

Grundlagen der Testtheorie

WS 2020/21

3. Itemgenerierung und Erstellung eines Testentwurfs
16.11.2020

Prof. Dr. Eunike Wetzel

Schritte der Testkonstruktion

1. Konstruktdefinition
2. Itemgenerierung & Erstellung eines Testentwurfs
3. Empirische Überprüfung des Testentwurfs & Testrevision
4. Validierung
5. Normierung

Schritte der Testkonstruktion

1. Konstruktdefinition
 1. Vorüberlegungen
 2. Eingrenzung des Merkmals
 3. Konstruktdefinition
2. **Itemgenerierung & Erstellung eines Testentwurfs**
 1. Konstruktionsstrategien
 2. Item- und Antwortformate
 3. **Fehlerquellen bei der Itembeantwortung**
 4. Itemformulierung
 5. Erstellung eines Testentwurfs
3. Empirische Überprüfung des Testentwurfs & Testrevision
4. Validierung
5. Normierung

2.3 Fehlerquellen bei der Itembeantwortung

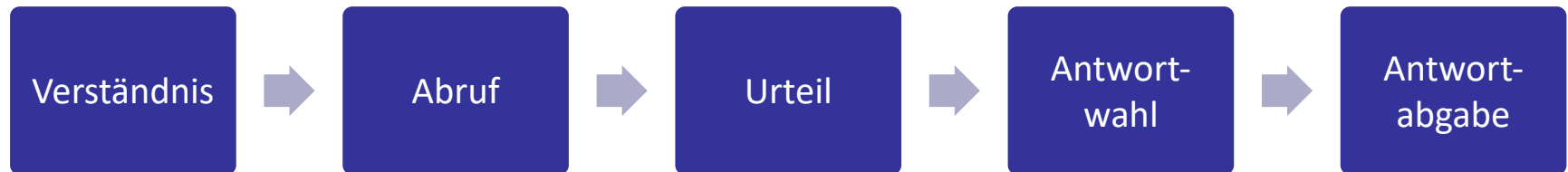
Fehlerquellen

- Proband*innen wählen nicht immer die Antwortkategorie, die ihrer Ausprägung auf dem latenten Konstrukt entspricht
- Man unterscheidet systematische und unsystematische Fehler
- Systematische Fehler erzeugen konstrukt-irrelevante Varianz und verringern dadurch die Validität
- Die systematischen Fehler entstehen während unterschiedlicher Stufen des Antwortprozesses

Kognitiver Prozess bei der Beantwortung von Items

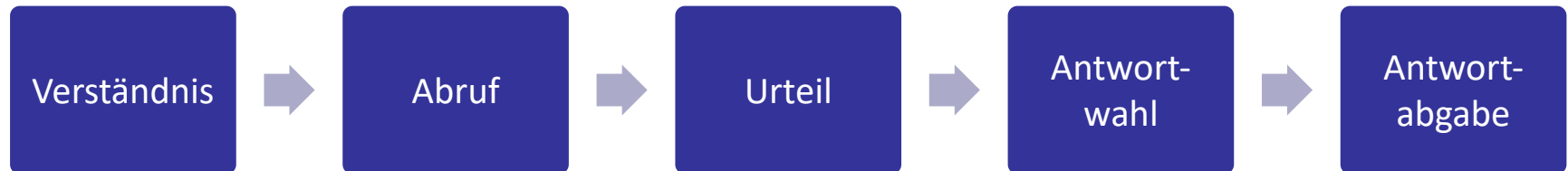
Podsakoff, MacKenzie, Lee & Podsakoff (2003)

- Die Beantwortung von Items erfordert unterschiedliche kognitive Prozesse, die sich 5 Stadien zuordnen lassen:



Kognitiver Prozess & Fehlerquellen

In jedem Stadium des kognitiven Prozesses können Fehler auftreten:



| Fehler | Fehler | Fehler | Fehler | Fehler |
|---|---|--|---|--|
| <ul style="list-style-type: none">• Mehrdeutige Items• Folge: Probanden interpretieren Items unterschiedlich oder antworten zufällig | <ul style="list-style-type: none">• Verschiedene Faktoren können Abruf beeinflussen (z.B. Priming, Itemkontext, Stimmungslage)• Folge: Proband ruft nur einen Teil der relevanten Informationen ab | <ul style="list-style-type: none">• Faktoren: Konsistenzeffekte, Priming, Stimmungslage• Folge: Verzerrung in der Bewertung der abgerufenen Informationen | <ul style="list-style-type: none">• Antwortstile (Bevorzugung/Vermeidung bestimmter Antwortkategorien)• Ankereffekte• Folge: Antwortwahl reflektiert Urteil nicht akkurat | <ul style="list-style-type: none">• Sozial erwünschtes Antworten• Konsistenzeffekte• Folge: Abgegebene Antwort entspricht nicht gewählter Antwort und Urteil |

Optimizing-Satisficing-Modell

- Modell, das erklärt warum manche Proband*innen sich bei der Itembeantwortung von Fehlerquellen beeinflussen lassen und andere nicht
- Krosnick (1991, 1999) unterscheidet zwei Gruppen von Proband*innen, die bei der Bearbeitung verschiedene Strategien verfolgen:
 - **Optimizing**
 - Gründliche Bearbeitung: alle Stadien des kognitiven Prozesses werden vollständig und gewissenhaft durchlaufen
 - Gründe: persönliche Motivation, Wille zu helfen, Belohnung
 - **Satisficing** (kombiniert aus *satisfying* und *sufficing*)
 - Oberflächliche Bearbeitung: die Stadien des kognitiven Prozesses werden nur oberflächlich durchlaufen oder es werden einzelne Stadien ausgelassen
 - Gründe: unfreiwillige oder beiläufige Teilnahme, Belohnung

