview
 - Ratingverfahren/ Fragebogenverfahren
2. Beobachtung
3. Psychologische Tests
4. Biopsychologische Verfahren

Selbstberichtsverfahren

Wie gelangt man zu Selbstauskünften ?

1. Interpretation der Frage
2. Bildung eines Urteils
3. Übersetzen in eine kommunizierte Auskunft

Selbstberichtsverfahren

Was kann sich auf den Prozess auswirken?

- Interpretationsspielraum der Frage
- Suggestivfragen
- Verzerrungen durch Erinnerungseffekte
- Antwortmöglichkeiten
- Antworttendenzen

Befragung

- Schriftliche vs. mündliche Befragung
- Standardisierte vs. nicht-standardisierte Befragung
- offene vs. geschlossene Fragen
- Strukturierte vs. unstrukturierte Befragung
- Einzelbefragung vs. Gruppenbefragung vs. Umfrage

Mündliche Befragung - Interview

- häufig weniger standardisiert als schriftliche Befragungen
- oft höhere Reaktivität als bei schriftlichen Befragungen
- stärkere Gefahr von VL-Effekten
- Nachteil: oft aufwändig und teuer

Schriftliche Befragung

- Frage vs. Aussage vs. Aufforderung
- Offenes vs. geschlossenes Antwortformat
- Vorteile gegenüber Interview: oft effizienter und anonym
- Nachteil: Lese- und Schreibkenntnisse notwendig

Quantitative schriftliche Befragung - Ratingskalen

- am häufigsten genutztes Format schriftlicher Befragungen
- erlauben Quantifizierung psychologischer Merkmale
- häufig Intervallskalenniveau

Ratingskalen

- uni- und bipolare Skalen

Wie oft gehen Sie ins Kino?
Nie – selten – manchmal – oft – sehr oft

Wie empfinden Sie die Atmosphäre in der Situation?
Gespannt 1 2 3 4 5 Gelöst

- Marken

5-stufige Skala mit Emoticons (nach Jäger, 2004)



Ratingskalen

Wie viele Skalenstufen sollte man wählen?

- Diskriminationsfähigkeit der Vp?
- Forced choice oder ungerade Anzahl?

Tipps zur Formulierung von Fragen

- Vermeidung von Fremdwörtern und Fachbegriffen
- Vermeidung langer und sehr komplexer Sätze
- Formulierung an Zielgruppe anpassen
- keine doppelten Verneinungen
- Nur ein Aspekt pro Item abfragen
- keine Fragen, die keine Varianz erzeugen
- mehrere Items zur Beantwortung eines Aspekts nutzen
- Instruktion, die klar verständlich und informativ ist
- nicht zu viele verschiedene Skalen

Qualitative schriftliche Befragungen

- unstrukturiert → Aufsatz schreiben (nicht-standardisiert)
- Teilstrukturiert → teilstandardisierter Fragebogen

Schwierigkeiten bei Selbstberichtsverfahren

Verfälschbarkeit:

- Soziale Erwünschtheit
- Selbstdarstellung → besonders bei als relevant wahrgenommenen Situationen

Schwierigkeiten bei Selbstberichtsverfahren

Antworttendenzen

- Halo-Effekt
- Tendenz zur Mitte
- Ja-Sage-Tendenz
- Nein-Sage-Tendenz
- Primacy-Effekt

Vor- und Nachteile von Fragebogenmethoden

Nachteile

- Geringere Flexibilität der Antwortmöglichkeiten
- Verfälschungen im Gruppensetting schwerer aufdeckbar
- Abhängigkeit von Lese- und Schreibkompetenzen
- Beschränkung auf wenige, klar umschriebene Befragungsinhalte
- Reaktivität

Vorteile

- In Gruppen durchführbar
- Hohe Vergleichbarkeit von Ergebnissen
- Hohe Durchführungsobjektivität
- Hohe Auswertungsobjektivität
- Kein großer Schulungsaufwand zur Durchführung notwendig
- Anonymität

Beobachtung

Beobachtung

- stellt Merkmalsausprägungen oder Auftreten von Ereignissen fest
- wissenschaftlicher Erkenntnisgewinn als Ziel
- Beobachtung unter natürlichen Bedingungen vs. Laborbedingungen
- Selbst- vs. Fremdbeobachtung

Beobachtung

unsystematisch vs. Systematisch:

- unsystematisch: als erste Orientierung
- systematisch: stärker festgelegt

Beteiligungsgrad des Beobachters:

- teilnehmend (aktiv vs. passiv)
- nicht-teilnehmend

Beobachtung

Transparenz der Beobachtung:

- offen
- verdeckt

→ Wann wäre eine verdeckte Beobachtung sinnvoll?

