

Einführung in die Forschungsmethoden der Psychologie und Psychotherapie

Einheit 4: Messen in der Psychologie - manifest

03.05.2023 | Prof. Dr. phil. Stephan Goerigk

Unterteilung von Erhebungsmethoden

- Verfahren des Beobachtens, Zählens und Messens
- Verfahren des Selbstberichts
- Psychologische Tests
- Biopsychologische bzw. neurowissenschaftliche Messungen

→ Verfahren sind nicht völlig distinkt, sondern weisen Überschneidungen auf

Auswahl und Einsatz von Erhebungsmethoden

Übergeordnete Perspektive

- **Ziel:** wissenschaftliche Fragestellungen in der Psychologie in Breite und Tiefe beantworten
- **Lösung:** Prinzipiell anstreben, Fragestellungen mit mehreren Datenerhebungsverfahren bzw. basierend auf unterschiedlichen Datenquellen zu untersuchen

Perspektive der Einzelstudie

- häufig nicht möglich oder sinnvoll (ökonomischen oder versuchsplanerische Gründe), innerhalb einer einzigen Studie verschiedene Methoden einzusetzen
- konkrete einzelne Untersuchung → gezielte Auswahl weniger Methoden

Qualitätskriterien von Erhebungsmethoden

Zur Auswahl und Bewertung psychologischer Erhebungsmethoden müssen Kriterien berücksichtigt werden:

Hauptgütekriterien:

- Objektivität
- Reliabilität
- Validität

Nebengütekriterien:

- Skalierung
- Normierung
- Ökonomie
- Nützlichkeit
- Zumutbarkeit
- Unverfälschbarkeit
- Fairness

Qualitätskriterien von Erhebungsmethoden

Objektivität

Definition: Eine Erhebungsmethode ist objektiv, wenn sie das Merkmal unabhängig von Testleiter:in, Testauswerter:in und von Ergebnisinterpretation misst.

3 Bereiche lassen sich unterscheiden:

1. Durchführungsobjektivität (Testleiterunabhängigkeit)
2. Auswertungsobjektivität (Verrechnungssicherheit)
3. Interpretationsobjektivität (Interpretationseindeutigkeit)

Qualitätskriterien von Erhebungsmethoden

Objektivität

Durchführungsobjektivität:

Definition:

Testergebnis soll nicht davon abhängen, welche Testleiter:in Test durchführt

→ Erhebung sollte unter möglichst standardisierten Bedingungen stattfinden

Standardisierung wird optimiert durch:

- Instruktionen der Testleiter schriftlich festgehalten
- soziale Interaktion zwischen Versuchsleiter und Testperson gering halten
- für möglichst ähnliche Untersuchungssituation sorgen

Qualitätskriterien von Erhebungsmethoden

Objektivität

Auswertungsobjektivität:

Definition:

Beim Vorliegen der Antworten/Beobachtungen einer Person soll jede Auswerter:in zum selben Ergebnis kommen

Kann durch statistische Kennzahlen zur Beurteilerübereinstimmung erfasst werden:

- Cohens Kappa
- Fleiss Kappa
- Konkordanzkoeffizienten nach Kendall

Auswertungsobjektivität wird optimiert durch:

- Vermeiden freier Antwortformate
- klare Auswertungsregeln

Qualitätskriterien von Erhebungsmethoden

Objektivität

Interpretationsobjektivität:

Definition:

Unterschiedliche Erheber:innen sollen beim Vorliegen der Ergebnisse zum selben Schluss kommen.

Interpretationsobjektivität kann erhöht werden durch:

- klare Regeln für die Interpretation
- Vorhandensein von Normen/Normwerten
- Verwendung von Prozenträngen

Qualitätskriterien von Erhebungsmethoden

Reliabilität

Definition:

- Eine Erhebungsmethode ist (vollständig) reliabel, wenn sie das Merkmal ohne Messfehler misst.
- Reliabilität gibt Grad der Genauigkeit an, mit der eine Erhebungsmethode misst

Es lassen sich 3-4 Arten der Reliabilität unterscheiden:

- **Retest-Reliabilität** → Erhebungsmethode kommt bei Wiederholung zum selben Ergebnis
- **Paralleltest-Reliabilität** → Erhebungsmethode kommt unter vergleichbaren Bedingungen zum selben Ergebnis
- **Innere Konsistenz** → Einzelteile der Erhebungsmethode kommen zu ähnlichen Ergebnissen wie andere Erhebungsmethode
- **(Testhalbierungs- (Split Half-)Reliabilität)** → analog zur inneren Konsistenz (nur Trennung in 2 Teile)

Qualitätskriterien von Erhebungsmethoden

Validität

Definition:

Eine Erhebungsmethode ist valide, wenn sie das Merkmal, das sie messen soll, auch wirklich misst.

