

Einführung in die Forschungsmethoden der Psychologie und Psychotherapie

Einheit 5: Messen in der Psychologie - latent

17.11.2023 | Dr. Caroline Zygar-Hoffmann

Heutige Themen

Selbstberichtsverfahren: Befragung und Rating

- Rahmenbedingungen
- Befragung
- Ratings

Testverfahren

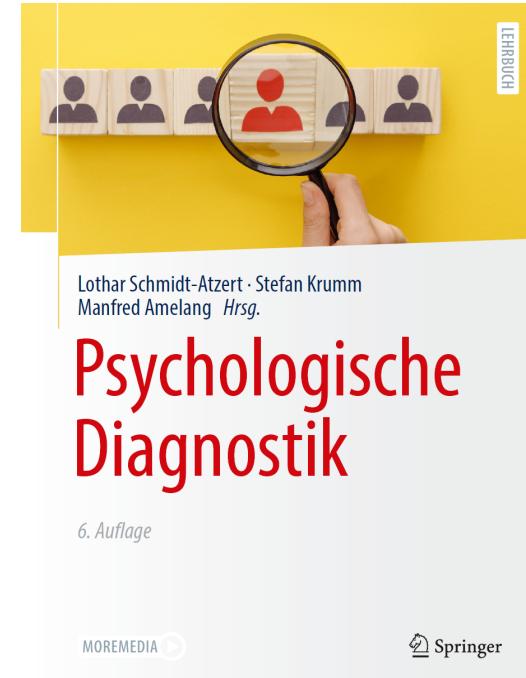
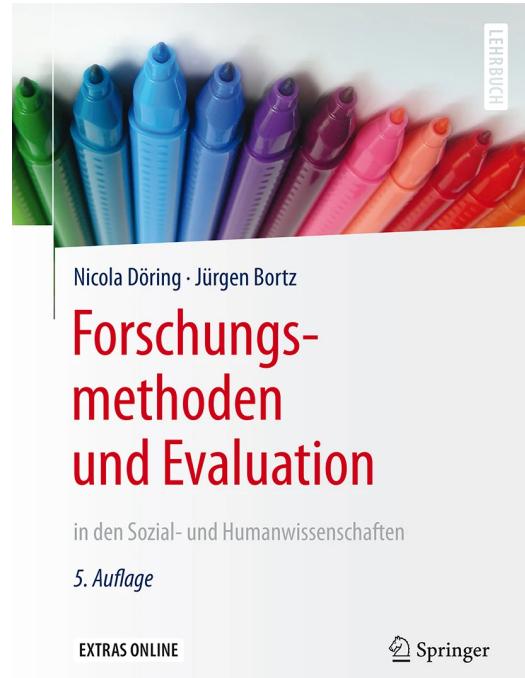
- Merkmale
- Leistungstests
- Persönlichkeitstests

Statistische Modellierung von Antwortprozessen

- Klassische Testtheorie
- Probabilistische Testtheorie

Take-Aways

Literatur für die heutige Sitzung



Kapitel 8.4.4. in Döring, N. & Bortz, J. (2016).
Forschungsmethoden und Evaluation in den Sozial- und
Humanwissenschaften. Pearson.

Kapitel 2.2.1. in Schmidt-Atzert, L., Krumm, S. & Amelang, M.
(2021). Psychologische Diagnostik. Springer.

Materialien: Vielen Dank an Prof. Dr. Stephan Goerigk für Bereitstellung der Grundlage für die Materialien

Selbstberichtsverfahren: Befragung und Rating

Rahmenbedingungen

Idee: Die Befragung ermöglicht einen Zugang zu psychischen Prozessen, Strukturen und Merkmalen, auch solchen die nicht direkt beobachtbar (manifest) sind, also nicht sichtbar (latent)

Drei Voraussetzungen:

1. Untersuchte Personen müssen **Zugang** zu den interessierenden psychischen Prozessen haben
2. Untersuchte Personen müssen interessierende psychische Prozesse **kommunizieren** können
3. Selbstauskünfte von Untersuchungsteilnehmer:innen müssen hinreichend **reliabel und valide** sein

Selbstauskünfte in Befragungen beinhalten **drei elementare kognitive bzw. mentale Prozesse**:

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

Selbstberichtsverfahren: Befragung und Rating

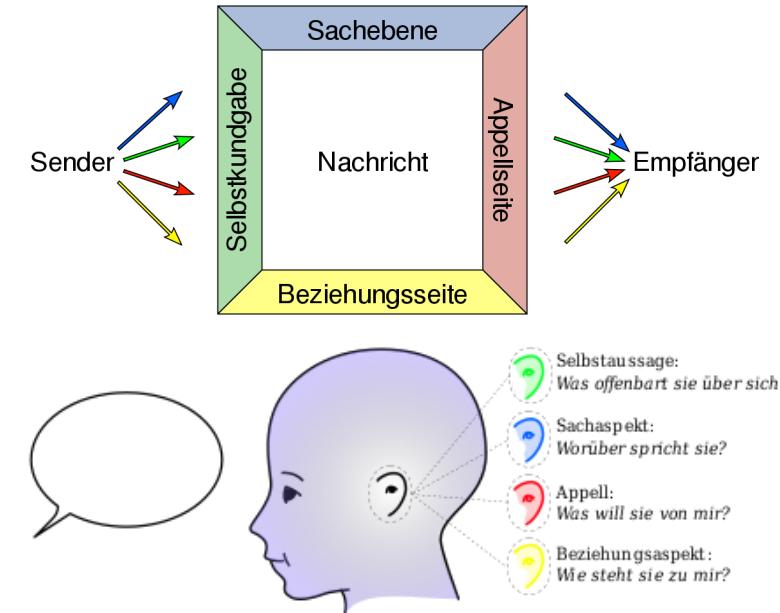
Rahmenbedingungen

Wie werden Selbstberichte kommuniziert?

