

## Richtung der Items

---

- Häufig wird die Richtung der Fragen variiert, um dem Problem der Ja-Sage-Tendenz zu begegnen und um stereotype Antwortmuster zu vermeiden (die Variation erfordert erhöhte Aufmerksamkeit).
- „Ich bin ein geselliger Mensch.“
- „Mich an andere anzupassen fällt mir schwer.“
- Vor der Auswertung müssen gegensinnig zum Konstrukt formulierte Items umgepolt werden.

## Reihenfolge der Items

---

- Items mit sehr ähnlichem Inhalt sollten nicht nahe beieinander auftauchen.
- Heikle Fragen sollten am Ende des Fragebogens stehen.
- Vorsicht, wenn vorangegangene Fragen Teilbereiche späterer Fragen ansprechen:
  - „Wie zufrieden sind Sie mit Ihrem Beruf?“
  - „Wie zufrieden sind Sie mit Ihrer Gesundheit?“
  - „Wie zufrieden sind Sie mit Ihrem Bekanntenkreis?“
  - „Wie zufrieden sind Sie mit Ihrem Leben insgesamt?“

---

besser:

- „Alle diese Punkte zusammengenommen, wie zufrieden sind Sie mit Ihrem Leben insgesamt?“

oder

- „Wie zufrieden sind Sie insgesamt mit allen weiteren wichtigen Aspekten Ihres Lebens?“

## Art und Anzahl der Antwortmöglichkeiten

---

„Wie gut finden Sie die Bundespolitik der SPD?“	sehr schlecht	eher schlecht	mittel-mäßig	eher gut	sehr gut
„Wie gut finden Sie die Bundespolitik der SPD?“	sehr schlecht	eher schlecht	eher gut	sehr gut	

## Gerade oder ungerade Anzahl?

---

- Eine ungerade Anzahl ermöglicht eine ausweichende Antwort, ist aber auch als echte, inhaltlich sinnvolle Antwortalternative zu verstehen.
- Bei einer geraden Anzahl von Antwortalternativen werden die Probanden gezwungen, sich für eine Antwortrichtung zu entscheiden und produzieren häufiger fehlende Werte.

## Anzahl der Antwortalternativen

---

- Bis 5 bzw. 7 Stufen verbessern sich Reliabilität, Faktorenreinheit und Informationsausschöpfung.
- Darüber hinaus wird die Differenzierungsfähigkeit der Probanden überfordert.

## Empfehlenswerte Antwortalternativen für gleichabständige 5-stufige Skalen (Rohrman)

<b>Häufigkeit</b>	nie	selten	gelegentlich	oft	immer
<b>Intensität</b>	nicht	wenig	mittelmäßig	ziemlich	sehr
<b>Wahr- scheinlichkeit</b>	keinesfalls	wahr- scheinlich nicht	vielleicht	ziemlich wahr- scheinlich	ganz sicher
<b>Zustimmung</b>	stimmt nicht	stimmt wenig	stimmt mittelmäßig	stimmt ziemlich	stimmt sehr

## Empfehlenswerte Antwortalternativen für 7-stufige Skalen (Ziffern mit Randmarkierung)

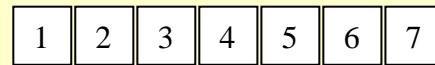
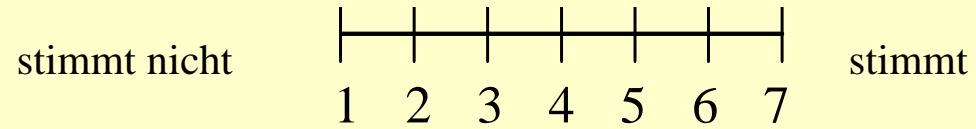
---

Häufigkeit	nie	1	2	3	4	5	6	7	immer
						oder			
	selten	1	2	3	4	5	6	7	oft
Intensität	gar nicht	1	2	3	4	5	6	7	sehr
Wahrscheinlichkeit	keinesfalls	1	2	3	4	5	6	7	sicher
Zustimmung	stimmt								stimmt
	nicht	1	2	3	4	5	6	7	völlig



