

Einführung in die Forschungsmethoden der Psychologie und Psychotherapie

Einheit 4: Latente Merkmale: Befragungen, Tests

02.05.2024 | Dr. Caroline Zygar-Hoffmann

Heutige Themen

Selbstberichtsverfahren: Befragung und Rating

- Rahmenbedingungen
- Befragung
- Ratings

Testverfahren

- Merkmale
- Leistungstests
- Persönlichkeitstests

Take-Aways und Schlüssel-/Fachbegriffe

- Take-Aways
- Schlüssel-/Fachbegriffe

Selbstberichtsverfahren: Befragung und Rating

Rahmenbedingungen

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

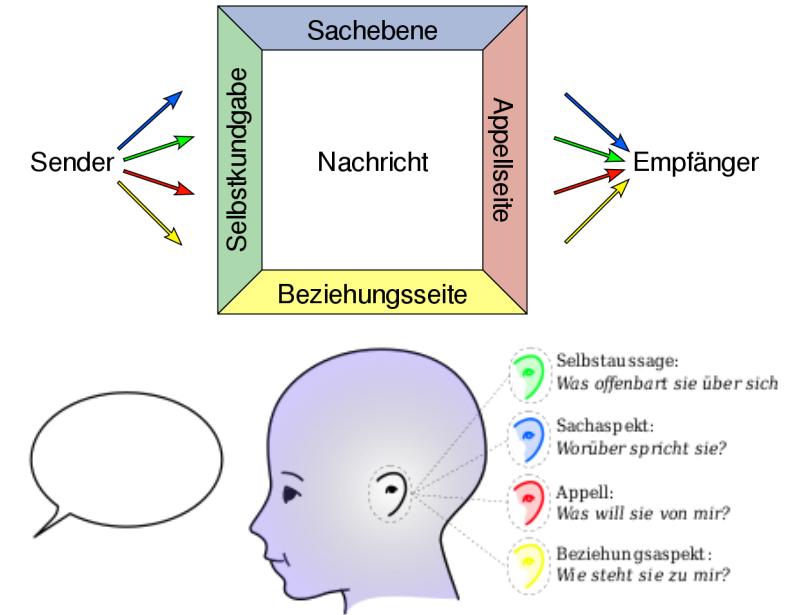
Rahmenbedingungen

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" und/oder "Sachebene" 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
 - (Voll-)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

- 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, dafür ggf. weniger Vorwissen nötig (zumindest wenn Interviews weniger standardisiert sind)

→ Entscheidung nach Forschungsziel und Ressourcenverfügbarkeit

Selbstberichtsverfahren: Befragung und Rating

Befragung - Varianten und Prinzipien der Konstruktion

Beispielthema: Warum sind Liebesbeziehungen oft nicht von Dauer?

Beispiel: Vollstrukturierte und standardisierte Befragung

1. Was kennzeichnet für Sie eine dauerhafte, funktionierende Beziehung?

- Offenheit Freiräume Verlässlichkeit gemeinsame Interessen Kompromissbereitschaft

2. Woran scheitern Ihrer Meinung nach viele Beziehungen?

- Untreue zu wenig Kommunikation Monotonie Erwartungsverletzungen Verwandtschaft

3. ...

Selbstberichtsverfahren: Befragung und Rating

Befragung - Varianten und Prinzipien der Konstruktion

Beispielthema: Warum sind Liebesbeziehungen oft nicht von Dauer?

Beispiel: Unstrukturiertes ("offenes") und unstandardisiertes Interview

Einstiegsfrage:

Denken Sie an ein Ihnen bekanntes Paar, das sich getrennt hat. Erzählen Sie mir, was Ihnen in der Zeit vor der Trennung aufgefallen ist (bzw. wie Sie sich in der Gegenwart des Paares gefühlt haben).

Potentielle Nachfragen (auch Verständnisfragen erlaubt):

- Beschreiben Sie eine Streitstuation des Ihnen bekannten Paares.
- Was hat Ihrer Meinung nach zur Trennung des Paares geführt?
- Was hätten die Betroffenen Ihrer Meinung nach tun können, um die Trennung zu verhindern?

→ in der qualitativen Forschung üblich, siehe Einheit 5 für mehr Details zu Interview

Befragung - Varianten und Prinzipien der Konstruktion

Standardisierte deutschsprachige Verfahren zur Diagnostik psychischer Störungen (Auswahl).

(Zit. nach Schneider & Margraf, 2002)

- Checklisten
 - IDCL: Internationale Diagnose-Checklisten für ICD-10 (Hiller et al., 1995)
 - IDCL-P: Internationale Diagnose-Checklisten für Persönlichkeitsstörungen (Bronisch et al., 1995)
 - AMDP: Manual zur Dokumentation psychiatrischer Befunde (Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie, 1997)
- Strukturierte Interviews
 - SKID-I: Strukturiertes Klinisches Interview für DSM-IV, Achse-I (Wittchen et al., 1997)
 - SKID-II: Strukturiertes Klinisches Interview für DSM-IV, Achse-II (Fydrich et al., 1997)
 - DIPS für DSM-IV-TR: Diagnostisches Interview bei psychischen Störungen für DSM-IV-TR (Schneider & Margraf, 2006)
 - Kinder-DIPS: Diagnostisches Interview bei psychischen Störungen im Kindes- und Jugendalter (Unnewehr et al., 1995; Schneider et al., 2009).
 - Mini-DIPS: Diagnostisches Kurzinterview bei psychischen Störungen (Margraf, 1994)
- Standardisierte Interviews
 - CIDI: Composite International Diagnostic Interview (Wittchen & Semler, 1990)
 - DIA-X: Diagnostisches Expertensystem für ICD-10 und DSM-IV (Wittchen & Pfister, 1997)

