Vorlesung Forschungsmethoden

25.10.2018

Urte Scholz

Lernziele der heutigen Veranstaltung

Am Ende der Veranstaltung …

… sind Sie in der Lage, besondere Herausforderung bei der Messung psychologischer Variablen zu benennen und mögliche Lösungen zu finden.

… wissen Sie, welches die drei Hauptgütekriterien quantitativer Erhebungsmethoden sind, können sie definieren, voneinander abgrenzen und die Zusammenhänge benennen.

… können Sie verschiedene Arten von Objektivität definieren und erklären, wozu diese notwendig sind.

… können Sie verschiedene Arten der Reliabilität definieren, erklären, wann man welche Art der Reliabilität verwenden kann und sollte sowie die jeweiligen Vor- und Nachteile bzw. Besonderheiten benennen.

Besonderheiten psychologischer Erhebungen

zentrales Ziel psychologischer Forschung: Erkenntnisgewinn bezüglich häufig nicht direkt beobachtbarer psychischer Prozesse

Probleme des Selbstberichts:

Zugänglichkeit

Verzerrungen

Reaktivität

Definition: „Reaktivität bei psychologischen Datenerhebungen bedeutet die Veränderung bzw. Verzerrung der erhobenen Daten alleine aufgrund der Kenntnis der untersuchten Personen darüber, dass sie Gegenstand einer Untersuchung sind.“ (Hussy et al., 2013, S. 57)

Hawthorne-Effekt (Roethlisberger & Dickson, 1939) https://www.youtube.com/watch?v=W7RHjwmVGhs

Beispiel reaktive Messverfahren

https://www.google.ch/search?q=fitbit&client=firefox-b&dcr=0&source=lnms&tbm=isch&sa=X&ved=0ahUKEwis1uyv2O3WAhWIvRQKHc6rDmIQ_AUICigB&biw=1536&bih=758#imgrc=i1H8jDLoLiVqEM:

http://abnehmen-tipps.me/wp-content/uploads/2013/03/Em%C3%A4hrungsprotokoll.pdf

Ernährungsprotokoll

Datum:_____

Frühstück

Nahrung	Uhrzeit	Menge	Zubereitung

Zwischenmahlzeit

Nahrung	Uhrzeit	Menge	Zubereitung

Mittagessen

Nahrung	Uhrzeit	Menge	Zubereitung

Zwischenmahlzeit

Nahrung	Uhrzeit	Menge	Zubereitung

Abendessen

Nahrung	Uhrzeit	Menge	Zubereitung

Health Psychology

0278-6133/14/$12.00 http://dx.doi.org/10.1037/hea0000104

The Question–Behavior Effect: Genuine Effect or Spurious Phenomenon? A Systematic Review of Randomized Controlled Trials With Meta-Analyses

Angela M. Rodrigues and Nicola O'Brien
Newcastle University

David P. French
University of Manchester

Liz Glidewell
University of Leeds

Falko F. Sniehotta
Newcastle University

Objective: Simply answering questions about a specific behavior may change that behavior. This is known as the mere-measurement or question–behavior effect (QBE). Our objective was to synthesize the evidence for the QBE on health-related behaviors. ***Method:*** Included studies were randomized controlled trials that tested the effect of questionnaires or interviews about health-related behaviors and/or related cognitions compared with a no-measurement control condition or another form of measurement. Subgroup analyses were conducted to identify potential moderators. ***Results:*** 41 studies were included assessing a range of health behaviors. Meta-analyses showed a small overall QBE effect (SMD = 0.09; 95% CI [0.04, 0.13]; $k = 33$). Studies showed moderate heterogeneity, variable risk of bias, and evidence of publication bias. No dose–response relationships were found from studies comparing more with less intensive measurement conditions. There were no significant differences in QBE by behavior, but QBEs for dental flossing, physical activity, and screening attendance were significantly different from 0. Findings were not altered by whether behavior or cognitions were measured, attitudes were or were not measured, studies used questionnaires or interviews, or outcomes were objective or self-reported. ***Conclusions:*** There is some evidence for the QBE on health-related behavior. However, risk of bias within studies and evidence of publication bias indicate that the observed small effect size may be overestimated, especially given that some studies included intervention techniques in addition to providing questionnaires. Preregistered high-quality trials with clear specification of intervention content are needed to confirm if and when measurement leads to behavior change.