- Selbstbericht = Akt der Kommunikation
- Person teilt sich bewusst mit (wendet sich an einen oder mehrere Adressaten)
- Menschliche Kommunikation = Informationen + Absicht des Senders

→ Einfluss auf die Validität: Was messe ich eigentlich? (z.B. ist ein Rückschluss auf das Merkmal basierend auf der gegebenen Information ggf. verzerrt, wenn sozial erwünschter geantwortet wird)

→ Beispiel 4-Seiten-Modell (F. Schulz von Thun): Oft ist man als Forscher daran interessiert, die Absicht des Senders auf die "Selbstkundgabe" zu richten



Selbstberichtsverfahren: Befragung und Rating

Befragung - Varianten und Prinzipien der Konstruktion

- **Schriftliche** (Fragebogen) vs. **mündliche** Befragung (Interview)
- **Standardisierte** vs. **nichtstandardisierte** Befragung → relevant für Auswertungs- und Interpretationsobjektivität
 - Standardisiert: Antwortmöglichkeiten vorgegeben (geschlossene Fragen)
 - Nichtstandardisiert: Befragte Person kann Antwort in ihren eigenen Worten formulieren (offene Fragen)
- **Strukturierte** vs. **unstrukturierte** Befragung → relevant für Durchführungsobjektivität
 - Strukturiert: Wortlaut und Reihenfolge der Fragen bzw. Items genau vorgegeben
 - Halbstrukturiert: Leitfaden mit vorformulierten Fragen zur Orientierung
 - Unstrukturiert: Keine Vorgaben hinsichtlich Frageformulierung und Reihenfolge der Items
- **Anzahl** der befragten Personen
 - Einzelbefragung (1 Person)
 - Gruppenbefragung (kleine Gruppe)
 - Survey (Umfrage mit großer Zahl an Befragten, wobei i.d.R. jede Person einzeln befragt wird)

Befragung - Varianten und Prinzipien der Konstruktion

Unterschiede zwischen schriftlicher und mündlicher Befragung

- Konstruktion schriftlicher Befragungen erfordert mehr Vorwissen (Interview = explorativer)
- Bei schriftlichen Befragungen ist das Reaktivitätsproblem geringer (der face-to-face-Kontakt bei Interviews kann Beeinflussungseffekte haben und Mitteilungsabsichten können stärker ausgeprägt sein)
- Befragte äußern sich bisweilen eher und ausführlicher in mündlichen Befragungen, Mitteilungsabsichten können erörtert werden
- Interviews i.d.R. aufwändiger und kostenintensiver

→ Entscheidung nach Forschungsziel und Ressourcenverfügbarkeit

Selbstberichtsverfahren: Befragung und Rating

Befragung - Varianten und Prinzipien der Konstruktion

Meist Formulierung des Items als Frage oder Aussage (manchmal auch einfach nur ein zu beurteilendes Adjektiv); zeitlicher Bezug kann variieren (z.B. "im Moment" vs. "im letzten Monat", vs. "im Allgemeinen")

Tipps zur Formulierung von Fragenbogenitems

Ein guter Fragebogen ist gekennzeichnet durch (Pelham und Blanton, 2007):

- einfache Formulierung und gute Verständlichkeit
- keine zu hohen Anforderungen an die mentale oder kognitive Leistungsfähigkeit der Befragten
- adressatenorientierte Formulierung
- keine (doppelten) Verneinungen in den Fragen
- keine überfrachteten Fragen
- keine »Forced Choice« Antworten bei unabhängig beantwortbaren Aspekten (z.B. "Sind Sie eher gewissenhaft oder offen?" → schließt sich nicht aus!)
- keine Fragen, die sehr ähnlich klingen
- Einsatz mehrerer Items zur Messung eines Konstrukt
- Beachtung der Ausgewogenheit in der Reihenfolge der Fragen
- eine klare und informative Instruktion

Selbstberichtsverfahren: Befragung und Rating

Ratings: Beurteilungen auf Skalen

Ratingskalen sind ein sehr häufiges Format in schriftlichen Befragungen:

- Beurteilung von Aspekten eines bestimmten Merkmals auf einer Skala
- Geben gleich große, georgnete, markierte Abschnitte des Merkmalskontinuums vor
- Personen sollen diejenige Stufe der Ratingskala ankreuzen, die persönlicher Auffassung am ehesten entspricht

Gründe für Beliebtheit in Psychologie:

1. liefern vergleichsweise direkt quantitative (in Zahlen übersetzte und damit statistisch auswertbare) Daten
2. unterteilen die Merkmalsausprägungen in gleich große Abschnitte → legen datenanalytisch günstiges Skalenniveau (Intervallskalenniveaus) nahe

Vorsicht: Intervallskalenniveau ist nicht genuin aus dem Format der Skala, sondern nur inhaltlich (psychologisch/empirisch) zu begründen (wird aber i.d.R. bei Ratingskalen akzeptiert)

Selbstberichtsverfahren: Befragung und Rating

Ratings: Beurteilungen auf Skalen

Aspekte bei der Konstruktion von Ratingskalen

Verwendung unipolarer oder bipolarer Endpunkte

Beispiel 1: Unipolare verbale Ratingskala

Wie oft gehen Sie in die Kirche?

- nie
- sehr selten
- selten
- gelegentlich
- oft
- sehr oft

In der Mode kehrt alles wieder.