Auswahl der Stichprobe:

- Zeitstichprobe
- Ereignisstichprobe

Beobachtung

	Zeitstichprobe	Ereignisstichprobe
Vorteile	Breiter/ umfassender Überblick über den Gesamtverlauf Dauer von Verhalten messbar	Gezielte Überprüfung von Zusammenhangshypothesen Gezielte Untersuchung ausgewählter Verhaltensweisen Auch relativ selten auftretende Ereignisse können erhoben werden
Nachteile	Aufwendige Kodierung Hohe Anforderung an Konzentrationsvermögen Unübersichtliche Ergebnisse Relativ selten auftretende Ereignisse werden vielleicht nicht erfasst	Keine Auskunft über Dauer

Quantitative Beobachtung

Formen:

- Erfassung einzelner Verhaltensweisen in Dauer, Häufigkeit oder Intensität
- Instrumente zur Beobachtung von Merkmalen und Verhaltensweisen mit höherem Komplexitätsgrad
- Beobachtung von Verhaltensspuren

Qualitative Beobachtung

- natürliches Umfeld
- Fokus auf sozialen Interaktionen
- Berücksichtigung der Prinzipien und Kriterien qualitativer Forschung
- Dokumentation und Auswertung der Beobachtungsprotokolle

Beobachtung (allgemein)

Vorteile	Nachteile
Möglich bei eingeschränkten Verbalisierungsfähigkeiten	zeitintensiv
Informationen über automatisierte und unbewusste Verhaltensweisen	kostenintensiv
Einsatz bei eingeschränkter Auskunftswilligkeit	Nicht alle Phänomene sind beobachtbar-> kann dort nicht eingesetzt werden
Non-reaktiv möglich	Geringere methodische Ausarbeitung
Abbildung von Zeitverlauf möglich	

Fazit

- Je nach Fragestellung ist der Einsatz von Beobachtung oder Selbstberichtsverfahren sinnvoller
- eine Kombination beider Verfahren ergänzt die Ergebnisse und hilft Fehler zu entdecken

Fragen

1. In welchem Zusammenhang fällt bei der qualitativen Beobachtung besonders viel Arbeit an, in welchem Zusammenhang bei der quantitativen Beobachtung?
2. Welche Varianten der Stichprobenziehung sind für quantitative Beobachtungsstudien relevant und wann werden sie eingesetzt?

Psychologische Tests

Psychologischer Test

- Erfassung individueller Merkmalsausprägung
- Items werden entsprechend Merkmalsausprägung beantwortet/gelöst
- besteht aus Testaufgaben und Manual

Psychologische Tests

- projektiv vs. psychometrisch
- eindimensional vs. mehrdimensional
- Abgrenzung zu Selbstberichten
 - Normierung
 - Teils andere Aufgabenstellung
 - Keine freie Veröffentlichung
 - Unterscheidung zwischen richtigen und falschen Antworten (bei Leistungstests)

Tests in den Anfängen der Psychologie

Francis Galton (1822-1911)

- Messung von Intelligenz über:
 - Reaktionszeiten
 - Unterscheidungsaufgaben
 - Gedächtnistests
- Klassifikation nach Normalverteilung

Anwendungsbereiche von Tests

Forschung

Praxis:

- Berufsberatung
- Betriebliche und institutionelle Personalauslese
- Verkehrspsychologie
- Bildungseinrichtungen
- Rehabilitation
- Rechtsprechung
- Klinische Diagnosestellung

Normorientiertes vs. kriteriumsorientiertes Testen

- normorientiert: Vergleich mit empirischen Normwerten
- kriteriumsorientiert: Vergleich mit einem Schwellenwert/Kriterium

Psychometrische Tests: Leistungstests

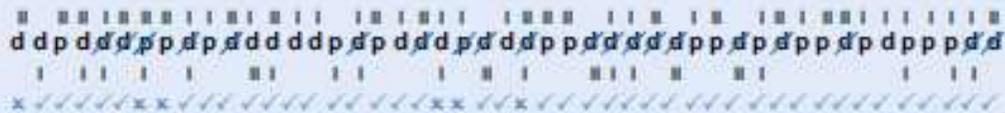
Speedtest

- Bearbeitungszeit zu knapp angesetzt
- Bsp. d2 (Brickenkamp, 2002)

Powertest

- Sukzessive Steigerung des Aufgabenniveaus
- Bsp. HAWIE (Tewes, 1991)

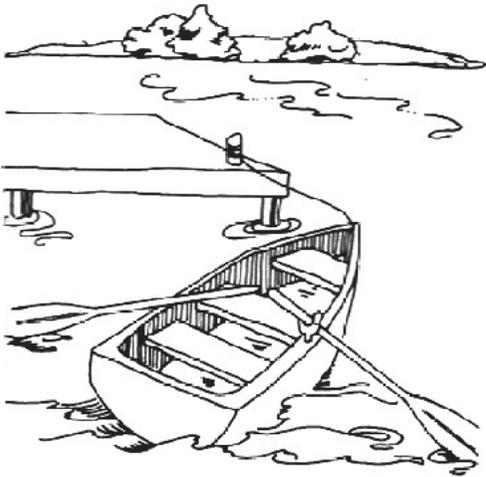
Leistungstests: Beispiel d2 (Brickenkamp, 2002)