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")

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 (Intervallskalenniveau) 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)

Ratings: Beurteilungen auf Skalen

Aspekte bei der Konstruktion von Ratingsskalen

Verwendung unipolarer oder bipolarer Endpunkte

- **Unipolare Skalen:** beginnen mit einem Nullpunkt (z.B. „nie“, „stimme überhaupt nicht zu“) und nehmen dann in der Ausprägung in eine Richtung zu (z.B. bis „immer“, „stimme voll und ganz zu“).
- **Bipolare Skalen:** bezeichnen in ihren Enden den negativen und positiven Pol eines Kontinuums (z.B. „lehne ab“ vs. „stimme zu“) oder gegensätzliche Eigenschaften eines Merkmals (z.B. „weich“ vs. „hart“)
- **Maßnahme:** Darbietung einer Nummerierung, die zu den verbalen Bezeichnungen passt, wirkt sich positiv auf die Reliabilität der Messungen aus (Rammstedt & Krebs, 2007).

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

Bipolare Skalen:

- Vorteil: Begriffe an den Endpunkten der Skala definieren einander wechselseitig
- Vorsicht: Nicht angemessen, um zwei distinkte Merkmale mit einem Item zu erfassen

Unipolare Skalen:

- Vorteil: Gegensatz eines Begriffs muss nicht klar sein (z.B. bei "schüchtern"); das ist häufig der Fall, daher sind unipolare Skalen auch häufiger
- Eignen sich gut 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? (Gegenposition: Review von Lewis & Sauro, 2021)
- 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

Ratings: Beurteilungen auf Skalen

Aspekte bei der Konstruktion von Ratingsskalen

Abstufung der Skala

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 Urteilsmöglichkeit 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

Frage 1) Wie oft (verbal beschrieben) haben Sie in der letzten Woche Fragen gestellt oder zu einer Diskussion auf andere Art beigetragen? Auswahl: Nie, manchmal, oft, sehr oft

Frage 2) Wie oft (absolute Häufigkeiten) haben Sie in der letzten Woche Fragen gestellt oder zu einer Diskussion auf andere Art beigetragen?

Selbstberichtsverfahren: Befragung und Rating

Ratings: Beurteilungen auf Skalen

Aspekte bei der Konstruktion von Ratingskalen

Bezeichnung der Abstufungen einer Skala

Table 2 Meaning of vague quantifier (i.e., median times per week) by item

Item	Times per week Median response (median absolute deviation)			
	Never	Sometimes	Often	Very often
Asked questions or contributed to course discussions in other ways	0 (0)	2.0 (1.3)	4.0 (2.0)	10.0 (7.0)
Explained course material to one or more students	0 (0)	1.0 (0.8)	3.0 (2.0)	5.0 (3.0)
Prepared for exams by discussing or working through course material with other students	0 (0)	0.5 (0.4)	2.0 (1.1)	3.0 (2.0)
Worked with other students on course projects or assignments	0 (0)	0.3 (0.2)	1.0 (0.9)	2.0 (1.3)
Talked about career plans with a faculty member	0 (0)	0.1 (0.1)	0.5 (0.4)	0.9 (0.6)
Worked with a faculty member on activities other than coursework (committees, student groups, etc.)	0 (0)	0.2 (0.1)	0.9 (0.7)	1.0 (0.9)
Discussed course topics, ideas, or concepts with a faculty member outside of class	0 (0)	0.2 (0.1)	1.0 (0.8)	2.0 (1.0)
Discussed your academic performance with a faculty member	0 (0)	0.2 (0.1)	0.7 (0.4)	1.0 (0.5)
Reached conclusions based on your own analysis of numerical information (numbers, graphs, statistics, etc.)	0 (0)	0.4 (0.3)	2.0 (1.3)	3.0 (2.0)
Used numerical information to examine a real-world problem or issue (unemployment, climate change, public health, etc.)	0 (0)	0.5 (0.4)	1.2 (0.9)	2.3 (1.7)
Reviewed your notes after class	0 (0)	1.2 (0.8)	4.0 (1.0)	5.0 (2.0)