- stimmt gar nicht
- stimmt wenig
- stimmt teils-teils
- stimmt ziemlich
- stimmt völlig

Die Gruppenatmosphäre in der ersten Videoaufzeichnung empfinde ich als

gespannt

1	2	3	4	5
---	---	---	---	---

 gelöst

Um die Polarisierung dieser bipolaren Skala besser zum Ausdruck zu bringen, können die Stufen auch in folgender Weise beziffert werden:

gespannt

-2	-1	0	+1	+2
----	----	---	----	----

 gelöst

Das folgende Beispiel verzichtet auf eine Bezifferung der 5-stufigen Skala und arbeitet mit Minus- und Plus-Symbolen.

gespannt

--	-	.	+	++
----	---	---	---	----

 gelöst

Ratings: Beurteilungen auf Skalen

Aspekte bei der Konstruktion von Ratingskalen

Verwendung unipolarer oder bipolarer Endpunkte

- Vorteil bipolare Skalen: Begriffe an den Endpunkten der Skala definieren einander wechselseitig
- Wann unipolare Skalen?

→ wenn der Gegensatz eines Begriffs unklar ist (z. B. bei "schüchtern"); das ist häufig der Fall, daher sind unipolare Skalen auch häufiger

→ zur Beurteilung von Merkmalen mit einem natürlichen Nullpunkt (z.B. bei Abfragen von Häufigkeiten)

Ratings: Beurteilungen auf Skalen

Aspekte bei der Konstruktion von Ratingskalen

Abstufung der Skala

1. Geringe vs. hohe Anzahl der Stufen

- Problem geringe Anzahl: relevante Unterschiede zwischen verschiedenen Beurteilungen nicht abbildbar
- Problem hohe Anzahl: Differenziertheit des Urteils nicht mehr möglich?
- In der Praxis i.d.R. 4- bis 9-stufige Ratingskalen, wobei es empirische Befunde gibt, dass die Reliabilität bei 5 (Revilla et al., 2014) bis 7 Stufen (Preston & Coleman, 2000) besonders gut ist

2. Gerade vs. ungerade Anzahl von Stufen

- Gerade Anzahl: Urteil in Richtung des einen oder anderen Pols der Skala wird erzwungen
- Ungerade Anzahl: Mittelkategorie vorhanden, d.h. neutrale Urteilmöglichkeit verfügbar
- Problem neutrales Urteil: nicht eindeutig interpretierbar (**Ambivalenz-Indifferenz-Problem**) → ggf. sinnvoll eine weitere Antwortoption neben der Mittelkategorie zur Angabe von Meinungslosigkeit zu ermöglichen (z.B. "ich weiß es nicht", "keine Meinung", "keine Angabe")

Selbstberichtsverfahren: Befragung und Rating

Ratings: Beurteilungen auf Skalen

Aspekte bei der Konstruktion von Ratingskalen

Bezeichnung der Abstufungen einer Skala

Zahlen (numerische Marker)

- Vorteil: Stufen sind eindeutig
- Vorteil: Abstände zwischen den Stufen der Ratingskala sind gleich

Wörter (verbale Marker) oder sprachfreie Zeichen (grafische Marker, z.B. Smileys oder Frownies)

- Vorteil: leichter verständlich (forschungssnaive Personen, Kinder)
- Nachteil: Verständnis variiert, gleicher Abstand zwischen den Stufen unklar

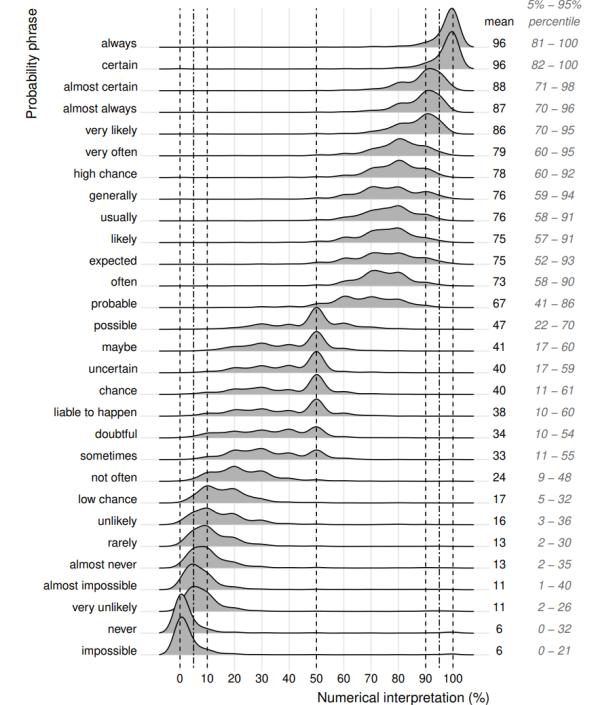


Figure 2. Density plots, mean values, and 5% and 95% percentiles of the numerical interpretations (in percentages) given by all participants for each phrase in the survey. Note that density plots are a smooth variant of histograms and may therefore be positive outside the data range of 0–100%.

Ratings: Beurteilungen auf Skalen

Probleme bei der Beantwortung von Ratingskalen

- **Antworttendenzen von Teilnehmer:innen**

- **Tendenz zur Mitte:** Extremurteile werden vermieden (insbesondere wenn Skalen an den Endpunkten nicht verankert sind, d.h. die Extreme unklar bleiben)
- **Extreme Antworttendenz:** Extremurteile werden bevorzugt
- Antworttendenzen sind besonders ausgeprägt, wenn Urteilsobjekte wenig bekannt sind
- Gegenmaßnahme: Untersuchungsteilnehmer:innen hinreichend über zu beurteilende Objekte informieren, alle Antwortkategorien verbal beschriften (Weijters et al., 2010; führt zu auch zu höherer Reliabilität: Saris & Gallhofer, 2007)

Selbstberichtsverfahren: Befragung und Rating

Ratings: Beurteilungen auf Skalen

Probleme bei der Beantwortung von Ratingskalen



Abb. 8.4 Ein weiterer Grund für eine Antworttendenz zur Mitte. (© Peter Roy / Search ID: pron128, Rights Available from CartoonStock.com)