Keywords: question-behavior effect, mere-measurement effect, health behavior, behavior change

Health Psychology
2014, Vol. 33, No. 7, 646–655

0278-6133/14/$12.00 http://dx.doi.org/10.1037/a0033505

Promoting the Return of Lapsed Blood Donors: A Seven-Arm Randomized Controlled Trial of the Question–Behavior Effect

Gaston Godin
Laval University

Marc Germain
Héma-Québec

Mark Conner
University of Leeds

Gilles Delage
Héma-Québec

Paschal Sheeran
University of Sheffield

Psychology & Health, 2014
Vol. 29, No. 4, 390–404, http://dx.doi.org/10.1080/08870446.2013.858343

Routledge
Taylor & Francis Group

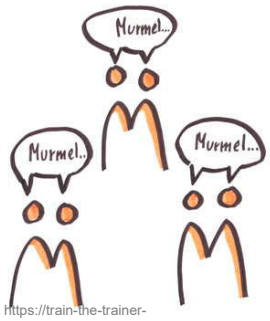
Why does asking questions change health behaviours? The mediating role of attitude accessibility

Chantelle Wood[a]*, Mark Conner[b], Tracy Sandberg[b], Gaston Godin[c] and Paschal Sheeran[a]

[a]Department of Psychology, University of Sheffield, Sheffield, UK; [b]Institute of Psychological Sciences, University of Leeds, Leeds, UK; [c]Faculty of Nursing, Research Group on Behaviours and Health, Laval University, Québec City, Canada

Massnahmen zur Reduzierung von Verzerrungen

→ Wie könnte man Verzerrungen bei der Erhebung von Daten verhindern?



https://train-the-trainer-seminar.de/trainingsmethoden/25_Murmelgruppen.html

Massnahmen zur Reduzierung von Verzerrungen (Hussy et al., 2013)

Konkrete / manifeste und abstrakte / latente Variablen

In Psychologie Rückschluss von manifesten (sichtbaren, messbaren) Variablen auf latente (unsichtbare, nicht direkt messbare) Variablen

Bsp.:

Konkrete / manifeste Variable → abstrakte / latente Variable

Operationalisierung → Hypothetisches Konstrukt

Kopfumfang? → Intelligenz

IQ Test? → Intelligenz

Wie stellen wir sicher, dass die manifeste, gemessene Variable eine gute Abbildung der latenten Variable ist?

Quantitative Gütekriterien

- Objektivität
- Reliabilität
- Validität

Quantitative Gütekriterien: Objektivität

Definition:

Die Objektivität bzw. Anwenderunabhängigkeit einer Untersuchung / eines Tests / eines Fragebogens gibt an, in welchem Ausmass die Testergebnisse von den Testanwendern /Testanwenderinnen unabhängig sind. (Döring & Bortz, 2016, S. 442)

- Durchführungsobjektivität
- Auswertungsobjektivität
- Interpretationsobjektivität

(Döring & Bortz, 2016, S. 443)

Aus Tewes, U., Schallberger, P., Rossmann, U. (Hrsg.) (1999). Hamburg-Wechsler-Intelligenztest für Kinder III (HAWIK-III). Bern: Huber

Abbruch

Wenn das Kind 5 aufeinander folgende Aufgaben nicht oder falsch beantwortet hat.

Anweisung

- Bevor man dem Kind das erste Bild vorlegt, sagt man:

  «Ich zeige dir nun einige Bilder, auf denen jeweils ein Teil fehlt. Sieh dir bitte jedes Bild genau an und sage mir dann, was darauf fehlt!»

- Dann legt man das Ringbuch mit den Bildern vor das Kind, schlägt die Beispielaufgabe auf und sagt:

  «Nun sieh dir dieses Bild an. Welcher wichtige Teil fehlt hier?»

Das Bild bleibt maximal 20 Sekunden vor dem Kind liegen.