3 Validitäten lassen sich unterscheiden:

- **Inhaltsvalidität**
 - Erhebungsmethode entspricht allen relevanten Bestandteilen des erhobenen Konstrukts (z.B. Depressionsfragebogen sollte alle relevanten Depressionssymptome und keine nicht für Depression relevanten Symptome enthalten)
- **Konstruktvalidität**
 - Erhebungsmethode erzeugt Daten, die von einer empirisch bewährten oder überzeugenden Theorie vorhergesagt werden (z.B. Testergebnis Diagnoseinstrument Autismus bildet erwartete Erblichkeit ab)
- **Kriteriumsvalidität**
 - Erhebungsmethode erzeugt Daten, die mit Kriterien des Merkmals korrelieren (z.B. Intelligenztest und Schulnoten)

Reaktivität

- Von Beginn an ist die Reaktivität der Erhebungsmethoden zu beachten.
- Definition: Veränderung/Verzerrung der erhobenen Daten aufgrund der Kenntnis der untersuchten Person darüber, dass sie Gegenstand einer Untersuchung ist

→ Datenerhebungen in Psychologie verändern oft bereits den Gegenstand

Beispiel: Hawthorne-Effekt

- klassische Studie in den Hawthorne-Werken der Western Electric Company (Roethlisberger und Dickson, 1939)
- bloße wissenschaftliche Untersuchung der Arbeiter:innen führte zu Steigerung der Produktivität
- Produktivitätssteigerung war **unabhängig** von den durch die Forscher implementierten Veränderungen der Arbeitsbedingungen
- **Bewertungserwartung** der untersuchten Personen aka. **Aufforderungscharakteristika** der Untersuchung

Maßnahmen zur Reduktion von Reaktivität

Maßnahme	Bewertung
Versuchspersonen in Unkenntnis darüber lassen, dass sie untersucht werden	Nur in Feld-, Archiv- oder Internetstudien praktikabel, nicht in Laborstudien; kann ethisch problematisch sein
Versuchspersonen Anonymität zusichern	Besonders wichtig bei der Erhebung von persönlichen/sensiblen Daten; reduziert sozial-erwünschtes Antworten
Coverstory über den Untersuchungszweck entwickeln	Wichtig in hypothesenprüfenden Studien, in denen die Versuchspersonen erforschtes Verhalten kontrollieren können; Täuschung ethisch zu reflektieren
Maße einsetzen, die von Versuchspersonen nicht kontrolliert oder beeinflusst werden können (nicht-reaktive Messverfahren)	Angenommen für biopsychologische Maße, die kaum kontrollierbare physiologische Vorgänge erfassen (z.B. Messung von Hormonspiegeln oder Verfahren zur Messung der Gehirnaktivität)
Indirekte/implizite Messverfahren einsetzen	Versuchspersonen können aus gemessenen Verhaltensweisen (oft Reaktionszeiten) nur schwer auf das untersuchte Konstrukt schließen

(Verhaltens-)beobachtung

- Jede Datenerhebung beinhaltet den Vorgang der visuellen Wahrnehmung, also Beobachtung (visuelle Wahrnehmung \neq wissenschaftliche Beobachtung)

Definition wissenschaftliche Beobachtung: Systematische und regelgeleitete Registrierung des Auftretens bzw. der Ausprägung von ausgewählten, psychologisch relevanten Merkmalen oder Ereignissen

Herausforderungen für Beobachtungen:

- **Selektivität**
 - Informationen müssen aus komplexem, sich ständig veränderndem Strom von ständig eintreffenden visuellen Daten ausgewählt werden
- **Konstruktivität**
 - Selektierte Informationen zu bedeutungsvollen, stabilen Einheiten zusammenfügen
 - Konstruktion abhängig von Einstellungen, Erwartungen und Vorerfahrungen der Beobachter:in