Döring & Bortz (2016), S.254

Ratings: Beurteilungen auf Skalen

Probleme bei der Beantwortung von Ratingskalen

- **Gedankenlose Reproduktion**

- bei ähnlich erscheinenden Items nach erster Antwort bei folgenden Fragen den selben Wert angeben
- Stellt Validität der Antworten in Frage
- Gegenmaßnahme: Mischung von Fragen
- Umpolung der Fragerichtung als alternative Gegenmaßnahme ist umstritten (Swain et al., 2008; Weijters & Baumgartner, 2012)

Merkmale Psychologischer Test

- **Definition:** wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer empirisch unterscheidbarer Merkmale
- **Ziel:** möglichst genaue quantitative Aussage über den relativen Grad der individuellen Merkmalsausprägung, v.a. zum Einsatz in psychologischer Diagnostik
- **Besteht in der Regel aus mehreren Aufgaben oder Fragen (Items)** → kann also auch ein Fragebogen sein ("Persönlichkeitstests"; ein Fragebogen wird meist erst dann als Test bezeichnet, wenn er in der Einzelfalldiagnostik eingesetzt werden kann und nicht nur in der Forschung, z.B. weil die Gütekriterien entsprechend gut sind und Normen vorliegen)
- Aufgaben/Fragen werden von Menschen mit unterschiedlichen Fähigkeiten oder Eigenschaften unterschiedlich gelöst/beantwortet
- Test = standardisierte Verhaltensstichprobe
- Testwert = Aggregation über die Einzelitems/-aufgaben des Tests

Leistungstests

- Leistungstests erfassen Merkmale im Hinblick auf einen objektiven Maßstab zur Beurteilung der Güte der Antworten
- Antworten können also »richtig« oder »falsch« sein
- Für Beantwortung sind kognitive Prozesse im Testverfahren Sinne und Leistungsmotivation entscheidend
- Um zwischen guten/schlechten Leistungen differenzieren zu können, müssen verschiedene Schwierigkeitsgrade realisiert sein

Beispiele:

- Intelligenztests
- Eignungstests

Leistungstests

Speed-Tests

- Bei Speed-Tests ist die Bearbeitungszeit zu knapp angesetzt
- Somit können in der Regel nicht alle Aufgaben bearbeitet werden

Beispiel: d2-Aufmerksamkeits-Konzentrationstest von Brickenkamp (2002)

- Testpersonen bearbeiten eine große Zahl von Zeichen, nämlich die Buchstaben d und p, die jeweils mit bis zu vier kleinen Strichen umgeben sind
- Anzustreichen sind nur jene d, die zwei Striche oberhalb aufweisen

1. b d b d b d b d b d b b b d d d d b d d q d b q d
2. d b d b b d b q b b d b d q d b d b d d d d b d
3. b d d d d q q d d d d b d d d d d q d d d b b d d
4. d d d d d q d q d d d b d d b q b d d b q d d d b
5. q d d d d b d d d q b d d b b d d q d d d q b d b
6. d d d b d b d d d b d d b q d d b d d q d d b

Leistungstests

Power-Tests

- Bei Power-Tests wird das Niveau der Aufgaben sukzessive gesteigert
- Adaptives Testverfahren, bis maximales Leistungsniveau gefunden ist

Beispiel: Wechsler-Intelligenz-Tests (HAWIE; seit 2013 WAIS-IV)

- HAWIE umfasst insgesamt 11 Untertests (Subskalen)
- Untertests sind Bereich der verbalen Intelligenz oder der Handlungsintelligenz zugeordnet

Leistungstests

Power-Tests

Beispiel: Wechsler-Intelligenz-Tests (HAWIE)

Verbalteil:

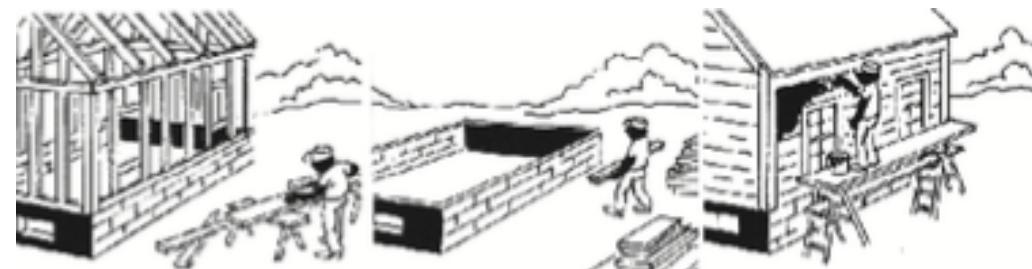
1. Allgemeines Wissen, z.B. »Was ist der Koran?«, »Wer erfand das Flugzeug?«
2. Zahlnachsprechen: Folgen von 3-9 Ziffern sind vorwärts und rückwärts nachzusprechen, z. B. »5-8-2« oder »4-2-7-3-1-8-2«
3. Wortschatz: »Was ist die Bedeutung von ...?« z. B. »anonym, Prestige, konkordant«
4. Rechnerisches Denken, z. B. »Ein Zug fährt 275 km in 5h. Wie groß ist seine Durchschnittsgeschwindigkeit in km/h?«
5. Allgemeines Verständnis, z. B. »Was bedeutet das Sprichwort, Stille Wasser sind tief?«

Leistungstests

Power-Tests

Beispiel: Wechsler-Intelligenz-Tests (HAWIE)

Bilder ordnen: »Ordnen Sie die Bilder bitte so, dass sich die sinnvollste Geschichte ergibt!«, z. B.