Gibt das Kind eine richtige Antwort, so deckt man als nächstes die Aufgabe auf, mit der die Altersgruppe dieses Kindes beginnt, und sagt:

«Nun, was fehlt auf diesem Bild?»

Die Frage kann verkürzt oder weggelassen werden, wenn eindeutig erkennbar ist, dass das Kind die Aufgaben verstanden hat.

- Wenn die Antwort des Kindes nicht klar verständlich oder mehrdeutig ist, sagt man:

  «Zeige mir bitte, was du meinst!»

- Das Kind kann sich jedes Bild maximal 20 Sekunden anschauen.

- Wenn das Kind das fehlende Detail in der Beispielaufgabe innerhalb von 20 Sekunden nicht erkennt, sagt man:

  «Sieh! Hier fehlt die Schreibmine in der Bleistiftspitze» **(zeigt auf die entsprechende Stelle)**.

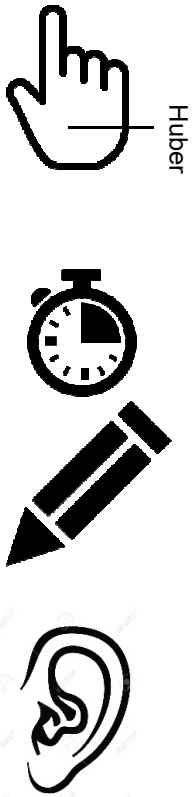
  Wenn das Kind bei Aufgabe 1 oder Aufgabe 2 die richtige Antwort nicht innerhalb von 20 Sekunden findet, bekommt es dafür jeweils 0 Punkte, und man sagt (bei Aufgabe 1):

  «Sieh! Hier fehlt das Ohr» **(zeigt auf die entsprechende Stelle)**.

  oder (für Aufgabe 2):

  «Sieh! Hier fehlt ein Deckel» **(zeigt auf die entsprechende Stelle)**.

- Für alle weiteren Aufgaben darf keine Hilfe mehr gegeben werden. Wenn das Kind das fehlende Detail nicht innerhalb von 20 Sekunden benennt oder nicht mit dem Finger darauf zeigt, erhält es für die betreffende Aufgabe null Punkte. Anschließend wird die nächste Aufgabe dargeboten. Bei einer falschen Antwort geht man sofort zur nächsten Aufgabe über und wartet nicht ab, ob sich das Kind innerhalb der Zeitgrenze von 20 Sekunden noch korrigiert.



Aus Tewes, U., Schallberger, P., Rossmann, U. (Hrsg.) (1999). Hamburg-Wechsler-Intelligenztest für Kinder III (HAWIK-III). Bern: Huber

Untertestaufgaben

Bildinhalt	fehlender Teil
1 Fuchs	Ohr
2 Karton	Deckel, Faltklappe
3 Katze	Schnurrhaare, Barthaare
4 Hand	Fingernagel des kleinen Fingers, Nagel, Nagellack
5 Elefant	Bein, Fuß
6 Mann	Uhrarmband
7 Tür	Türangel, Scharnier
8 Spiegel	Spiegelbild der Puppe (Wenn das Kind nur mit «Puppe» antwortet, sagt man: «Zeige mir bitte, was du meinst!»)
9 Uhr	die Ziffer «Elf» …die zweite «Eins» der Ziffer «Elf»… dort, wo die «Elf» sein sollte, ist nur eine «Eins»

Aus Tewes, U., Schallberger, P., Rossmann, U. (Hrsg.) (1999). Hamburg-Wechsler-Intelligenztest für Kinder III (HAWIK-III). Bern: Huber

Bewertung und Protokollierung

Für jede richtige Antwort, die innerhalb der Zeitgrenze von 20 Sekunden gegeben wird, gibt es einen Punkt. Ferner gibt es jeweils einen Punkt für jede Aufgabe, die unterhalb der Aufgabe liegt, mit dem die betreffende Altersgruppe beginnt. Null Punkte gibt es, wenn die Antwort des Kindes falsch ist oder nicht innerhalb von 20 Sekunden erfolgt.