# Auswahl und graphisches Design der Antwortskala

---



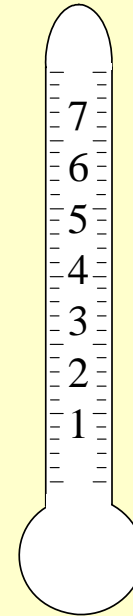
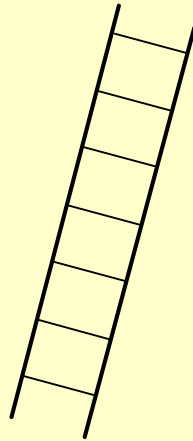
1 2 3 4 5 6 7

○ ○ ○ ○ ○ ○ ○

# Auswahl und graphisches Design der Antwortskala

---

7 stimmt völlig  
6  
5  
4  
3  
2  
1 stimmt nicht



- 
- Senkrechte Antwortskalen sind nicht empfehlenswert, da eine Wertung nahegelegt wird.

## Unipolar oder bipolar?

---

stimmt nicht    1    2    3    4    5    6    7    stimmt völlig

stimmt nicht    -3   -2   -1   0   +1   +2   +3    stimmt völlig

## Unipolar oder bipolar?

---

stimmt nicht	1	2	3	4	5	6	7	stimmt völlig
stimmt nicht	-3	-2	-1	0	+1	+2	+3	stimmt völlig

- Aufgrund der Varianzeinschränkung sind bipolare Skalen in der Regel nicht zu empfehlen, es sei denn, es handelt sich beim zu messenden Konstrukt um eine echte bipolare Variable, wie „Depression und Manie“ oder „Zwanghaftigkeit und Unterkontrolliertheit“.

## Weiß-nicht-Alternative?

---

- Diese ist nur notwendig, wenn Wissen im engeren Sinne abgefragt wird, nicht bei Einstellungs-, Persönlichkeits- oder Selbstkonzeptfragebogen.

„Wer hat ein höheres Einkommen, Sie oder Ihr (Ehe-)partner?“

- Ich habe ein höheres Einkommen.
- Unser Einkommen ist in etwa gleich.
- Mein Partner hat ein höheres Einkommen.
- Ich kenne das Einkommen meines Partners nicht.

## Zustimmung links oder rechts?

---

gar nicht	1	2	3	4	5	6	7	sehr
sehr	1	2	3	4	5	6	7	gar nicht
sehr	7	6	5	4	3	2	1	gar nicht

## Zustimmung links oder rechts?

---

gar nicht	1	2	3	4	5	6	7	sehr
sehr	1	2	3	4	5	6	7	gar nicht
sehr	7	6	5	4	3	2	1	gar nicht

- Viele Befragte empfinden es als angenehmer, sich mit der Leserichtung von links nach rechts gleichzeitig die Dimension „weniger“ bzw. „mehr“ ebenfalls von links nach rechts vorzustellen.



## Gestaltungsfehler

---

nie

eher häufig

sehr selten

sehr häufig

eher selten

immer

## Gestaltungsfehler

---

- |                                   |                                   |
|-----------------------------------|-----------------------------------|
| <input type="radio"/> nie         | <input type="radio"/> eher häufig |
| <input type="radio"/> sehr selten | <input type="radio"/> sehr häufig |
| <input type="radio"/> eher selten | <input type="radio"/> immer       |

- Hierbei ist nicht gewährleistet, dass alle Befragten die Antwortalternativen spaltenweise (statt zeilenweise) lesen.

## Gestaltungsfehler

---

trifft nicht zu

trifft zu

1 2 3 4 5 6 7

## Gestaltungsfehler

---

trifft nicht zu

trifft zu

1 2 3 4 5 6 7

- Hier reicht die verbale Markierung „trifft nicht zu“ von Ziffer „1“ fast bis Ziffer „4“, während die Markierung „trifft zu“ lediglich die Ziffern „6“ und „7“ umfasst.