Selbstberichtsverfahren: Befragung und Rating

Ratings: Beurteilungen auf Skalen

Aspekte bei der Konstruktion von Ratingskalen

Bezeichnung der Abstufungen einer Skala

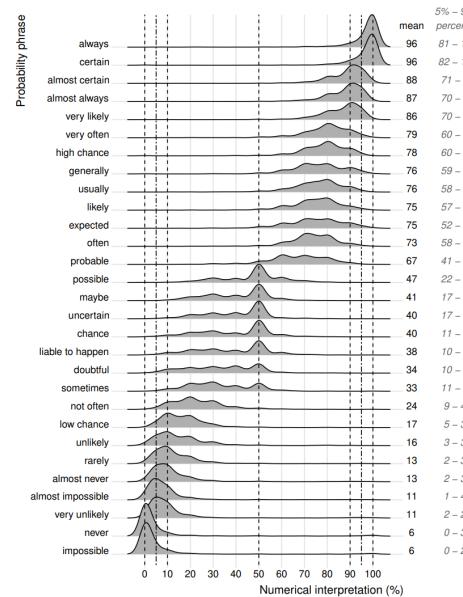
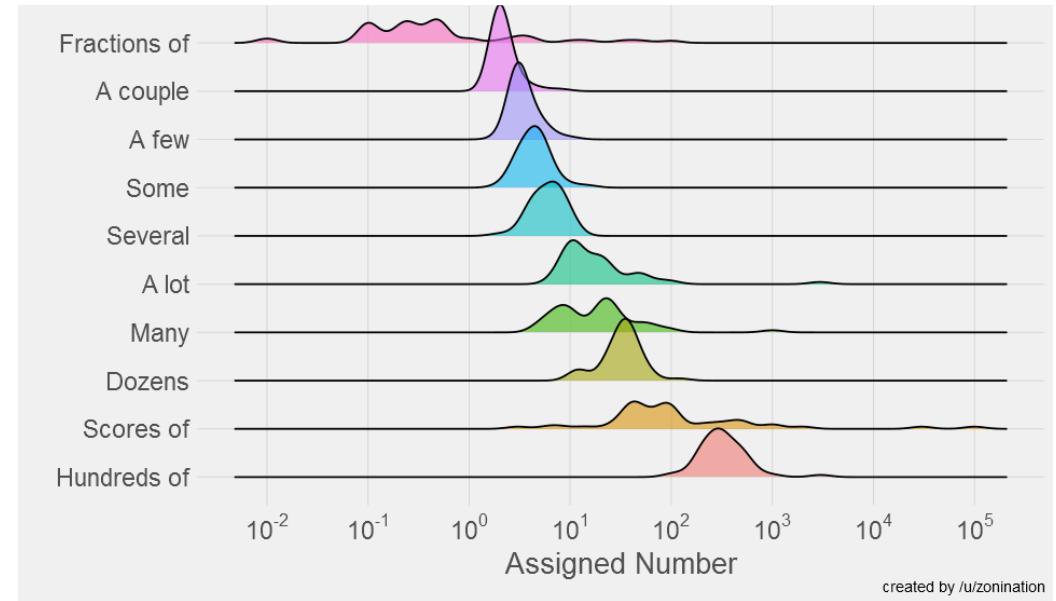


Figure 2. Density plots, mean values, and 5% and 95% percentiles of the numerical interpretations (in percentages) given by all participants for each phrase in the survey. Note that density plots are a smooth variant of histograms and may therefore be positive outside the data range of 0–100%.



created by /u/zonation

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
 - **"Ja-Sage-Tendenz" (Akquieszenz) oder „Nein-Sage-Tendenz“**
 - Antworttendenzen sind besonders ausgeprägt, wenn Urteilsobjekte wenig bekannt sind → **Gegenmaßnahme:** Personen 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)

Ratings: Beurteilungen auf Skalen

Probleme bei der Beantwortung von Ratingskalen

- Motivierte Verzerrungen

- Selbstdarstellungstendenz
- Selbstäuschung
- Soziale Erwünschtheit
- "faking good" (aber auch "faking bad")
- provokativer Antwortstil

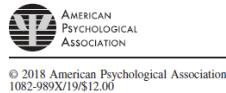
- Unmotivierte Verzerrungen

- soziale Vergleichsprozesse
- Verankerungseffekte
- Verfügbarkeitsheuristik
- falsche bzw. mangelhafte Repräsentation des zu messenden Merkmals
- Missverständnis des Fragebogens als Leistungstest

Selbstberichtsverfahren: Befragung und Rating

Ratings: Beurteilungen auf Skalen

Wie könnte es in Zukunft laufen?



Semantic Measures: Using Natural Language Processing to Measure, Differentiate, and Describe Psychological Constructs

Oscar N. E. Kjell and Katarina Kjell
Lund University

Danilo Garcia
Blekinge County Council, Karlskrona, Sweden, and University
of Gothenburg

Sverker Sikström
Lund University

Abstract

Psychological constructs, such as emotions, thoughts, and attitudes are often measured by asking individuals to reply to questions using closed-ended numerical rating scales. However, when asking people about their state of mind in a natural context (“How are you?”), we receive open-ended answers using words (“Fine and happy!”) and not closed-ended answers using numbers (“7”) or categories (“A lot”). Nevertheless, to date it has been difficult to objectively quantify responses to open-ended questions. We develop an approach using open-ended questions in which the responses are analyzed using natural language processing (Latent Semantic Analyses). This approach of using open-ended, semantic questions is compared with traditional rating scales in nine studies ($N = 92\text{--}854$), including two different study paradigms. The first paradigm requires participants to describe psychological aspects of external stimuli (facial expressions) and the second paradigm involves asking participants to report their subjective well-being and mental health problems. The results demonstrate that the approach using semantic questions yields good statistical properties with competitive, or higher, validity and reliability compared with corresponding numerical rating scales. As these semantic measures are based on natural language and measure, differentiate, and describe psychological constructs, they have the potential of complementing and extending traditional rating scales.

Translational Abstract

We develop tools called semantic measures to statistically measure, differentiate and describe subjective psychological states. In this new method, natural language processing is used for objectively quantifying words from open-ended questions, rather than the closed-ended numerical rating scales traditionally used today. Importantly, the results suggest that these semantic measures have competitive, or higher, validity and reliability compared with traditional rating scales. Using semantic measures also brings along advantages, including an empirical description/definition of the measured construct and better differentiation between similar constructs. This method encompasses a great potential in terms of improving the way we quantify and understand individuals’ states of mind. Semantic measures may end up becoming a widespread alternative applied in scientific research (e.g., psychology and medicine) as well as in various professional contexts (e.g., political polls and job recruitment).