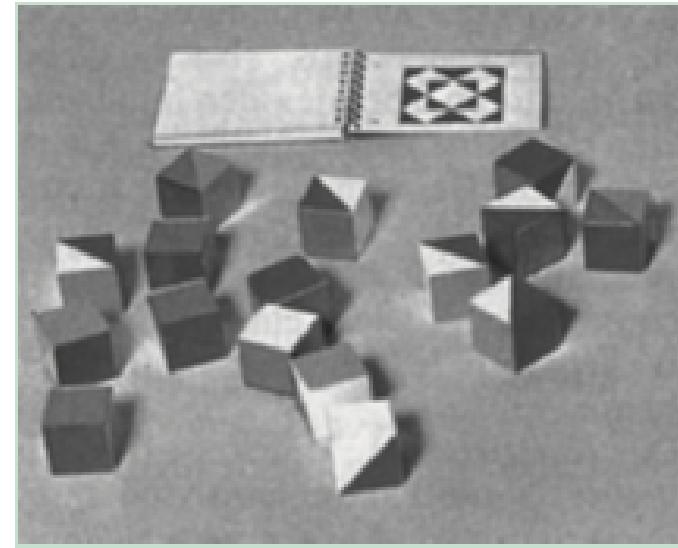


Leistungstests

Power-Tests

Beispiel: Wechsler-Intelligenz-Tests (HAWIE)

Mosaik-Test: »Legen Sie die Würfel so zusammen, dass sie ein Muster zeigen, wie das auf der Karte«, z. B.

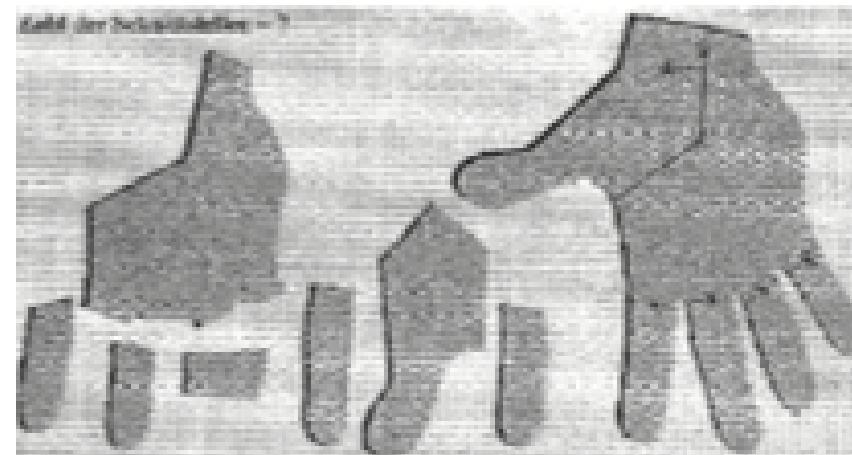


Leistungstests

Power-Tests

Beispiel: Wechsler-Intelligenz-Tests (HAWIE)

Figurenlegen: »Setzen Sie die Teile so zusammen, dass sie etwas darstellen!«, z. B.



Persönlichkeitstests

- Persönlichkeitstests liefern Daten im Hinblick auf emotional, motivational und sozial relevante Persönlichkeitseigenschaften
- Es wird zwischen subjektiven und objektiven Persönlichkeitstests unterschieden, je nachdem ob sie auf Selbstauskunft basieren oder nicht
 - Bei subjektiven Persönlichkeitstests ist der Zweck des Tests für die getesteten Personen leicht durchschaubar
 - Bei objektiven Persönlichkeitstests wird versucht, den Zweck zu verschleiern, um so die Reaktivität der Datenerhebung zu minimieren und die Validität der Ergebnisse zu erhöhen

Persönlichkeitstests

Beispiele für subjektive Persönlichkeitstests:

- NEO Five Factor Inventory (NEO-FFI; Costa & Macrae, 1992) -
mit Variante als Fremdbericht über andere Personen
- Freiburger Persönlichkeitsinventar (FPI-R, Fahrenberg, Hampel & Selg, 1994)

Beispiele für objektive Persönlichkeitstests:

- Implizite Assoziationstests (IAT) → basieren auf Reaktionszeiten, Kritik an Validität
- Objektiver Leistungsmotivationstest (OLMT, Schmidt-Atzert, 2007) → Idee: Wer leistungsmotivierter ist, strengt sich bei den dargestellten Aufgaben mehr an, und legt eine weitere Strecke zurück (= drückt öfters die notwendige Taste)

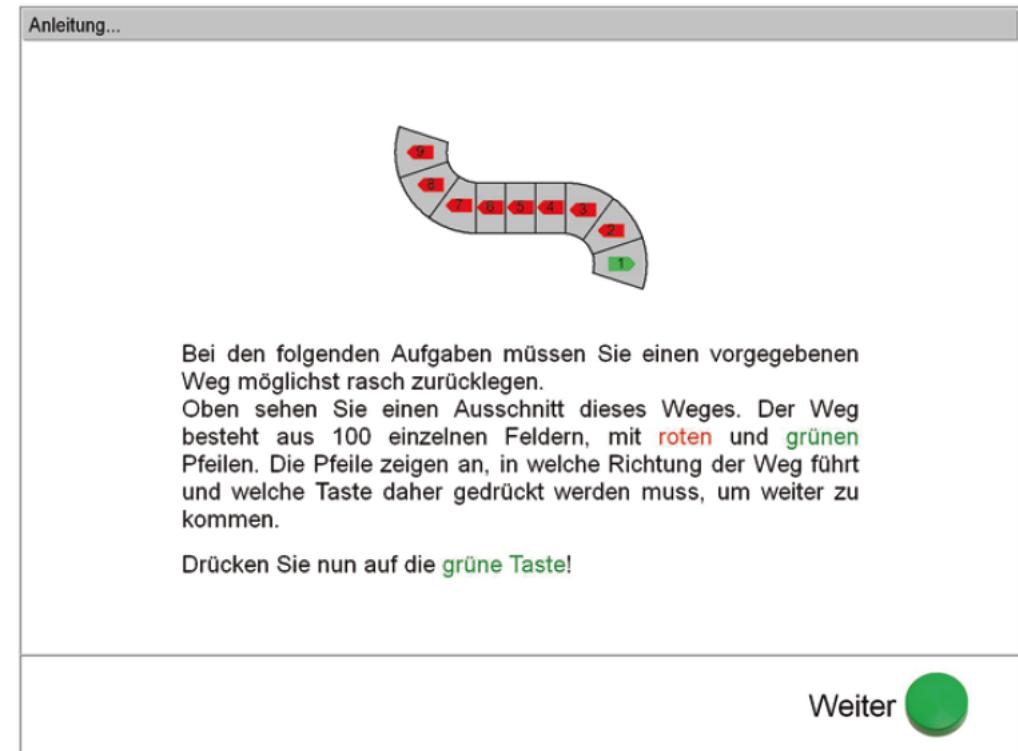


Abb. 3.22 Aufgabe im OLMT. (Aus Schmidt-Atzert 2007, mit freundlicher Genehmigung der SCHUHFRIED GmbH)