## Gestaltungsfehler

---

„Gehen Sie bei der nächsten

Bundestagswahl wählen?“     ja     nein

## Gestaltungsfehler

---

„Gehen Sie bei der nächsten  
Bundestagswahl wählen?“     ja    nein

- Hier wäre es besser, die Antwortalternativen weiter auseinander oder untereinander zu platzieren.

## Gestaltungsfehler

---

„Benötigen Sie  
einen Wahlschein – Briefwahlunterlagen  
(Nichtzutreffendes bitte streichen)“.

„Benötigen Sie  
einen Wahlschein – Briefwahlunterlagen  
(Nichtzutreffendes bitte streichen)“.

- Bei derartigen Antwortvorgaben kommt es vor, dass Befragte stattdessen das Zutreffende unterstreichen, da sie die Instruktion nicht genau gelesen haben. Unterstreichung und Durchstreichung sind nicht immer klar zu unterscheiden.



## Gestaltungsfehler

---

besser:

„Welche Unterlagen benötigen Sie?“

Wahlschein                       ja  nein

Briefwahlunterlagen         ja  nein

## Formulierung der Instruktion

---

Die übliche Instruktion weist auf folgende zu beachtende Punkte hin:

- alle Sätze durchzulesen und mit „stimmt/stimmt nicht“ etc. zu beantworten,
- die Sätze nacheinander zu bearbeiten und keinen auszulassen,
- so aufrichtig wie möglich zu antworten – Anonymität wird garantiert,
- auch dann eine Entscheidung zwischen den Antwort-Alternativen zu treffen, wenn dies einmal schwierig erscheinen sollte – es soll stets angekreuzt werden, was vergleichsweise „eher“ zutrifft,
- möglichst zügig zu antworten.

## Formulierung der Instruktion

---

- Da Versuchspersonen in der Regel mit psychologischen Untersuchungen den Begriff „Intelligenztest“ verknüpfen, kann man in der Instruktion zu Persönlichkeits-, Einstellungs- oder Selbstkonzept-Fragebogen diesem Eindruck entgegenwirken, indem man weiter ausdrücklich darauf hinweist, dass es bei dieser Art von Untersuchung keine richtigen oder falschen Antworten gebe.

## Formulierung der Instruktion

---

In bestimmten Fällen, in denen es als notwendig erscheint, die Probanden zusätzlich an konstruktiver Mitarbeit zu interessieren, kann es als günstig erscheinen darauf hinzuweisen,

- dass die Untersuchung lediglich zu Forschungszwecken ausgeführt wird und nicht der Bestimmung individueller Diagnosen mit Konsequenzen für das Individuum dient, sondern dass man sich für die Durchschnittswerte einer größeren Gruppe von Personen interessiert sowie,
- dass ausreichender Datenschutz gewährleistet ist.

## Beispiel 1

---

- „Im folgenden finden Sie eine Reihe von Feststellungen. Bitte antworten Sie, je nachdem ob Sie persönlich meinen, dass eine Feststellung eher zutrifft oder eher nicht zutrifft, mit „ja“ oder „nein“. Sie können dabei gar nichts falsch machen, denn es gibt keine für jede Person zutreffenden Antworten (kein Intelligenztest o.ä.). Antworten Sie bitte aufrichtig und ohne lange zu überlegen; lassen Sie bitte keine Antworten aus. Die Untersuchung dient allein wissenschaftlichen Forschungszwecken. Datenschutz ist gewährleistet; geben Sie bitte ihren Namen nicht an, sondern machen Sie bitte lediglich die Angaben zur Person!“