Merkmale Psychologischer Test

- **Definition Test:** wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer empirisch unterscheidbarer Merkmale
- v.a. zum Einsatz in psychologischer Diagnostik, um eine möglichst genaue Aussage über den relativen Grad der individuellen Merkmalsausprägung zu bekommen
- **Annahme:** Aufgaben/Fragen des Tests werden von Menschen mit unterschiedlichen Fähigkeiten oder Eigenschaften unterschiedlich gelöst/beantwortet
- **Testwert** = Aggregation über die Einzelitems/-aufgaben des Tests

Merkmale Psychologischer Test

Für einen psychologischen Test gilt:

- a) Es handelt sich um eine Messmethode, bei der Personen auf **standardisierte Reizvorlagen** (Aufgaben, Fragen etc.) reagieren.
- b) Reaktionen werden durch die **spezifischen, im Test realisierten Bedingungen** hervorgerufen.
- c) Die Reaktionen erlauben einen **wissenschaftlich begründbaren Rückschluss auf die individuelle Ausprägung** eines psychologischen Merkmals (oder auch mehrere Merkmale).
- d) Das Vorgehen ist **standardisiert**.
- e) Ziel ist eine **quantitative (Ausprägung des Merkmals) und/oder eine qualitative Aussage** (Vorhandensein oder Art des Merkmals) über das psychologische Merkmal.

→ 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)

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 weitesten 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 d b d
 3. b 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 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 b d d d q b d d b b d d q d 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
- Spezialfall: **Adaptives Testverfahren**, welches eine Teilmenge von Aufgaben vorgibt 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:

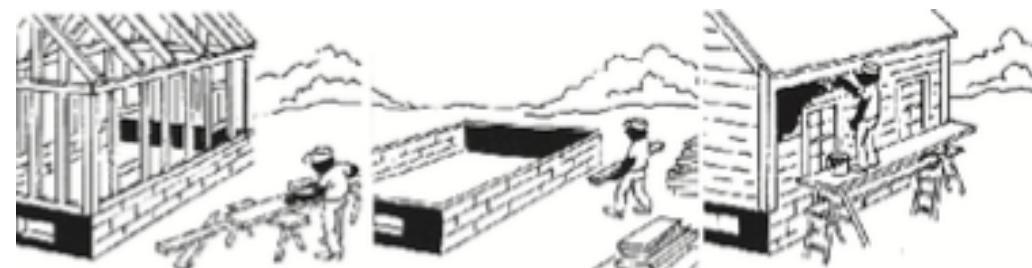
1. Allgemeines Wissen, z.B. »Was ist der Koran?«, »Wer erfand das Flugzeug?«
2. Zahlnachsprechen: Folgen von 3–9 Ziffern sind vorwärts und rückwärts nachzusprechen, z. B. »5-8-2« oder »4-2-7-3-1-8-2«
3. Wortschatz: »Was ist die Bedeutung von ...?« z. B. »anonym, Prestige, konkordant«
4. Rechnerisches 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.

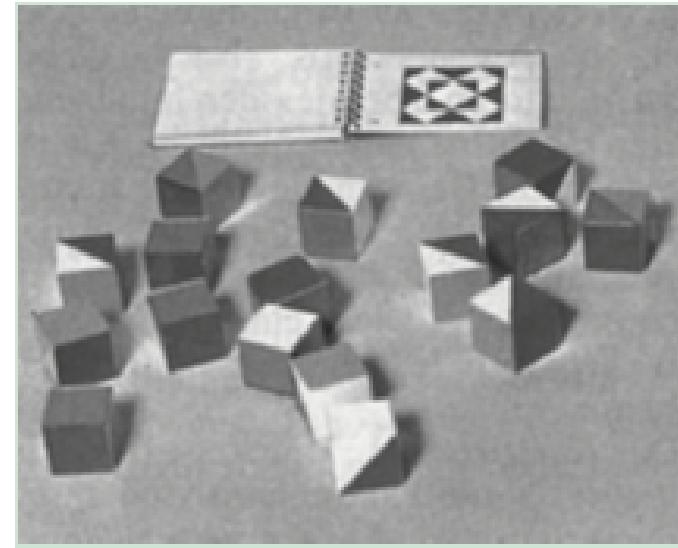


Leistungstests

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.

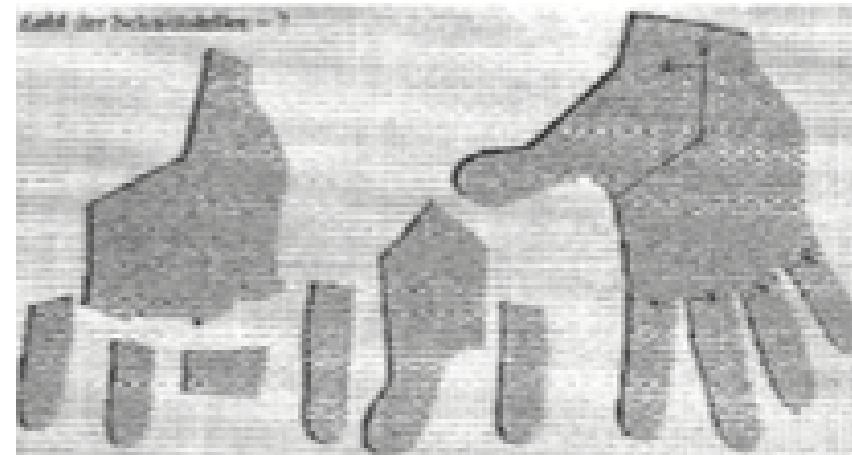


Leistungstests

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 Selbst- oder Fremdauskunft auf Fragebogenitems basieren (-> subjektiv) oder nicht (-> objektiv)
 - 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 Persönlichkeitsinventar (NEO-PI-R; Costa & Macrae, 1985) oder Kurzform NEO Five Factor Inventory (NEO-FFI; Costa & Macrae, 1992) - *mit Variante als Fremdbericht über andere Personen*
- Freiburger Persönlichkeitsinventar (FPI-R, Fahrenberg, Hampel & Selg, 1994)