Lösung: → Wissenschaftliche Beobachtung nach a priori festgelegtem Beobachtungsplan

(Verhaltens-)beobachtung

Wissenschaftliche Beobachtung nach a priori festgelegtem Beobachtungsplan:

1. Was soll beobachtet werden? (Kategorien für das/die interessierende/n Ereignis/se oder Merkmal/e)
2. Wer beobachtet? (teilnehmend vs. nicht-teilnehmend)
3. Wo wird beobachtet? (Feld vs. Labor)
4. Wie wird beobachtet? (Grad der Strukturierung)
5. Wann wird beobachtet? (Zeitpunkte, Zeiträume, Situationen, gleichzeitige vs. zeitversetzte Registrierung)
6. Auf welche Weise wird das Beobachtete registriert, protokolliert, interpretiert?

(Verhaltens-)beobachtung

Beispiele für (Verhaltens-)beobachtung in der Psychologie:

Das Marshmellow Experiment (Mischel, 1989) ([Link](#))

- Experiment zum Belohnungsaufschub und zur Impulskontrolle
- Fähigkeit kurzfristig auf etwas Verlockendes für die Erreichung langfristiger Ziele zu verzichten

Bandura's Bobo Doll ([Link](#))

- Eine der wichtigsten Studien zum Beobachtungslernen aka Modell-Lernen
- Kinder sehen Film (Person "Rocky schlägt Bobo Puppe")
- 3 Varianten für Ende des Films:
 - Andere Person kommt hinzu, die „Rocky“ für sein Verhalten lobt und ihn mit Süßigkeiten belohnt.
 - Andere Person kommt hinzu, tadeln jedoch „Rocky“ und bestraf ihn mit Schlägen und Drohungen.
 - Geschehen bleibt unkommentiert, keine weitere Person tritt auf.

(Verhaltens-)beobachtung

Annahmen der (Verhaltens-)beobachtung:

1. Typisches Verhalten einer Person wird durch Eigenschaften der Person bewirkt.
2. Eine höhere Ausprägung einer bestimmten Eigenschaft macht das Auftreten des von ihr bewirkten Verhaltens wahrscheinlicher.
3. Es gibt Interindividuelle Unterschiede im Verhalten.
4. Beobachtungsraum = „Bildausschnitt“ (Verhaltensstichprobe)
5. Interindividuelle Unterschiede sind über die Zeit hinweg stabil.
6. Interindividuelle Unterschiede sind über verschiedene Situationen hinweg konsistent.

→ Gewohnheiten sind auf die Wirkung von einer oder mehreren Eigenschaften zurückzuführen

(Verhaltens-)beobachtung

Annahmen der (Verhaltens-)beobachtung:

Was wird beobachtet? - Definition von Beobachtungseinheiten:

- Abgrenzung des Beobachtungsobjekts: z.B. einzelne Personen vs. Familie vs. Paare...
- Abgrenzung der Variablen / Verhaltensmerkmale: umfassendes Bild vs. einzelnes Merkmal
- Abgrenzung der Dauer der Beobachtung: kurz/einmalig vs. lang/wiederholt
- Definition von Anfang, Ende und Inhalt einer Beobachtungseinheit:
 - Wann beginnt / endet ein Verhalten, das einen Hinweis auf das Verhaltensmerkmal gibt?
 - Welches Verhalten gibt Hinweis auf Verhaltensmerkmal?

(Verhaltens-)beobachtung

Annahmen der (Verhaltens-)beobachtung:

Was wird beobachtet? - Kriterien für gute Beobachtungseinheiten:

- Jede Beobachter:in kann Beobachtungseinheit **zuverlässig identifizieren**.
- Das beschriebene Verhalten **kommt** in der Beobachtungssituation **vor**.
- Zu beobachtendes Merkmal wird von Beobachtungseinheiten in **allen Facetten** erfasst.
- Liste der Verhaltensweisen kann von Beobachtenden während der Beobachtung **überblickt** werden.
- **Grenzwerte** für Beobachtende: max. 3 Eigenschaften oder max. 10 Einzelverhaltensweisen pro Beobachtung (abhängig von Erfahrung der Beobachter:in)
- Beobachtungseinheiten können klar voneinander **abgegrenzt** werden („unterricht andere“ vs. „redet dazwischen“)