## Beispiel 2

---

- „Sie nehmen an einer Studie im Rahmen eines Forschungsprojekts in der Fakultät für Psychologie der Universität xy teil. Zunächst bedanken wir uns herzlich für Ihre Bereitschaft, den Fragebogen auszufüllen. Thema der Studie ist die Zufriedenheit der Studierenden mit ihrem Studium. Die Ergebnisse der Studie sollen helfen, Lehre und Beratungsangebote zu verbessern. Im Folgenden sind verschiedene Aussagen aufgeführt. Schätzen Sie bitte ein, wie gut diese Aussagen für Sie persönlich zutreffen. Dabei bedeutet die „1“, dass die Aussage völlig unzutreffend ist, die „7“ bedeutet, dass die Aussage völlig zutreffend ist. Mit den übrigen Ziffern können Sie Ihre Antwort abstufen. Bitte bearbeiten Sie den Fragebogen in der vorgegebenen Reihenfolge. Es ist wichtig, dass Sie keine Antwort auslassen – kreuzen Sie im Zweifelsfall diejenige Ziffer an, die noch am ehesten zutrifft.“

## Zum Abbau von Testangst der Probanden

---

- „Mit dieser Studie wollen wir eine neue Theorie überprüfen.“
- „Manche Fragen werden Ihnen sicherlich sehr ähnlich erscheinen. Dies liegt daran, dass wir mit dieser Studie erst herausfinden wollen, welche Fragen sich am besten eignen.“

## Zusammenstellung größerer Fragebogenbatterien

---

- Mehrere Konstrukte, die in einer Studie gemessen werden, können differenzierter bei geblockter statt vermischter Itemvorgabe gemessen werden.
- Manchmal müssen Antwortvorgaben vereinheitlicht werden, z.B. wenn ein Autor die zustimmende Alternative links, der andere sie rechts gesetzt hat.
- Demographische Angaben an den Anfang oder an das Ende; heikle Fragen zum Schluss.



## Angaben zur Person

---

- Als demographische Daten werden in der Regel zumindest Alter und Geschlecht erhoben, häufig auch Schulabschluss und Beruf.
- Zusätzlich werden untersuchungsspezifische demographische Daten erhoben, die in engem Zusammenhang mit der Forschungsfrage bzw. dem mit dem Fragebogen gemessenen Konstrukt stehen.
- In einer Untersuchung über Paarbeziehungen würden z.B. Fragen nach der Dauer der Partnerschaft gestellt, nach dem Vorhandensein von Kindern, ob man zusammen wohnt und ob man verheiratet ist oder nicht.
- Diese Angaben dienen der Beschreibung der Stichprobe, aber auch der Überprüfung, ob diese Variablen mit den Fragebogenergebnissen in Zusammenhang stehen.

## Befragung einer Stichprobe

---

Bei psychologischen Untersuchungen und Experimenten behilft man sich in der Regel mit anfallenden Stichproben, in der Soziologie legt man mehr Wert auf Repräsentativität:

- einfache Zufallsauswahlen, bei denen aus der kompletten Grundgesamtheit Untersuchungseinheiten ausgewählt werden, was bei Grundgesamtheiten z.B. aus ganz Deutschland jedoch gar nicht möglich ist, da Einwohnerlisten nur auf Gemeindeebene existieren,
- mehrstufige Zufallsstichproben, bei denen z.B. zunächst eine Reihe von Stimmbezirken, daraus dann Haushalte und schließlich einzelne Personen zufällig ausgewählt werden, sowie
- Quotenstichproben, bei denen beliebige Personen befragt werden, die allerdings bestimmte Ausprägungen auf einer oder auf mehreren Variablen haben müssen. Stellt man z.B. fest, dass die Grundgesamtheit aus 50% Frauen und 50% Männern besteht sowie zu je einem Viertel aus vier bestimmten Altersgruppen, muss die Stichprobe diese Verhältnisse abbilden.