Persönlichkeitstests

Projektive Verfahren als Sonderform von objektiven Persönlichkeitstests

- Wie bisher: Standardisierte Testsituation
- Neu: Projektion auf diffuses Testmaterial
- Es ist nicht erforderlich, dass Personen eine explizite Repräsentation von dem zu messenden Konstrukt haben → geringere Verfälschbarkeit
- Große Variation in der Gültigkeit der Gütekriterien!! z.B. mangelhaft beim Rorschach-Test

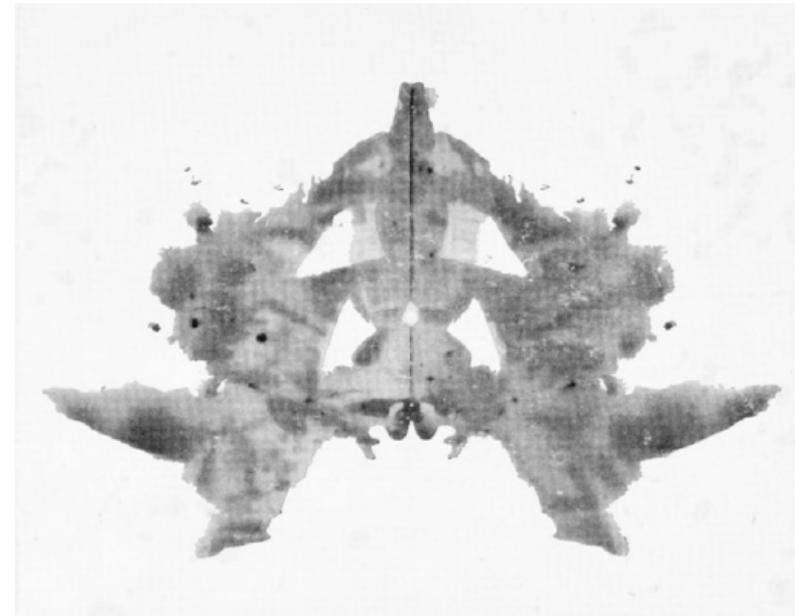


Abb. 3.24 Tafel aus dem Rorschach-Test. (Hermann Rorschach, Psychodiagnostik. Der Rorschach®-Test (Tafel I). © Verlag Hans Huber, Hogrefe AG, Bern/Switzerland)

Schmidt-Atzert et al. (2016), S.386

Persönlichkeitstests

Projektive Verfahren als Sonderform von objektiven Persönlichkeitstests

Beispiel: Picture Story Exercise (PSE) zur Messung von impliziten (nicht bewusst zugänglichen) Motiven *in der Forschung* (für diagnostische Praxis nicht reliabel genug)

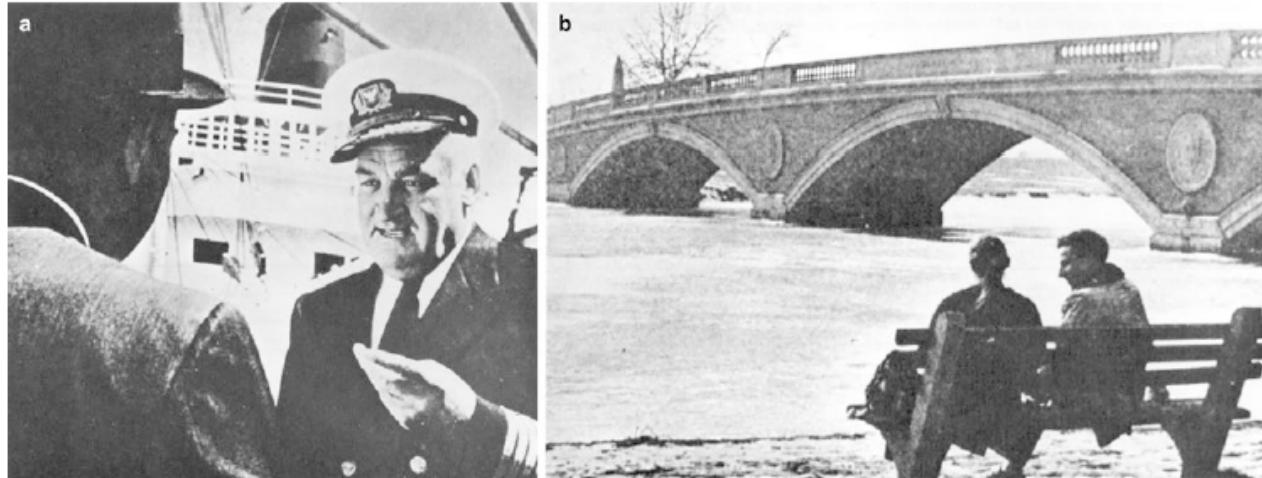


Abb. 3.27 Ausgewählte Bilder, bei denen deutsche Testpersonen ($n=428$) die höchsten Ausprägungen für ein Motiv zeigten (Kapitän: Macht; Paar vor Brücke: Affiliation). (Nach Pang 2010, © Oxford University Press)