Die meisten Kinder geben eine mündliche Antwort zum fehlenden Detail. Manchmal zeigt ein Kind jedoch nur auf das fehlende Detail. Wenn das Kind nur auf die richtige Stelle zeigt, erhält es einen Punkt für eine richtige Antwort. Wenn es jedoch auf die richtige Stelle zeigt und durch eine erklärende Erläuterung erkennen lässt, dass es etwas Falsches meint, gilt die Aufgabe als nicht gelöst. So kann ein Kind beispielsweise bei Aufgabe 16 auf das Innere der Badewanne zeigen und dazu sagen: «Hier fehlt das Wasser.» Wenn das Kind beispielsweise bei Aufgabe 17 mit dem Finger auf die Glühbirne zeigt, sollte man klären, ob das Kind erkannt hat, dass der Glühfaden fehlt und dass es nicht die Fassung meint.

Aus Tewes, U., Schallberger, P., Rossmann, U. (Hrsg.) (1999). Hamburg-Wechsler-Intelligenztest für Kinder III (HAWIK-III). Bern: Huber

Interpretationsobjektivität: Normwertetabellen

IQ-Wert		IQ-Wert		IQ-Wert	
unter 55	schwere bis schwerste Retardierung/ Behinderung	85 - 99	Grenzbereich niedriges Niveau im Normalbereich	115 - 129	überdurch-schnittliche Intelligenz
55 - 69	leichte Retardierung/ Behinderung	100	Normwert (mittlerer Durchschnitt)	130 - 145	Hochbegabung
70 - 84	unterdurch-schnittliche Intelligenz	101 - 114	Grenzbereich hohes Niveau im Normalbereich	über 145	Höchstbegabung

http://www.hochbegabten-homepage.de/intelligenztest_fuer_kinder.html

Quantitative Gütekriterien: Objektivität

Objektivität durch Standardisierung von Durchführung, Auswertung und Interpretation der Untersuchung / des Tests / des Fragebogens.

→ Instruktionen im Testhandbuch / Manual / Handanweisung
→ einfach zu erreichendes Gütekriterium

Objektivität = Voraussetzung für die weiteren Gütekriterien

Quantitative Gütekriterien: Reliabilität

Definition:

„Die Reliabilität […] („reliability“) gibt an, wie gering oder stark ein Test durch Messfehler verzerrt ist.“

(Döring & Bortz, 2016, S. 442).

Synonyme für Reliabilität:

- Zuverlässigkeit
- Präzision
- Messgenauigkeit

Quantitative Gütekriterien: Reliabilität

Basiert auf Annahmen der Klassischen Testtheorie:

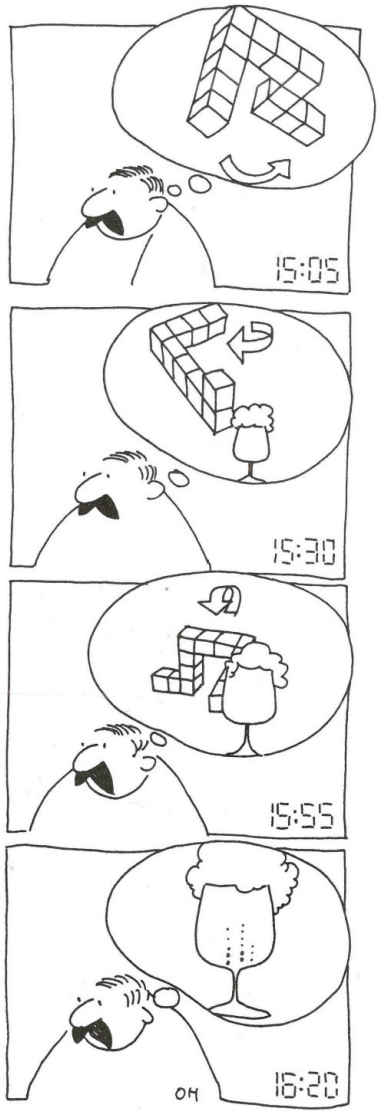
- Testwert X = wahrer Wert T + Messfehler E
- Reliabilität umso höher, je kleiner der zu Testwert X gehörende Messfehler E

→ Perfekte Reliabilität: X = T

- Fehlervarianz = unsystematische Abweichungen von wahren Werten

(Döring & Bortz, 2016, S442-443; Gravetter & Forzano, 2018, S.61ff.)