Tipp: Beobachtungsplan in einem „Probendurchgang“ testen

(Verhaltens-)beobachtung

Formen und Rahmenbedingungen für Beobachtungen

Wer beobachtet? - Teilnehmende vs. nicht-teilnehmende Beobachtung:

Aktiv-teilnehmende Beobachtung vs. passiv-teilnehmende Beobachtung

- keine Geräte für Beobachtung notwendig
- bei aktiv-teilnehmender Beobachtung: „natürliche“ Situation möglich

Nicht-teilnehmende Beobachtung:

- Beobachtende sind in der Situation nicht anwesend bzw. werden von Beobachtetem nicht gesehen
- Situation wird nicht verändert

(Verhaltens-)beobachtung

Formen und Rahmenbedingungen für Beobachtungen

Wo wird beobachtet? - Feld- vs. Laborbeobachtung:

Bei Beobachtung im Feld wird in der „natürlichen“ Situation beobachtet

- Situation ist „echt“
- höhere ökologische Validität

Bei Beobachtung im Labor wird in einer „künstlich erzeugten“ Situation beobachtet. Rahmenbedingungen und Aufgaben sind genau geplant

- Kontrolle von Störfaktoren
- gezielte Planung der Bedingungen

(Verhaltens-)beobachtung

Formen und Rahmenbedingungen für Beobachtungen

Wann wird beobachtet? - Direkte (gleichzeitige) vs. indirekte (zeitversetzte) Beobachtung:

Direkte Beobachtung: Verhalten wird direkt zum Zeitpunkt des Auftretens beobachtet (setzt verhaltensahe Beobachtung voraus)

- keine Verzerrung durch Erinnerungseffekte
- wenig Interpretationsspielraum

indirekte Beobachtung: zeitversetzte „Beobachtung“

- Datenaggregation vor Beurteilung
- längere Verhaltensausschnitte

(Verhaltens-)beobachtung

Formen und Rahmenbedingungen für Beobachtungen

Wie wird beobachtet? - Strukturierte vs. unstrukturierte Beobachtung:

Strukturierte Beobachtung wird anhand eines vorher festgelegten Beobachtungsplans durchgeführt

- höhere Objektivität und Reliabilität
- Quantifizierung des beobachteten Verhaltens möglich

Unstrukturierte Beobachtung findet ohne genauen Beobachtungsplan statt

- exploratives Vorgehen wenn zu wenig Informationen zur Erstellung eines Beobachtungsplans vorliegen

(Verhaltens-)beobachtung

Formen und Rahmenbedingungen für Beobachtungen

Wie wird Beobachtung registriert? - Vermittelte vs. unvermittelte Beobachtung:

Vermittelte Beobachtung: Verhalten wird aufgezeichnet und später ausgewertet

- Fehlerminimierung durch wiederholtes Abspielen

Unvermittelte Beobachtung: eine Person beobachtet direkt oder indirekt

- unaufwändiger weil keine Geräte notwendig
- in allen Situationen möglich

(Verhaltens-)beobachtung

Formen und Rahmenbedingungen für Beobachtungen

Wie wird Beobachtung registriert? - Verhaltensahe Beobachtung vs. Verhaltensbeurteilung:

Verhaltensahe Beobachtung: Verhaltensmerkmale oder Situationsmerkmale werden hinsichtlich Intensität, Häufigkeit oder Dauer registriert

- Keine Interpretationsfehler
- Können später von mehreren Rater:innen beurteilt werden (Objektivität)

Verhaltensbeurteilung: Beobachtungsziel wird direkt eingeschätzt

- ökonomischer
- höherer Grad an subjektiver Interpretierbarkeit

(Verhaltens-)beobachtung

Formen und Rahmenbedingungen für Beobachtungen

Wie wird Beobachtung registriert? - Time vs. Event Sampling

Event Sampling:

Registrierung von Häufigkeit und/oder Dauer des Verhaltens- oder Situationsmerkmals über den gesamten Beobachtungszeitraum hinweg