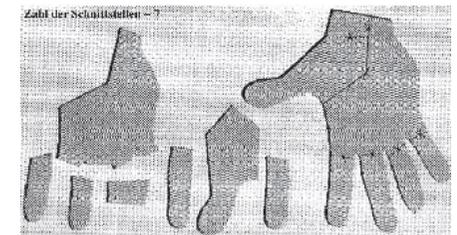
- Verfahren weist eine hohe Reliabilität und Validität auf
- Aber eine geringe externe Validität

Leistungstests: Beispiel HAWIE (Tewes, 1991)

1. Bilder ergänzen



2. Bilder ordnen



3. Mosaik-Test

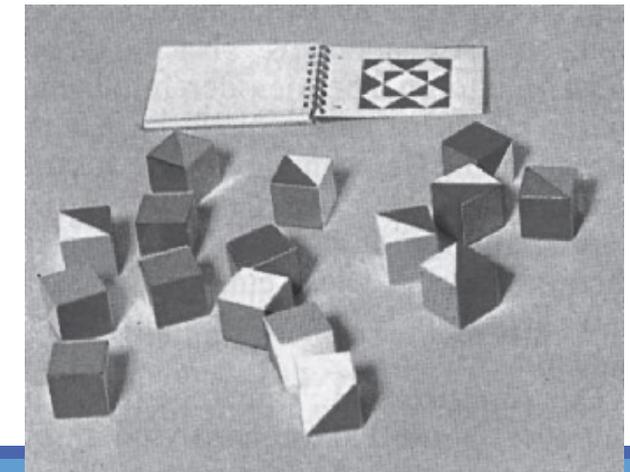
4. Figurenlegen

5. Zahlen-Symbol-Test

1	2	3	4	5	6	7	8	9
-	n	□	L	U	O	^	x	=

Beispiel

2	1	3	1	2	4	3	5



Leistungstests: Adaptives Testen

- Bearbeitung aller Items durch Getesteten ist redundant/ Erhebung irrelevanter Informationen
- Anpassen des Tests an die Fähigkeiten der Personen
- Tailored testing vs. branched testing

Psychometrische Tests: Persönlichkeitstests

- keine objektiven Leistungsmaße, sondern Beurteilung hinsichtlich Persönlichkeitsmerkmalen
→ typisches Verhalten und Erleben
- Selbstbeurteilung
- Bsp. FFI (Costa & Macrae, 1992)

Persönlichkeitstests: Beispiel FFI (Costa & Macrae, 1992)

Erfassung der Big Five

- Neurotizismus
- Extraversion
- Offenheit
- Gewissenhaftigkeit
- Verträglichkeit

5-stufige Ratingskala

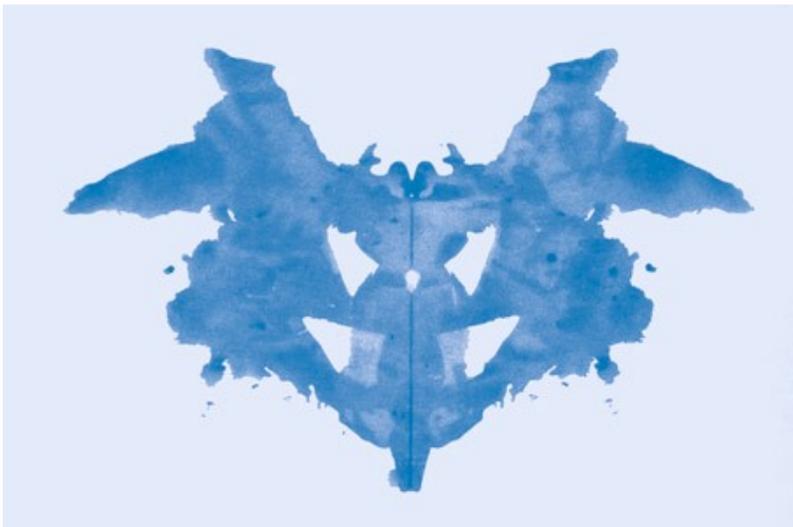
Testverfälschung

- raten
- soziale Erwünschtheit
- Simulation/ Dissimulation
- Antworttendenzen
- Selbsttäuschung

Vorgehen zur Vermeidung von Testverfälschung

- raten → geeignete Distraktoren, Ratekorrektur
- soziale Erwünschtheit → Anonymität, Kontrollskalen
- Simulation → Bogus Pipeline, Faking good/bad Instructions
- Verfälschung → Intransparenz der Untersuchung

Projektive Tests



Rorschach-Test (aus Bortz & Döring. (2016). *Forschungsmethoden und Evaluation für Human-und Sozialwissenschaftler*. S.451)

TAT (aus Bortz & Döring. (2016). *Forschungsmethoden und Evaluation für Human-und Sozialwissenschaftler*. S.452)



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, 8.6.4, 10.1, 10.2, 10.3, 10.4, 10.5.2, 10.5.4, 10.5.6