Beispiel für Messfehlerquelle (Huber, 2013)



Quantitative Gütekriterien: Reliabilität

- Reliabilität = Anteil wahre Varianz / beobachtete Varianz
- Werte: 0 bis 1
- zwischen 0.8 und 0.9 = ausreichend; > 0.9 = hoch (Fisseni, 1990; Bühner, 2011 zit. nach Döring & Bortz, 2016)
- Aber auch Empfehlungen, dass Reliabilität > .70 akzeptabel ist (z.B. Kline, 1999; Field, 2005)

Reliabilitätsarten

1. Test-Retest-Reliabilität (Stabilität)
2. Paralleltest-Reliabilität
3. Testhalbierungs-Reliabilität
4. Interne Konsistenz

5. Interrater-Reliabilität

Reliabilitätsarten: Test-Retest-Reliabilität (z.B. Döring & Bortz, 2016)

= Ausmass der Übereinstimmung bei einer wiederholten Anwendung des Tests / Fragebogens / der Untersuchung bei der gleichen Stichprobe

- Korrelation der Testwerte des ersten und zweiten Messzeitpunkts

2. Wie häufig werden Sie und diese Person ärgerlich oder sauer aufeinander?

	selten oder nie	manchmal	häufig	besonders häufig	extrem häufig
Mutter	1	2	3	4	5
Vater	1	2	3	4	5
Freund(in)	1	2	3	4	5
Partner(in)	1	2	3	4	5
Geschwister	1	2	3	4	5
Kind	1	2	3	4	5

2. Wie häufig werden Sie und diese Person ärgerlich oder sauer aufeinander?

	selten oder nie	manchmal	häufig	besonders häufig	extrem häufig
Mutter	1	2	3	4	5
Vater	1	2	3	4	5
Freund(in)	1	2	3	4	5
Partner(in)	1	2	3	4	5
Geschwister	1	2	3	4	5
Kind	1	2	3	4	5

1. Messzeitpunkt (t1) — 4 Wochen — 2. Messzeitpunkt (t2) — Zeit

Reliabilitätsarten: Test-Retest-Reliabilität (z.B. Döring & Bortz, 2016)

Bei stabilen Merkmalen führt eine reliable Testung bei Wiederholung unter gleichen Bedingungen zu gleichen / sehr ähnlichen Ergebnissen (Gravetter & Forzano, 2018; Hussy et al., 2013)

- r_{t1t2} = 0.89 → 89% der Merkmalsvarianz = wahre Varianz
- Probleme: Erinnerungseffekte, aufwendig in der Durchführung
- nicht geeignet bei instabilen Merkmalen

Reliabilitätsarten: Test-Retest-Reliabilität (z.B. Döring & Bortz, 2016)

Häufige Quellen für Messfehler (Gravetter & Forzano, 2018)

Reliabilitätsarten: Paralleltest-Reliabilität (z.B. Döring & Bortz, 2016)

= Übereinstimmung zweier Versionen (Äquivalenz) des gleichen Tests innerhalb einer Stichprobe

- Korrelation der Testwerte von Version A und Version B (r_{tAtB})

Version A (tA)

Frage 1 Version A

	selten oder nie	manchmal	häufig	besonders häufig	extrem häufig
uA	1	2	3	4	5
vA	1	2	3	4	5
wA	1	2	3	4	5
xA	1	2	3	4	5
yA	1	2	3	4	5
zA	1	2	3	4	5

Version B (tB)

Frage 1 Version B

	selten oder nie	manchmal	häufig	besonders häufig	extrem häufig
uB	1	2	3	4	5
vB	1	2	3	4	5
wB	1	2	3	4	5
xB	1	2	3	4	5
yB	1	2	3	4	5
zB	1	2	3	4	5

- Sehr aufwendig in der Entwicklung
- Anwendung: z.B. bei Gruppentestungen im Leistungsbereich oder bei wiederholter Testung gleicher Personen

Zeit

1. Messzeitpunkt (t1)

Reliabilitätsarten: Testhalbierungs-Reliabilität (split half)