- es können auch seltene Ereignisse erfasst werden
- erlaubt Überprüfung von Hypothesen zum Zusammenhang aufeinanderfolgender Ereignisse (Kontingenzen)

Time Sampling (Spezialfall des event sampling):

- Unterteilung des Beobachtungszeitraums in gleich lange Zeitintervalle
- Registrierung ob Verhaltensmerkmal im Zeitintervall gezeigt wird
- Möglichkeit, Pausen einzubauen (z.B. nur in jedem 2. Zeitintervall beobachten)
- Beobachterrotation möglich (z.B. Beobachter:in 1 im 1. Zeitintervall, Beobachter:in 2 im 2. Zeitintervall)
- oft akustischer oder vibrierender Signalgeber notwendig

(Verhaltens-)beobachtung

Formen und Rahmenbedingungen für Beobachtungen

Wie wird Beobachtung registriert? - Konstruktion von Ratingskalen

- unipolare vs. bipolare Skala
- Anzahl der Abstufungen: abhängig von Differenzierungsfähigkeit der Beobachter:in und Variation des Merkmals
- Je mehr Merkmale zu beurteilen sind, desto geringer sollte die Anzahl an Stufen sein
- Markierung und verbale Verankerung der Skalen: um Abstufungen der Skalen deutlich zu machen, werden diese numerisch/grafisch markiert oder verbal verankert.

unipolar:

|-----| feindseelig

numerisch verankert:

0 - 1 - 2 - 3

bipolar:

freundlich |-----| feindseelig

verbal verankert:

sehr niedrig - eher niedrig - eher hoch - sehr hoch

(Verhaltens-)beobachtung

Beobachtungsfehler und -verzerrungen (Beispiele)

Beobachtungsfehler	Beschreibung
Beobachterdrift	Im Verlauf zunehmend weniger/mehr Auftreten registrieren, obwohl sich am Verhalten nichts verändert hat (oft wegen Ermüdung, mangelnder Übung)
Erinnerungseffekte	Verzerrungen, wenn nicht unmittelbar protokolliert wird
Haloeffekt	Eine Eigenschaft einer Person wird durch eine andere Eigenschaft „überstrahlt“
Logischer Fehler	Von einem Hinweisreiz wird auf eine Eigenschaft geschlossen (obwohl kein Zusammenhang besteht)
Primacy Effekt	Überbetonung der ersten Information
Recency Effekt	Überbetonung der letzten Information
Mildeeffekt	Verzerrung zum Positiven
Strengeeffekt	Verzerrung zum Negativen
Zentrale Tendenz	Bevorzugung der Skalenmitte (häufig wegen Unsicherheit beim Beurteilen)
Tendenz zu Extremurteilen	Bevorzugung der Skalenpole
Kontrasteffekt	Einschätzung einer Person hängt von jener der zuvor eingeschätzten Person ab

(Verhaltens-)beobachtung

Verringern von Beobachtungsfehlern und -verzerrungen

Beobachtungsfehler	Beschreibung
Grenzwerte	Limit an zu beobachtenden Verhaltensmerkmalen und Beobachtungsdauer für BeobachterInnen
Verhaltensahe Beobachtung	Beobachtbare Einheiten, Vermeidung von Interpretation
konkreter/eindeutiger Beobachtungsplan	Missverständnisse vermeiden, klare Kategorien
Beobachterschulung	Vertrautmachen mit Protokollbogen, Definieren der Verhaltensmerkmale, Probedurchgänge
Beobachterrotation	verschiedene Beobachter:innen einsetzen; Versuchsperson in verschiedenen Situationen durch unterschiedliche Beobachter:innen beobachten

- auch bei Berücksichtigung sind Beobachtungsfehler nicht auszuschließen.
- bei jeder Verhaltensbeobachtung Beobachterübereinstimmung empirisch prüfen!

→ auch Beobachterübereinstimmung garantiert noch nicht, dass keine Beobachtungsfehler aufgetreten sind

(Verhaltens-)beobachtung

Beispiel: Aggressives Verhalten von Kindergartenkindern

Kategoriessystem:

1. schlägt, tritt
2. schlägt zurück
3. wirft mit Gegenstand
4. wirft mit Gegenstand zurück
5. droht, beschimpft
6. droht, schimpft zurück
7. zerstört beim Partner
8. zerstört Eigenes
9. zerstört beim Partner als Reaktion
10. zerstört Eigenes als Reaktion

- Jedes Kategoriensystem besteht aus mindestens zwei Kategorien
- Die optimale Zahl von Kategorien liegt zwischen 3 und 10 Kategorien.