## Ausschöpfungsquote erhöhen

---

(Anteil der Stichprobe, die den Fragebogen ausfüllt)

- Nennung eines seriösen Auftraggebers
- Nennung eines interessanten Untersuchungsthemas
- bei schriftlichen Befragungen: telefonische Ankündigung
- hochwertiges Briefpapier
- Rückumschläge mit aufgeklebter Briefmarke

## Vorbereitung der Auswertung

---

Codierung: Nach der Befragung muss als erstes jeder Antwort eine Zahl zugewiesen werden, damit diese in die Datenmaske eingegeben werden kann.

- Ihr Alter \_\_\_\_\_ Jahre
- Ihr Geschlecht    O Mann            O Frau
- Stimmen Sie der Einführung von Studiengebühren zu?  
  O ja            O nein
- Wie stark werden Studiengebühren die Lehre verbessern?  
  gar nicht    1   2   3   4   5    sehr
- Hier: Alter in Jahren, 1 = Mann, 2 = Frau; 1 = ja, 0 = nein;  
  Ziffer 1 bis 5

## Datenkorrektur

---

- Univariate Häufigkeitsverteilungen: Stellt man z.B. fest, dass bei einer Skala von 1 bis 7 eine „8“ eingegeben worden ist oder bei der Variable „Geschlecht“ eine „3“, kann ein solcher Fehler durch einen Blick in den Originalfragebogen korrigiert werden.
- Bivariate Auswertungen: Ein Zehnjähriger kann nicht verheiratet sein bzw. bei einem Erwerbslosen sollte keine Dauer der Betriebszugehörigkeit in den Daten erscheinen.
- Wenn Inkonsistenzen nicht auf Tippfehler zurückzuführen sind, sondern von den Befragten verursacht worden sind: Löschung der Werte und Behandlung als fehlende Werte.

## Umpolung gegensinnig zum Konstrukt formulierter Items

---

„Ich bin ängstlich.“

stimmt nicht    1   2   3   4   5    stimmt völlig

„Ich bin ruhig.“

stimmt nicht    1   2   3   4   5    stimmt völlig

Recode It2 (1=5) (2=4) (3=3) (4=2) (5=1).

## Gesamtwert berechnen

---

- (Vorläufige) Mittelung der Items, die ein Konstrukt messen sollen (statt Summierung)
- Der theoretisch mögliche Wert liegt im gleichen Wertebereich, wie es bei jedem einzelnen Item der Fall ist.
  - Durch die Mittelung werden fehlende Werte kompensiert.

## Item-Analyse

---

- Berechnung der Schwierigkeit (bzw. Leichtigkeit: Mittelwert) und der Varianz
- Bei zweistufigen Items („stimmt nicht / stimmt“) werden Items mit sehr hoher bzw. sehr geringer Schwierigkeit entfernt (unter  $p = 20\%$  und über  $p = 80\%$ ). Bei mehrstufigen Items solche mit einer extremen Schwierigkeit oder Leichtigkeit (keine Faustregel).
- Als optimal wird ein mittlerer Schwierigkeitsgrad angesehen, da dieser die Trennschärfe erhöht. Im Zweifelsfall bei der Itemselektion auch Varianz (sollte hoch sein) und Trennschärfe (sollte hoch sein) beachten.



## Trennschärfe

---

- Korrelation zwischen dem Items und dem Gesamtscore (Summe aller *anderen* Items).
- Der Trennschärfe-Index gibt an, wie stark jedes Item den endgültigen Fragebogen-Wert bereits vorhersagt, wie gut es das gesamte im Fragebogen gemessene Konstrukt repräsentiert.
- Items mit einem niedrigen Trennschärfe-Index werden eliminiert.
- Zuweilen findet man Empfehlungen, nach denen die Trennschärfe jedes Items  $.30$ ,  $.40$  oder  $.50$  überschreiten soll. Solche Empfehlungen sind problematisch, weil das zu messende Konstrukt schon unterschiedlich homogen definiert sein kann.