(z.B. Döring & Bortz, 2016)

= Übereinstimmung zweier Hälften (Äquivalenz) des gleichen Tests innerhalb einer Stichprobe

- Korrelation der Testwerte von Hälfte A und Hälfte B ($r_{t1/2A\ t1/2B}$)
- Testhalbierung z.B. durch Zufallsauswahl / gerade vs ungerade Fragen / erste vs letzte Hälfte
- Reliabilität steigt mit der Anzahl der Items

→ Unterschätzung durch Testhalbierung

- Korrektur durch Spearman-Brown-Prophecy-Formula (s. Döring & Bortz, 2016)

2. Wie häufig werden Sie und diese Person ärgerlich oder sauer aufeinander?

	selten oder nie	manchmal	häufig	besonders häufig	extrem häufig
Mutter	1	2	3	4	5
Vater	1	2	3	4	5
Freund(in)	1	2	3	4	5
Partner(in)	1	2	3	4	5
Geschwister	1	2	3	4	5
Kind	1	2	3	4	5

Hälfte A (t1/2A)

Hälfte B (t1/2B)

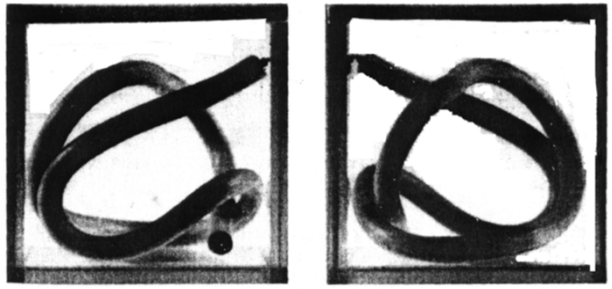
Zeit

1. Messzeitpunkt (t1)

Schlauchfigurentest (Stumpf & Fay, 1983)

- Die Schlauchfiguren sind ein Aufgabentyp zur Erfassung des räumlichen Vorstellungsvermögens bei Jugendlichen und Erwachsenen (Normwerte für das Alter von 15 bis 20 Jahren). Die Durchführungszeit beträgt ca. 12 Minuten.
- Es liegen zwei Parallelformen des Tests vor
- Die Halbierungszuverlässigkeit liegt zwischen .70 und .79, die Äquivalenz der beiden Paralleltests zwischen .71 und .78 (http://www.testzentrale.de/programm/schlauchfiguren.html)

Beispiel:



Reliabilitätsarten: Interne Konsistenz

(z.B. Döring & Bortz, 2016)

- Erweiterung der Testhalbierung: Teilung des Tests in kleinste Einheiten (→ Items)
- Jedes Item = Paralleltest
- Korrelation zwischen Items: wahre Varianz
- Gebräuchlichstes Mass der internen Konsistenz: Cronbach‘s Alpha

→ Mittlere Testhalbierungsreliabilität für alle möglichen Testhalbierungen

– Indikator der Homogenität eines Tests
– Bei mehrdimensionalen Tests: Unterschätzung
– Cronbach‘s Alpha höher je mehr Items und je höhere Iteminterkorrelationen

2. Wie häufig werden Sie und diese Person ärgerlich oder sauer aufeinander?

	selten oder nie	manchmal	häufig	besonders häufig	extrem häufig
Mutter	1	2	3	4	5
Vater	1	2	3	4	5
Freund(in)	1	2	3	4	5
Partner(in)	1	2	3	4	5
Geschwister	1	2	3	4	5
Kind	1	2	3	4	5

Zeit

1. Messzeitpunkt (t1)

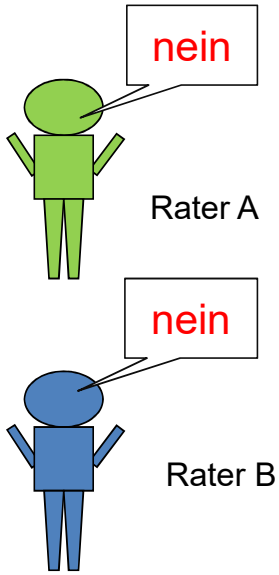
Vor- und Nachteile der verschiedenen Reliabilitätsarten
(angelehnt an Martin, 2008)