Erhebungsmethoden der psychologischen Forschung

(Verhaltens-)beobachtung

Beispiel: Aggressives Verhalten von Kindergartenkindern (Time Sampling und Rating Skala)

Vehalten	Min. 1-5	Min. 6-10	Min. 11-15	Min. 16-20
1. schlägt, tritt	++	++	0	-
2. schlägt zurück	++	++	0	+
3. wirft mit Gegenstand	-	0	0	--
4. wirft mit Gegenstand zurück	0	-	--	0
5. droht, beschimpft	+	-	0	-
6. droht, schimpft zurück	--	++	0	++
7. zerstört beim Partner	-	0	++	++
8. zerstört Eigenes	++	--	0	--
9. zerstört beim Partner als Reaktion	--	++	++	++
10. zerstört Eigenes als Reaktion	0	++	+	++

(Verhaltens-)beobachtung

Maße für die Beobachterübereinstimmung

- Um Beobachterübereinstimmung und Reliabilität einer Verhaltensbeobachtung zu bestimmen gibt es eine Vielzahl von Maßen
- Wahl des angemessenen Übereinstimmungs- bzw. Zuverlässigkeitsmaßes abhängig vom Skalenniveau der Daten
- Nachfolgend werden nur die wichtigsten kurz vorgestellt:
 - prozentuale Übereinstimmung
 - Cohen's Kappa (mindestens nominal)
 - Kendalls Konkordanzkoeffizient (mindestens ordinal)
 - Intra-Class-Korrelation (ICC) (metrisch)

(Verhaltens-)beobachtung

Maße für die Beobachterübereinstimmung

Beispiel: nominales Urteil (ja/nein)

Vehalten	Rater1	Rater2
1. schlägt, tritt	Ja	Nein
2. schlägt zurück	Ja	Ja
3. wirft mit Gegenstand	Nein	Nein
4. wirft mit Gegenstand zurück	Ja	Ja
5. droht, beschimpft	Ja	Ja
6. droht, schimpft zurück	Ja	Ja
7. zerstört beim Partner	Nein	Ja
8. zerstört Eigenes	Ja	Ja
9. zerstört beim Partner als Reaktion	Nein	Nein
10. zerstört Eigenes als Reaktion	Nein	Nein

(Verhaltens-)beobachtung

Maße für die Beobachterübereinstimmung

Beispiel: metrisches Urteil (Skala 1-6)

Vehalten	Rater1	Rater2
1. schlägt, tritt	3	4
2. schlägt zurück	3	2
3. wirft mit Gegenstand	5	5
4. wirft mit Gegenstand zurück	1	2
5. droht, beschimpft	4	5
6. droht, schimpft zurück	1	1
7. zerstört beim Partner	6	6
8. zerstört Eigenes	5	3
9. zerstört beim Partner als Reaktion	2	2
10. zerstört Eigenes als Reaktion	1	1

(Verhaltens-)beobachtung

Übungsaufgabe:

- Überlegen Sie sich einen geeigneten Event Sampling Beobachtungsplan (wie oft wird Verhalten gezeigt) für **Werbungsverhalten von Säuglingen um die Aufmerksamkeit ihrer Eltern**

1. Ableitung geeigneter Kategorien (max. 10)

Link zur Beobachtungssituation

- teilnehmend vs. nicht-teilnehmend?
- Feld vs. Labor?
- Überlegen Sie sich, inwiefern Sie die Hauptgütekriterien erfüllt sehen.