starke Ablehnung Ablehnung neutral Zustimmung starke Zustimmung

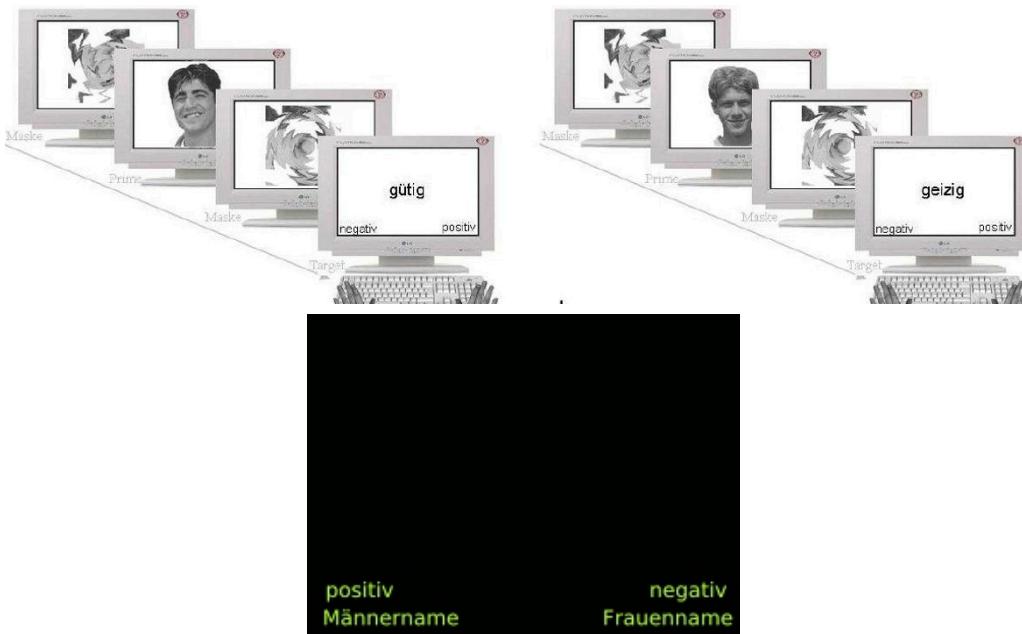
1. Ich bin nicht leicht beunruhigt.	<input type="radio"/> SA	<input type="radio"/> A	<input type="radio"/> N	<input type="radio"/> Z	<input type="radio"/> SZ
2. Die meisten Menschen, die mir begegnen, sind mir wirklich sympathisch.	<input type="radio"/> SA	<input type="radio"/> A	<input type="radio"/> N	<input type="radio"/> Z	<input type="radio"/> SZ
3. Ich habe eine sehr lebhafte Vorstellungskraft.	<input type="radio"/> SA	<input type="radio"/> A	<input type="radio"/> N	<input type="radio"/> Z	<input type="radio"/> SZ
4. Im Hinblick auf die Absichten anderer bin ich eher zynisch und skeptisch.	<input type="radio"/> SA	<input type="radio"/> A	<input type="radio"/> N	<input type="radio"/> Z	<input type="radio"/> SZ
5. Ich bin für meine Umsicht und meinen gesunden Menschenverstand bekannt.	<input type="radio"/> SA	<input type="radio"/> A	<input type="radio"/> N	<input type="radio"/> Z	<input type="radio"/> SZ
6. Ich ärgere mich oft darüber, wie andere Leute mich behandeln.	<input type="radio"/> SA	<input type="radio"/> A	<input type="radio"/> N	<input type="radio"/> Z	<input type="radio"/> SZ
7. Ich gehe Menschenansammlungen aus dem Weg.	<input type="radio"/> SA	<input type="radio"/> A	<input type="radio"/> N	<input type="radio"/> Z	<input type="radio"/> SZ
8. Ästhetik und Kunst bedeuten mir nicht sehr viel.	<input type="radio"/> SA	<input type="radio"/> A	<input type="radio"/> N	<input type="radio"/> Z	<input type="radio"/> SZ
9. Ich halte mich nicht für jemanden, der listig oder gerissen ist.	<input type="radio"/> SA	<input type="radio"/> A	<input type="radio"/> N	<input type="radio"/> Z	<input type="radio"/> SZ
10. Ich lasse mir lieber Entscheidungsmöglichkeiten offen, anstatt alles im Voraus zu planen.	<input type="radio"/> SA	<input type="radio"/> A	<input type="radio"/> N	<input type="radio"/> Z	<input type="radio"/> SZ

Auszug aus der deutschen Übersetzung des NEO-PI-R von Ostendorf & Angleitner (2004)

Persönlichkeitstests

Beispiele für objektive Persönlichkeitstests:

- Implizite Assoziationstests (IAT) basieren auf Reaktionszeiten, Kritik an Validität



- Objektiver Leistungsmotivationstest (OLMT, Schmidt-Atzert, 2007) bedient sich der Idee: Wer leistungsmotivierter ist, strengt sich bei den dargestellten Aufgaben mehr an, und legt eine weitere Strecke zurück (= drückt öfters die notwendige Taste)