Reliabilität	Vorteile	Nachteile	Besonderheiten
Test-Retest	- Gleiche Testitems - Keine extra Arbeit bei Entwicklung	- Erinnerungseffekte - Nicht geeignet für instabile Merkmale - Erfordert zwei Messzeitpunkte - Abhängig vom gewählten Zeitintervall	- Mass für Stabilität eines Merkmals
Paralleltest	- Minimiert Wiederholungseffekte - In Gruppensettings anwendbar - Auch geeignet für Prä-Posttest-Designs	- Verwendung verschiedener Items verringert Reliabilität - Aufwendig in der Entwicklung	- z.B. für Leistungstests in Gruppensettings
Test-halbierung Spezialfall Interne Konsistenz	- Minimiert Wiederholungseffekte - unaufwendig	- Verwendung verschiedener Items verringert Reliabilität - Benötigt längeren Test - Geringere Anzahl Items verringert Reliabilität - Reliabilität auch abhängig von Art der Halbierung	- Mass für die Homogenität

Reliabilitätsarten: Interrater-Reliabilität (z.B. Gravetter & Forzano, 2018)

= Höhe der Übereinstimmungen der Einschätzungsergebnisse unterschiedlicher Beobachter / Testanwender (Rater)

– Interrater-Reliabilität ist hoch, wenn verschiedene Rater bei den gleichen Testpersonen zu gleichen oder ähnlichen Einschätzungen (Ratings) kommen



- Erfordert umfassendes Training der Rater
- Berechnung als Prozent der Übereinstimmung oder Cohen's Kappa (Korrektur für Zufallsübereinstimmung; nur für zwei Rater/Raterinnen)
- Bei mehr als 2 Beurteilenden z.B. Krippendorff's Alpha

1. Messzeitpunkt (t1)

Beispiel Interrater-Reliabilität (Gravetter & Forzano, 2018, S.416)

TABLE 15.8

Data That Can Be Used to Evaluate Inter-Rater Reliability Using Either the Percentage of Agreement or Cohen's Kappa

Two observers record behavior for the same individual over 25 observation periods and record whether they observe aggressive behavior during each period.

Observation Period	Observer 1	Observer 2	Agreement
1	Yes	Yes	Agree
2	Yes	Yes	Agree
3	No	Yes	Disagree
4	No	No	Agree
5	Yes	Yes	Agree
6	Yes	Yes	Agree
7	Yes	Yes	Agree
8	Yes	Yes	Agree
9	Yes	Yes	Agree
10	No	No	Agree
11	No	No	Agree
12	No	No	Agree
13	Yes	No	Disagree
14	Yes	Yes	Agree
15	Yes	Yes	Agree
16	Yes	Yes	Agree
17	Yes	Yes	Agree
18	Yes	No	Disagree
19	Yes	Yes	Agree
20	Yes	Yes	Agree
21	Yes	Yes	Agree
22	Yes	No	Disagree
23	Yes	Yes	Agree
24	Yes	Yes	Agree
25	Yes	Yes	Agree

Prozentuale Übereinstimmung: 84%

Cohen's Kappa = 56.5%

(Formeln und Herleitung siehe Gravetter & Forzano, 2018, S. 412-416)

Quantitative Gütekriterien

- ✓ Objektivität
- ✓ Reliabilität
- → Validität nächstes Mal

Lernziele erreicht?

Am Ende der Veranstaltung …

… sind Sie in der Lage, besondere Herausforderung bei der Messung psychologischer Variablen zu benennen und mögliche Lösungen zu finden.

… wissen Sie, welches die drei Hauptgütekriterien quantitativer Erhebungsmethoden sind, können sie definieren, voneinander abgrenzen und die Zusammenhänge benennen.

… können Sie verschiedene Arten von Objektivität definieren und erklären, wozu diese notwendig sind.

… können Sie verschiedene Arten der Reliabilität definieren, erklären, wann man welche Art der Reliabilität verwenden kann und sollte sowie die jeweiligen Vor- und Nachteile bzw. Besonderheiten benennen.