## Praktische Durchführung

---

- Zunächst wird das Item mit der geringsten Trennschärfe eliminiert, anschließend müssen sämtliche Trennschärfen erneut berechnet werden, da sich mit der Änderung des Gesamtfragebogens (der ja mit den Items korreliert wird) die Trennschärfe jedes Items verändert. Anschließend wird das Item mit der nun geringsten Trennschärfe entfernt usw., bis die angestrebte Itemanzahl erreicht ist bzw. alle Items zufrieden stellende Trennschärfen haben.

## Zu beachten

---

- Die Trennschärfe eines Items hängt nicht allein von seiner Qualität, sondern von der Zusammensetzung des Fragebogens ab. Besteht z.B. ein Fragebogen über konservative Erziehungsstile aus 6 Items zum Thema „Umgang mit Bestrafung“ und 14 Items zum Thema „Wichtigkeit der Einhaltung von Regeln“, kann es sein, dass beide Itemgruppen in sich homogen sind, die Items, deren Inhalte weniger vertreten sind, sich jedoch aufgrund geringerer Korrelationen mit dem Gesamtwert nicht „durchsetzen“ können, weil der Gesamtwert sich vor allem aus den häufig vertretenen Inhalten zusammensetzt. Daher sollte man bereits bei der Itemzusammenstellung auf Ausgewogenheit der Inhalte achten.
- Zudem kann man prüfen, ob nicht eine Konstruktion zweier in sich homogener, aber voneinander verschiedener Fragebogen vorzuziehen wäre. Ob die Items tatsächlich zwei oder mehr voneinander abgrenzbare Dimensionen messen, erfährt man mittels einer Faktorenanalyse.

## Reliabilitätsprüfung

---

- Die beiden wichtigsten Gütekriterien von Fragebogen sind die Reliabilität (Zuverlässigkeit) und die Validität (Gültigkeit). Während sich die Prüfung der Validität mit der Frage beschäftigt, inwieweit der Fragebogen tatsächlich das misst, was er messen soll, beschäftigt sich die Prüfung der Reliabilität mit der Frage, inwieweit der Fragebogen das, was er misst (was immer das sei), genau misst. Man unterscheidet zwei Prinzipien der Reliabilitätsprüfung: die Prüfung der inneren Konsistenz (messen mehrere Fragebogenteile dasselbe?) und die Prüfung der zeitlichen Stabilität (misst derselbe Fragebogen zu mehreren Zeitpunkten dasselbe?).

## Innere Konsistenz

---

Grundidee: Ein Fragebogen misst dann genau, wenn die verschiedenen Fragebogenteile dasselbe Ergebnis erbringen.

- Split-Half-Reliabilität (Korrelation zweier Testhälften)
- Parallelttest-Reliabilität (Korrelation zweier Parallelförmungen)
- Cronbach Alpha (Maß für den Zusammenhang aller Items)

## Retest-Reliabilität

---

Grundidee: Ein Test misst dann genau, wenn er zu mehreren Zeitpunkten dasselbe Ergebnis liefert.

- Korrelation desselben Fragebogen-Gesamtwerts zu verschiedenen Zeitpunkten mit denselben Probanden.  
(ungeeignet bei vorübergehenden Merkmalen, z.B. Stimmung)
- Leistungstests haben Reliabilitäten von .90, Persönlichkeitstests eher .80, bei sehr kurzen Fragebogen ist noch .70 akzeptabel.

## Validitätsprüfung

---

- Ein Fragebogen ist dann valide, wenn er das misst, was er messen soll. In der Regel stehen Fragebogen-Ergebnisse ja nicht „für sich“, sondern sie sollen Rückschlüsse auf das Verhalten der Befragten erlauben, die über die Befragungssituation hinausgehen. Z.B. erwartet man von einer Untersuchung der Wahlabsichten vor einer politischen Wahl, dass das tatsächliche Wahlergebnis in etwa so ausfällt wie in der Umfrage zuvor. Man erwartet von den Teilnehmern eines Berufseignungstests, dass die Personen, die in dem Test gute Ergebnisse erzielt haben, sich später tatsächlich als geeignet für den Beruf erweisen.
- Üblicherweise werden daher noch weitere Variablen ermittelt, die mit den Ergebnissen eines Fragebogens in Zusammenhang stehen sollen, und mit diesen korreliert.