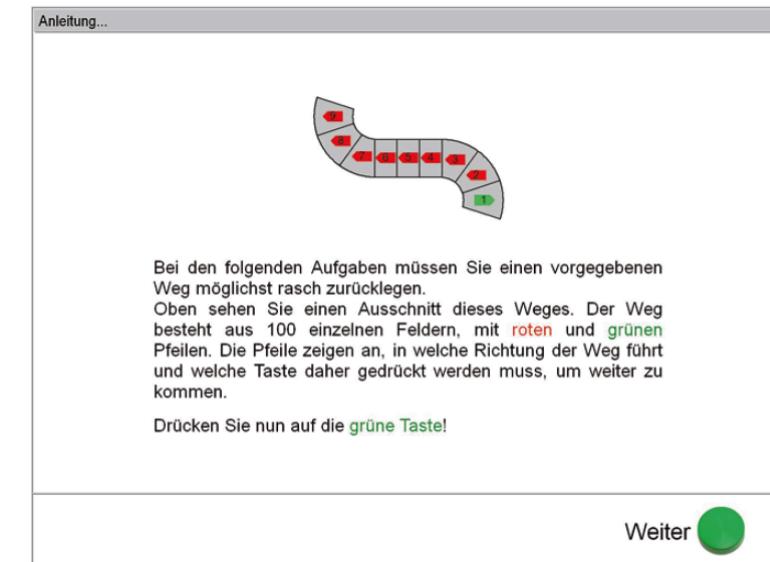


Abb. 3.22 Aufgabe im OLMT. (Aus Schmidt-Atzert 2007, mit freundlicher Genehmigung der SCHUHFRIED GmbH)

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
- "Objektiver Persönlichkeitstest" weil nicht auf Selbst- oder Fremdbericht basierend, nicht weil Gütekriterium der Objektivität zwingend erfüllt (bei Interpretationsobjektivität eher im Gegenteil)
- Große Variation in der Gültigkeit der Gütekriterien!! z.B. mangelhaft beim Rorschach-Test

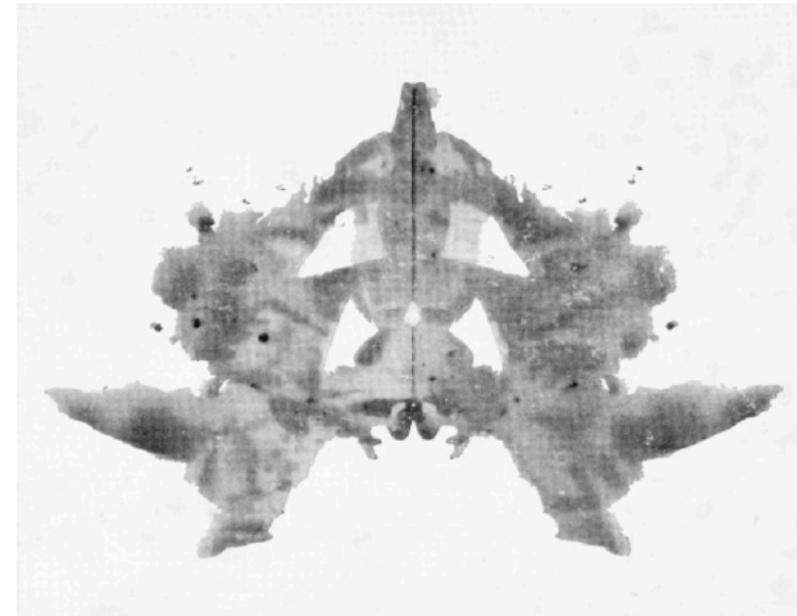


Abb. 3.24 Tafel aus dem Rorschach-Test. (Hermann Rorschach, Psychodiagnostik. Der Rorschach®-Test (Tafel I). © Verlag Hans Huber, Hogrefe AG, Bern/Switzerland)

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

Persönlichkeitstests

Projektive Verfahren als Sonderform von objektiven Persönlichkeitstests

Forens Psychiatr Psychol Kriminol (2022) 16:114–124
<https://doi.org/10.1007/s11757-022-00718-8>

ÜBERSICHT



Zur Verwendung projektiver Verfahren in der Familienrechtspychologie: ein Diskussionsbeitrag

Charis R. Neuerburg¹ · Rainer Banse²

Angenommen: 21. April 2022 / Online publiziert: 18. Mai 2022
© Der/die Autor(en) 2022

Zusammenfassung

Projektive Verfahren werden auch heute noch in der familienrechtspychologischen Diagnostik eingesetzt. Erfahrungsgemäß lassen sich 3 unterschiedliche Anwendungsarten unterscheiden: a) als klassisch diagnostisches Instrument, b) als informelle Verhaltensprobe oder c) als Explorationshilfe bzw. Gesprächseinsteig. In der folgenden Übersicht werden diese 3 Anwendungsarten dargestellt sowie ihr Nutzen für die familienrechtspychologische Diagnostik kritisch diskutiert. Hierfür wird zunächst ein Überblick über die empirische Fundierung der am meisten genutzten projektiven Verfahren (des Thematischen Apperzeptionstests und zeichnerische Verfahren allgemein) gegeben. Es wird dann diskutiert, ob die klassischen psychometrischen Gütekriterien auf projektive Verfahren anwendbar sind. Hinsichtlich der Verwendung projektiver Verfahren als informelle Verhaltensprobe wird auf mögliche Urteilsverzerrungen hingewiesen, insbesondere den „confirmation bias“, den Effekt der illusorischen Korrelation und den möglichen Einfluss irrelevanter Informationen auf den diagnostischen Prozess. Angesichts der potenziell negativen Auswirkungen auf die Validität der Diagnostik empfehlen wir, projektive Verfahren nicht in der Einzelfalldiagnostik einzusetzen, wenn keine direkten empirischen Belege für die Validität des spezifischen Verfahrens für die genutzte Auswertungsart und das zu diagnostizierende Konstrukt vorliegen.