Schmidt-Atzert et al. (2016), S.395

Klassische Testtheorie

- Die meisten Fragebögenskalen und Tests basieren auf der klassischen Testtheorie (daher werden Sie dazu bereits im Bachelor eine eigene Vorlesung hören)
- Zentrale Annahme der Klassischen Testtheorie: Jeder Wert einer Person auf einem konkreten Item ist aus zwei Komponenten zusammengesetzt

1. Wahrer Wert:

- mittlerer Testwert, den eine Person in einer unendlichen Serie von Testwiederholungen erzielen würde
- keine praktisch erzielbare, sondern eine theoretische Größe
- kann durch die konkrete empirische Antwort einer Person geschätzt werden

2. Fehleranteil (Messfehler):

- Abweichung dieses empirischen Schätzwerts vom wahren Wert

Klassische Testtheorie

- Anwendung auf Ratingskalen möglich (aber eigentlich für stetige Variablen entwickelt, z.B. Reaktionszeiten)
- Ziel: möglichst direkte und präzise Schätzung des wahren Werts auf einem Item, um darauf basierend einen Rückschluss auf eine latente Variable machen zu können
- Durch den Einsatz mehrerer Testitems soll der Fehleranteil insgesamt minimiert werden
- Anders ausgedrückt: Mehrere Items ermöglichen eine bessere Annäherung an die latente Variable einer Person
- klassische Testtheorie setzt somit voraus, dass wahre Werte und Fehlerwerte (und ihre jeweiligen Varianzen in Stichproben) getrennt bestimmt werden können → wertvoll zur Bestimmung der Reliabilität: Setzt die geschätzte Varianz der wahren Werte ins Verhältnis zur Gesamtvarianz der empirischen Schätzwerte

Klassische Testtheorie

Axiome (definitorische Festlegungen/Annahmen) der klassischen Testtheorie und die Folgerungen daraus

1. Ein **Testergebnis** setzt sich zusammen aus der **Summe von wahrem Wert** (Erwartungswert über unendlich viele Messungen) **und Messfehler** (z.B. Störeinflüsse der Umwelt).
2. **Der mittlere Messfehler ist gleich null** (→ Messfehler sind unsystematisch). Bei wiederholten Testanwendungen gleichen sich die verschiedenen Messfehler sozusagen aus.
3. **Wahre Werte und Messfehler eines Items/Tests sind nicht miteinander korreliert** (voneinander unabhängig) → Es werden nicht in bestimmten Ausprägungsbereichen des Items/Merkals mehr oder weniger Messfehler gemacht.
4. **Messfehler in einem Test sind nicht mit den wahren Werten in einem anderen Test korreliert** → Die Messfehler hängen nicht von bestimmten Eigenschaften ab.
5. **Die Messfehler aus verschiedenen Tests sind nicht miteinander korreliert** (voneinander unabhängig).

→ Die KTT berücksichtigt nur unsystematische Messfehler. Es kann jedoch auch systematische Messfehler geben, z.B. extremer Antwortstil, sozial erwünschte Antworten, mangelnde Motivation bei der Bearbeitung des Tests, ...

Probabilistische Testtheorie aka. Item-Response-Theorie (IRT, Embretson & Reise, 2000; Rasch, 1980)

Vergleich:

- **Klassische Testtheorie:**

- Schätzung des wahren Werts unter Berücksichtigung des Messfehlers
- Es existieren verschiedene testtheoretische Modelle je nach Annahme wie sich die Items in Hinblick auf die Messung der latenten Variable unterscheiden

- **Probabilistischen Testtheorie:**

- Schätzung der Wahrscheinlichkeit, mit der eine Person mit einer bestimmten Merkmalsausprägung ein Item auf eine bestimmte Art beantwortet (bzw. ein Item löst)
- Wahrscheinlichkeit, dass eine Person ein Testitem zur Erfassung des latenten Merkmals X positiv beantwortet, hängt von der Ausprägung von X bei dieser Person ab
- Es existieren verschiedene testtheoretische Modelle je nach Skalenniveau der Antworten (d.h. auch für dichotome Antworten einsetzbar)
- Ausführliche Behandlung im Masterstudium

Probabilistische Testtheorie

Folgen der Annahmen der probabilistischen Testtheorie:

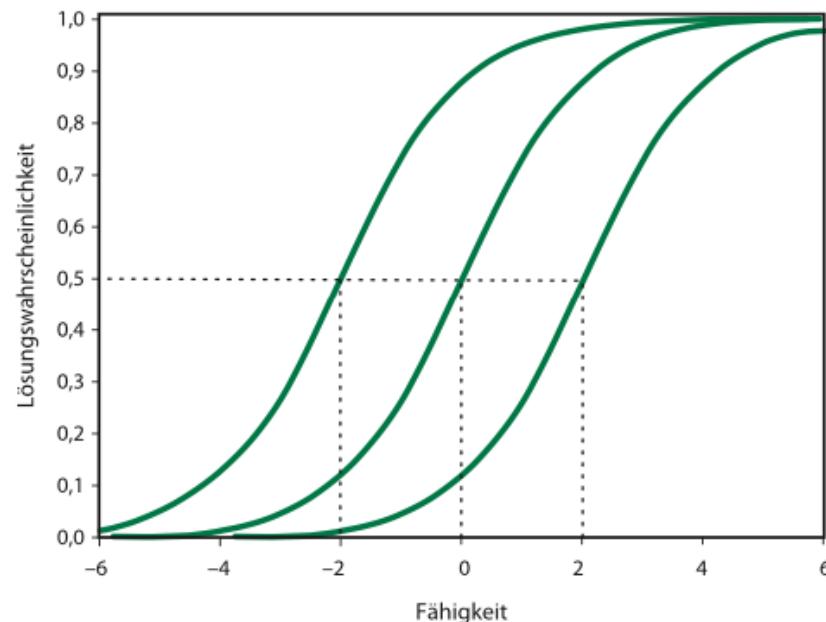
1. Eine Person mit höherer Fähigkeit löst ein geeignetes Item in einem Test mit höherer Wahrscheinlichkeit als eine Person mit niedrigerer Fähigkeit
2. Eine Person löst mit größerer Wahrscheinlichkeit ein Testitem, das von vielen Personen gelöst wird, als ein Testitem, das nur von wenigen Personen gelöst wird.