## Inhaltsvalidität

---

- Ein Fragebogen oder Test wird dann als inhaltsvalide angesehen, wenn die Reaktionen in einer Befragungssituation repräsentativ sind für das, was mit dem Fragebogen gemessen werden soll. Soll in einem Betrieb beispielsweise eine Schreibkraft eingestellt werden, die schnell und fehlerfrei tippen kann, müssen die Bewerber in einer Arbeitsprobe demonstrieren, wie schnell und fehlerfrei sie tippen können. In einem solchen Fall entspricht die Aufgabe im Test genau dem Typ von Aufgaben, die der Bewerber auch in Zukunft bewältigen soll.
- Inhaltsvalidität wird auch als Augenscheinvalidität bezeichnet, die oft durch Experten beurteilt wird.



## Kriteriumsvalidität

---

- Bei der Überprüfung der Kriteriumsvalidität wird außer dem Fragebogen (mindestens) eine weitere Variable erhoben, mit der man einen Hinweis auf das zu messende oder vorherzusagende Konstrukt erhält, das so genannte „Kriterium“. Anschließend werden Test und Kriterium miteinander korreliert, die Höhe der Korrelation ist das Maß für die Validität (Anforderung: mindestens Signifikanz, nennenswerte Varianzaufklärung)

### Arten von Außenkriterien

- Verhaltensbeobachtungen
- Fremdeinschätzungen oder
- weitere Fragebogen (Quasi-Kriterium)

## Beispiel

---

- Ein Fragebogen, der Durchsetzungsfähigkeit in Verhandlungssituationen messen soll, kann mittels einer Verhaltensbeobachtung validiert werden, wenn man die befragten Personen in Verhandlungssituationen beobachtet und auch dort ihre Durchsetzungsfähigkeit erhebt. Hierfür gibt es allerdings unterschiedliche Möglichkeiten. Man könnte die Redezeit erfassen, man könnte auszählen, wie oft die betreffende Person die anderen unterbricht oder wie oft Vorschläge der Person von anderen akzeptiert werden. Es gibt also unterschiedliche Möglichkeiten der Validitätsbestimmung, nämlich so viele, wie es sinnvolle Kriterien dafür gibt, was der Fragebogen messen soll.

- 
- Man unterscheidet Kriterien, die gleichzeitig mit dem Fragebogen oder später erhoben werden. Wenn das Kriterium zum selben Zeitpunkt wie der Fragebogen erfasst wird, spricht man von der Überprüfung der *konkurrenten Validität*. Das ist z.B. der Fall, wenn ein Fragebogen zur Aggressivität und eine Fremdbeurteilung eines Freundes über aggressives Verhalten in etwa gleichzeitig erhoben und miteinander korreliert werden. Wird das Kriterium erheblich später erfasst, überprüft man die *prädiktive Validität*. Ein Beispiel dafür ist die Validierung eines Leistungstests, der die Eignung für einen Beruf vorhersagen soll, anhand des Berufserfolgs, der mehrere Jahre später ermittelt wird, etwa an der Note in der Abschlussprüfung, der Beurteilung durch den Vorgesetzten oder dem Einkommen.

- 
- Problem: Auch die Validität der Außenkriterien kann angezweifelt werden.
  - Bei Selbstkonzeptvariablen ist es schwierig bis unmöglich, valide „Außen“kriterien zu finden, denn sie betreffen die vollständig subjektive Sicht einer Person auf sich selbst.

## Experimentelle Validierung

---

- Abhängigkeit von Fragebogenergebnissen von experimentellen Bedingungen, z.B. Instruktionen. Ein Fragebogen misst z.B. dann gut die Tendenz, sozial erwünscht zu antworten, wenn die Ergebnisse unter Ideal-Instruktion („Stellen Sie sich so dar, wie Sie idealerweise sein möchten“) positiver ausfallen als unter Normal-Instruktion.