Schlüsselwörter Explorationshilfe · Diagnostik bei Kindern · TAT · Rorschach · „Confirmation bias“ · Illusorische Korrelation

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)

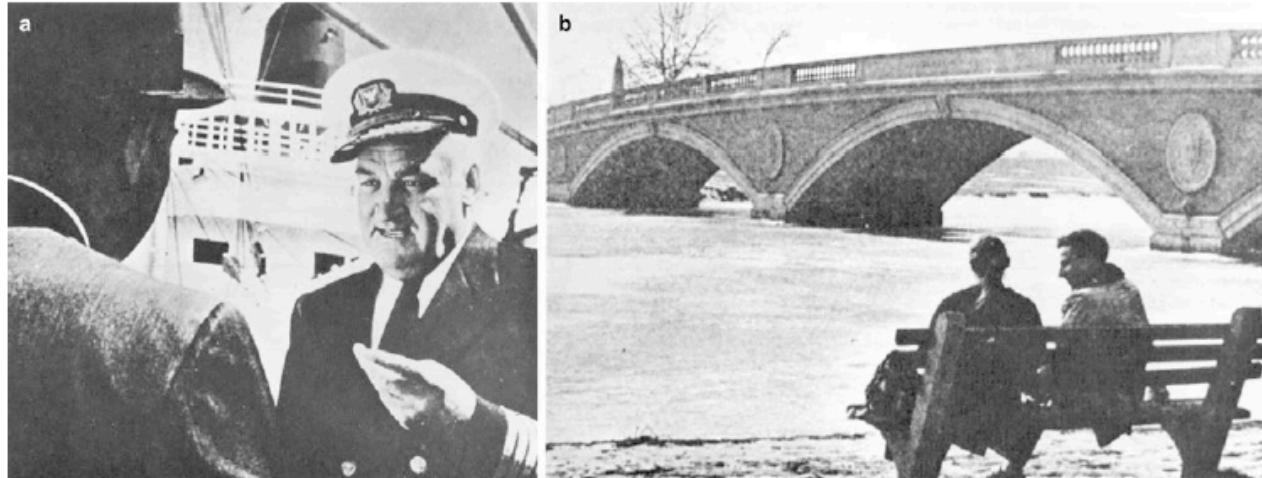


Abb. 3.27 Ausgewählte Bilder, bei denen deutsche Testpersonen ($n=428$) die höchsten Ausprägungen für ein Motiv zeigten (Kapitän: Macht; Paar vor Brücke: Affiliation). (Nach Pang 2010, © Oxford University Press)

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

Fragebögen und Tests in der Psychologie

Abstract

We examined the extent to which constructs and measures have proliferated in psychological science. We integrated two large databases obtained from the American Psychology Association (APA) that they have used to keep track of constructs, measures, and research in the psychological science literature for the past 30 years. Our descriptive analyses finds that (i) thousands of new constructs and measures are published each year, (ii) most measures are used very few times, and (iii) there is no trend towards consensus or standardization in the use of constructs and measures; in fact, there is a slight trend towards even greater fragmentation over time. That is, constructs and measures are proliferating. We conclude that measurement in the psychological science literature is fragmented, creating problems such as redundancy and confusion, and stifling cumulative scientific progress. We conclude by providing suggestions for what researchers can do about this problem.

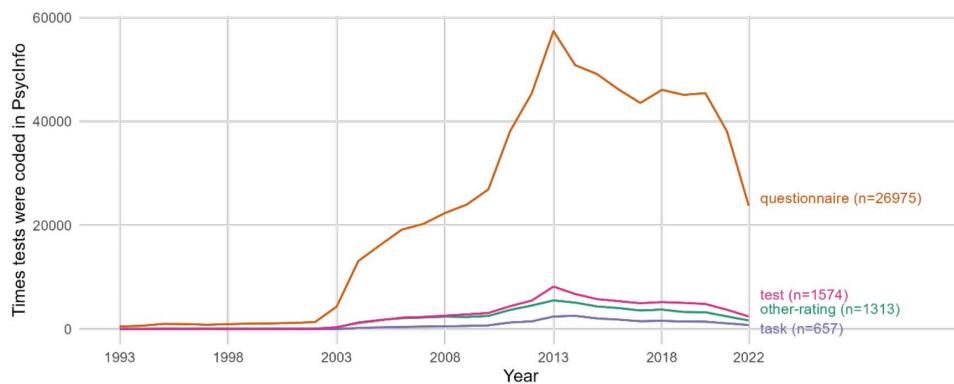


Figure 2. PsychInfo records how many times each measure was used in each year. The y-axis reflects the number of times a measure was recorded as being used in a given year, and the x-axis reflects each year of records. The Figure thus shows the trends in the number of times each measurement type was recorded in PsychInfo as having been used. Questionnaire = self-report ratings and inventories; other-rating = ratings for other (e.g., partner-reports or parent-reports); task = a behavioural task; test = other types of measures coded as being a test of some sort.

→ Sinnhaftigkeit so vieler Fragebögen und Tests?