Quantifizierung von Antwortverhalten in der IRT:

- Beziehung zwischen dem Antwortverhalten (der Lösungswahrscheinlichkeit) und dem latenten Merkmal (der Fähigkeit) wird statistisch bestimmt
- Detailliertere Abbildung des Antwortprozesses möglich als bei der KTT (z.B. Berücksichtigung von Ratewahrscheinlichkeiten)
- Graphische Darstellung in sog. Item-Characteristic-Curves (ICC)

Probabilistische Testtheorie

Item-Characteristic-Curves (ICC):



- Antwortverhalten (der Lösungswahrscheinlichkeit) auf der y-Achse
- Ausprägung des latenten Merkmals (der Fähigkeit) auf der x-Achse
- pro Item eine Kurve (hier: 3 Items abgebildet)

Take-Aways

- Selbstratings sind in der Psychologie beliebte, kostengünstige Messverfahren, haben aber Probleme hinsichtlich der Verfälschbarkeit
- Die Itemformulierung und die Antwortskala von Fragebögen beeinflusst die Gütekriterien des Fragebogens (insbesondere Reliabilität und Validität)
- Ein psychologischer Test ist ein wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer empirisch unterscheidbarer Merkmale
- Leistungstests erfassen Merkmale im Hinblick auf einen objektiven Maßstab
- Klassische Testtheorie: Schätzung des wahren Werts unter Berücksichtigung des Messfehlers
- Probabilistischen Testtheorie: Schätzung der Wahrscheinlichkeit für eine bestimmte Itemantwort

[zurück zur heutigen Übersicht der Vorlesung →](#)

[zum Quiz zur Wissensprüfung →](#)

Einladung der Fachgruppe für Sozialpsychologie zur Paneldiskussion zum Nahostkonflikt

The executive board of the German Psychological Society recently wrote: "Many members of our professional society have a long tradition of cooperative projects with Israeli scientists. These cooperations are characterized by common values such as peace, understanding, democracy and freedom of science and many friendships have developed from them."

This is also true for the Social Psychology section of the German Psychological Society. There are many collaborations between German and Israeli social psychologists, and several of them deal with intergroup conflict, and the Israeli-Palestinian conflict in particular.

You are cordially invited to the online discussion between German and Israeli social psychologists Halabi, Hirschberger, Imhoff, Knab and Shnabel, on the consequences of October 7.

The webinar is organized by the Social Psychology section of the German Psychological Society and takes place on 21. November, 7:00 – 8:30 pm (Berlin/Zurich). Please find the link to the event here: [https://uzh.zoom.us/j/69562851420?
pwd=Um5xK0JINmdYMWs2QXYyczB0SjIEUT09](https://uzh.zoom.us/j/69562851420?pwd=Um5xK0JINmdYMWs2QXYyczB0SjIEUT09)

We ask you to share the invitation responsibly with interested non-members (e.g., students).

With kind regards Johannes Ullrich (Sprecher, Fachgruppe Sozialpsychologie)

Einladung der Fachgruppe für Sozialpsychologie zur Paneldiskussion zum Nahostkonflikt

Information about the panelists:

Slieman Halabi Slieman is a post-doctoral researcher at the University of Wuppertal. He earned his degree from Friedrich Schiller University of Jena. His research primarily focuses on "in-between groups", people who hold conflicting social identities and how they navigate their relations with the multiple groups they belong to. Besides being a researcher, Slieman has been facilitating dialogue groups between Palestinians and Israelis.

Gilad Hirschberger Gilad is a professor of social and political psychology at Reichman University, Israel, associate dean of the Baruch Ivcher School of Psychology. His work focuses on collective existential threats, and on how threat perceptions influence and shape political cognitions. Using group survival theory, he conducts research on conflict resolution and the prospects of reaching a sustainable agreement between Israel and the Palestinians.

Roland Imhoff Roland holds the chair for social and legal psychology at the University of Mainz, after receiving his PhD from the university of Bonn in 2010 and serving as assistant Professor at the Social Cognition Center Cologne. His research interests include representations of history, stereotyping and categorization, antisemitism and conspiracy beliefs. Before becoming a social psychologist, he facilitated exchange programs of antiracist NGOs between Israel and the Cologne-based Jugendclub Courage e.V..

Einladung der Fachgruppe für Sozialpsychologie zur Paneldiskussion zum Nahostkonflikt

Information about the panelists:

Nadine Knab Nadine is a post-doctoral researcher and lecturer at Tel Aviv University. She received her PhD from University Kaiserslautern-Landau and investigates psychological determinants and tools to transform intergroup conflict. Nadine is the chair of the German Peace Psychology Association and co-editor of the Handbook for Peace Psychology. She gained experience as a trainer for democracy building capacities, co-founded several NGOs to establish positive intergroup relations and served as a consultant for human rights organizations.

Nurit Shnabel Nurit is a professor of psychology at Tel Aviv University, where she received her PhD in 2008, and the current president of EASP (the European Association of Social Psychology). Her research interests include interpersonal and intergroup reconciliation, and subtle mechanisms that reinforce traditional gender roles. She is also trained as a facilitator of dialogue groups between Jews and Arabs in Israel.

Podcast-Empfehlung zur Hypnose als Methode in der Therapie (hier: nicht aus psychodynamischen Blickwinkel)

Es spricht Dr. phil. Barbara Schmidt (<https://www.uniklinikum-jena.de/mpsy/TWI0YXJiZWI0ZXI%3D.html>)

Zum Podcast: <https://www.miriamvanlijnden.de/podcast/folge19>