(Verhaltens-)beobachtung

Maße für die Beobachterübereinstimmung

Prozentuale Übereinstimmung

- Anzahl der von allen Ratern gleich beurteilten Personen geteilt durch Anzahl der insgesamt beurteilten Personen x 100
- Problem: hohe a-priori Ratewahrscheinlichkeit

(Verhaltens-)beobachtung

Maße für die Beobachterübereinstimmung

Prozentuale Übereinstimmung

Vehalten	Rater1	Rater2
1. schlägt, tritt	Ja	Nein
2. schlägt zurück	Ja	Ja
3. wirft mit Gegenstand	Nein	Nein
4. wirft mit Gegenstand zurück	Ja	Ja
5. droht, beschimpft	Ja	Ja
6. droht, schimpft zurück	Ja	Ja
7. zerstört beim Partner	Nein	Ja
8. zerstört Eigenes	Ja	Ja
9. zerstört beim Partner als Reaktion	Nein	Nein
10. zerstört Eigenes als Reaktion	Nein	Nein

- Anzahl Kategorien: 10
- Gleich beantwortete Kategorien: 8
- Prozentuale Übereinstimmung = $8 / 10 \times 100 = 80\%$

(Verhaltens-)beobachtung

Maße für die Beobachterübereinstimmung

zufallskorrigiertes Maß: Cohen's Kappa

- Maß für die Beobachterübereinstimmung bei kategorialen Daten
- liegt zwischen -1 und +1 (prüfen auf Signifikanz für k möglich)
- es lässt sich kein allgemeiner Wert bestimmen, ab welchem k als hohe Übereinstimmung zu klassifizieren ist
- abhängig von untersuchtem Gegenstand, von Anzahl der Kategorien etc.

Für die Praxis (Landis & Koch, 1977; jedoch kritisch zu hinterfragen):

- $k < 0$ schlechte Übereinstimmung
- $0 < k < .20$ etwas Übereinstimmung
- $.21\text{--}.40$ ausreichende Übereinstimmung
- $.41\text{--}.60$ mittelmäßige Übereinstimmung
- $.61\text{--}.80$ beachtliche Übereinstimmung
- $.81\text{--}1.00$ (fast) vollkommene Übereinstimmung

(Verhaltens-)beobachtung

Maße für die Beobachterübereinstimmung

Beispiel: Cohen's Kappa

Vehalten	Rater1	Rater2
1. schlägt, tritt	Ja	Nein
2. schlägt zurück	Ja	Ja
3. wirft mit Gegenstand	Nein	Nein
4. wirft mit Gegenstand zurück	Ja	Ja
5. droht, beschimpft	Ja	Ja
6. droht, schimpft zurück	Ja	Ja
7. zerstört beim Partner	Nein	Ja
8. zerstört Eigenes	Ja	Ja
9. zerstört beim Partner als Reaktion	Nein	Nein
10. zerstört Eigenes als Reaktion	Nein	Nein

$$\kappa = \frac{P_0 - P_e}{1 - P_e}$$

P_0 : relativer Anteil übereinstimmender Urteile P_e : relativer Anteil Zufallsübereinstimmungen

- $P_0 := 8 / 10 \times 100 = 80\%$
- Rater 1: 8/10 Personen mit Ja bewertet = 80%
- Rater 2: 8/10 Personen mit Ja bewertet = 80%
- Rater 1: 2/10 Personen mit Nein bewertet = 20%
- Rater 2: 2/10 Personen mit Nein bewertet = 20%

$$P_e = \frac{8}{10} \cdot \frac{2}{10} + \frac{8}{10} \cdot \frac{2}{10} = 0.32$$

$$\kappa = \frac{0.8 - 0.32}{1 - 0.32} = 0.71$$

Take-Aways

- **Hauptgütekriterien** für psychologische Erhebungen sind Objektivität, Reliabilität und Validität
- **Objektivität:** Erhebung unabhängig von Testleiter:in, Testauswerter:in und Ergebnisinterpretation
- **Reliabilität:** Fehlerfreie und zuverlässige Erhebung des Merkmals
- **Validität:** Methode erhebt wirklich das interessierende Merkmal und nicht etwas anderes
- Wissenschaftliche Beobachtung erfolgt nach a priori festgelegtem **Beobachtungsplan**
- Häufig nutzt man zur Bewertung von Beobachtungen die Konstruktion von **Ratingskalen**
- **Beobachtungsfehler und -verzerrungen** sollten möglichst bewusst vermieden werden
- Zur Eindeutigkeit von Beobachtungen können **Maße der Beobachterübereinstimmung** berechnet werden