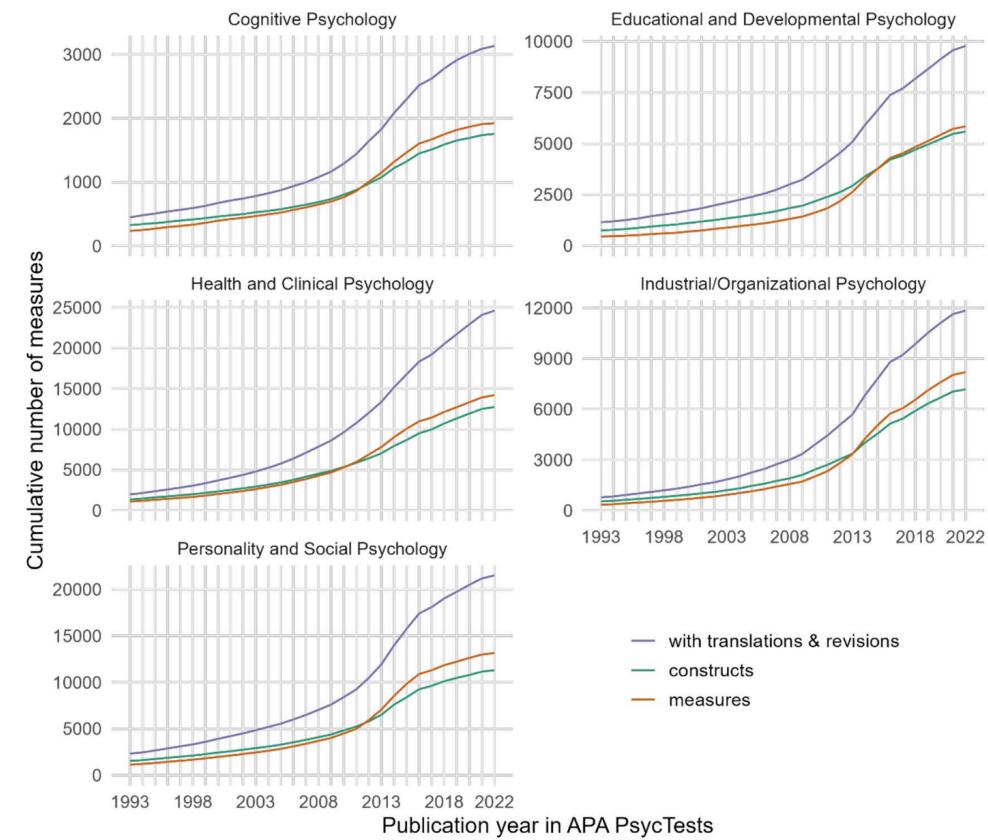


Figure 7. Cumulative number of distinct measures and constructs for each subdiscipline.

Take-Aways

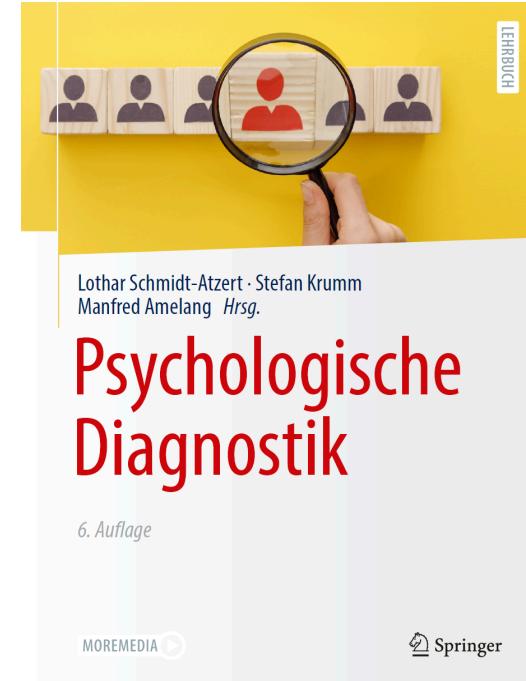
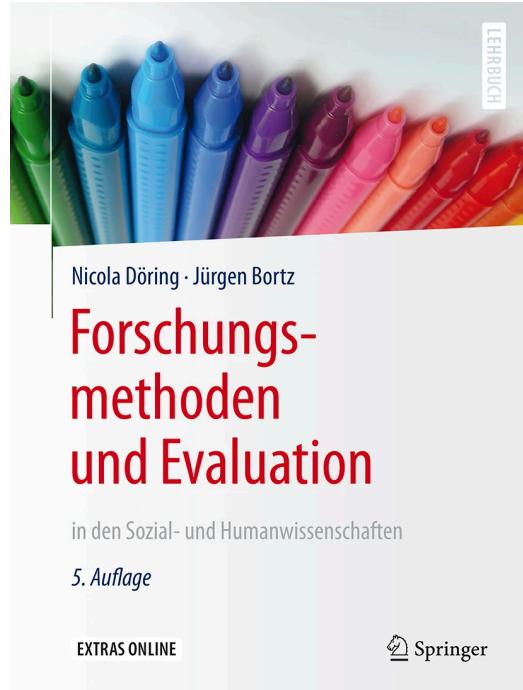
- Befragungen unterscheiden sich in ihrer **Darbietung** (schriftlich oder mündlich), **Standardisierung** (hinsichtlich der Antwortmöglichkeiten), **Strukturierung** (hinsichtlich der Vorgaben zur Fragenstellung), und der **Anzahl der befragten Personen**
- Befragungen in Form von **schriftlichen, standardisierten, und strukturierten 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**
- Man unterscheidet **subjektive** und **objektive Persönlichkeitstests**

Schlüssel-/Fachbegriffe der heutigen Vorlesung

standardisierte Befragung	unipolare Skalen	Ja-Sage-Tendenz = Akquieszenz	Verfügbarkeitsheuristik
unstandardisierte Befragung	bipolare Skalen	Nein-Sage-Tendenz	Leistungstests
(Voll-)strukturierte Befragung	Ambivalenz-Indifferenz-Problem	Gedankenlose Reproduktion	Speed-Tests
Teil-/Halbstrukturierte Befragung	numerische Marker	Motivierte Verzerrungen	Power-Tests
Forced Choice Antworten	verbale Marker	Unmotivierte Verzerrungen	subjektive Persönlichkeitstests
Ratingskalen	grafischer Marker	Soziale Erwünschtheit	objektive Persönlichkeitstests
	Tendenz zur Mitte	faking good / faking bad	projektive Verfahren
	Extreme Antworttendenz		

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

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 3.2. - 3.5. 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