## Interne Validität

---

- Üblichstes Verfahren einer „internen“ Validitätsanalyse ist die Faktorenanalyse der Fragebogen-Items, also die mathematische Analyse der Interkorrelationen aller Elemente des Fragebogens. Da es sich bei dem Persönlichkeits-, Einstellungs- oder Selbstkonzept-Fragebogen erklärtermaßen um ein eindimensionales Messinstrument handelt, wird die Faktorenanalyse meistens einen sehr starken ersten Faktor ergeben. Wenn sich ein übermächtiger Hauptfaktor ergibt, der wenig Platz für weitere interpretierbare Faktoren lässt, so bestätigt sich die Eindimensionalität des Fragebogens.
- Konstruiert man einen größeren Fragebogen von vornherein so, dass mehrere unterschiedliche, in sich eindimensionale Fragebogen aus dem Item-Pool ableitbar sind, ist also absichtlich von vornherein eine sehr heterogene Item-Sammlung gegeben, so dient die Faktorenanalyse der nachträglichen Identifizierung und damit Validierung mehrerer unterschiedlicher Fragebogen-Dimensionen.

## Konstruktvalidität

---

- Unter einem Konstrukt im hier verwendeten Sinne kann man eine Art Netzwerk von Assoziationen zu einem psychologischen Problem verstehen, die in ihrer Gesamtheit das Problem zu umschreiben vermögen. Man formuliert möglichst viele Hypothesen über Zusammenhänge zwischen dem Konstrukt und anderen Konstrukten / Variablen und überprüft diese empirisch; möglichst mit unterschiedlichen Methoden, nicht nur mit Fragebogen.

## Beispiel: Nationalismus

---

- Es wird vermutet, dass der Nationalismus mit dem Alter steigt und mit dem Bildungsstand sinkt. Es wird vermutet, dass Offiziere der Bundeswehr höhere Werte erzielen als Beamte und diese höhere Werte als Studenten. Es ist zu erwarten, dass der Nationalismus positiv mit Ethnozentrismus, Autoritarismus, Dogmatismus, einer „rechten“ politischen Haltung und Religiosität sowie negativ mit Nonkonformismus korreliert. Man kann die Befragten bitten, Nationalflaggen nach Sympathie in eine Rangreihe zu bringen und den Rangplatz der eigenen Nation ermitteln, oder die Anzahl der als „unsympatisch“ empfundenen fremden Länder oder Völker feststellen, beide Variablen müssten positiv mit Nationalismus korrelieren. Die Gesamtheit der bestätigten Zusammenhänge ergibt das Ausmaß der Konstruktvalidität, es wird also nicht eine einzige Maßzahl ermittelt, sondern eine Sammlung unterschiedlichster Ergebnisse.



## Beispiel: Intimitätsmotiv

---

- Personen mit einem hohen Intimitätsmotiv werden von anderen Personen (in der Fremdbeurteilung) als natürlich, freundlich, verständnisvoll sowie als wenig selbstzentriert beschrieben. In Rollenspielen verhalten sie sich weniger dominant, bringen die Mitspieler häufiger zum Lachen, sagen häufiger „wir“ und positionieren sich selbst in größerer räumlicher Nähe zu den Mitspielern als Personen mit einem geringen Intimitätsmotiv. Außerdem sind sie zufriedener mit ihrer Partnerschaft und aufgrund ihrer sozialen Kompetenz auch erfolgreicher im Beruf.

- 
- Die Hypothesen beziehen sich sowohl darauf, mit welchen Variablen der Fragebogenwert korrelieren soll (konvergente Validität), als auch darauf, mit welchen sie *nicht* korrelieren sollen (diskriminante Validität).
  - Ein Berufseignungstest, den die Teilnehmer vorher üben können, sollte z.B. nicht zu hoch mit der Gedächtnisleistung korrelieren.