2.3 Fehlerquellen bei der Itembeantwortung

Fehlerquellen

- 1. Verzerrungstendenzen**
 - 1. Antwortstile**
 - 2. Sozial erwünschtes Antworten**
 - 3. Unaufmerksames Antworten**
2. Motivation
3. Reihenfolgeeffekte
4. Negativ gepolte Items

2.3.1 Verzerrungstendenzen

- Engl. *response biases*
- Antwort wird (intentional oder unbewusst) verzerrt und reflektiert daher nicht akkurat die Ausprägung der Testperson auf dem latenten Konstrukt
- 3 Arten
 1. Antwortstile
 2. Sozial erwünschtes Antworten
 3. Unaufmerksames Antworten

2.3.1.1 Antwortstile

- Def.: systematische interindividuelle Unterschiede in der Verwendung der Antwortskala, die unabhängig von dem Iteminhalt sind
- Beispiele:

Extreme Response Style (Extremkreuzen)

| | Starke Ablehnung | Ablehnung | Neutral | Zu- stimmung | Starke Zustimmung |
|---------------------------|----------------------------------|-----------------------|-----------------------|-----------------------|----------------------------------|
| Ich finde leicht Freunde. | <input checked="" type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> |

Non-Extreme Response Style (Mittelkreuzen)

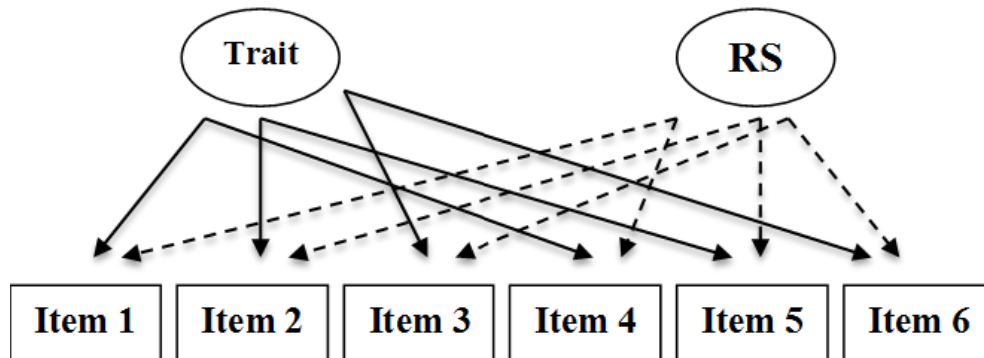
| | Starke Ablehnung | Ablehnung | Neutral | Zu- stimmung | Starke Zustimmung |
|---------------------------|-----------------------|----------------------------------|----------------------------------|----------------------------------|-----------------------|
| Ich finde leicht Freunde. | <input type="radio"/> | <input checked="" type="radio"/> | <input checked="" type="radio"/> | <input checked="" type="radio"/> | <input type="radio"/> |

Akquieszenz (Ja-Sage Tendenz)

| | Starke Ablehnung | Ablehnung | Neutral | Zu- stimmung | Starke Zustimmung |
|---------------------------|-----------------------|-----------------------|-----------------------|----------------------------------|----------------------------------|
| Ich finde leicht Freunde. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> | <input checked="" type="radio"/> |

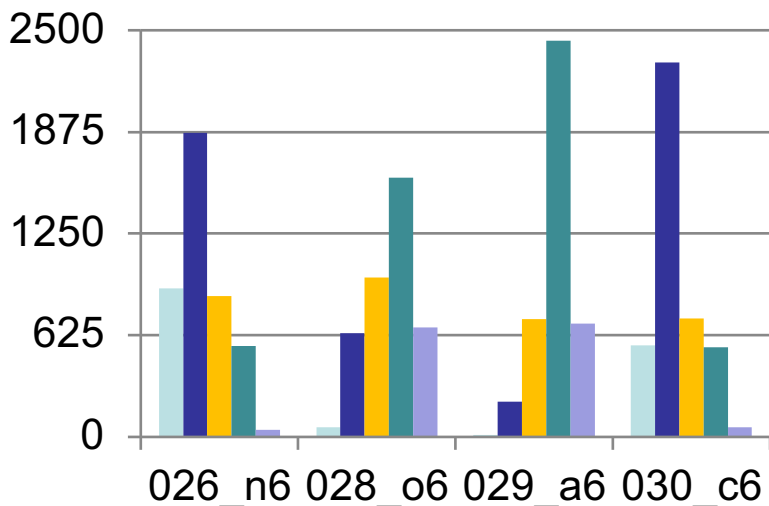
2.3.1.1 Antwortstile

- Wahl der Antwortkategorie wird nicht allein durch die Traitausprägung bestimmt
- Häufiges Auftreten in Fragebögen mit Ratingskalen
- Mögliche Folgen:
 - Verfälschung der Summenscores
 - Verzerrung der Faktorstruktur
 - Verzerrung von Korrelationen



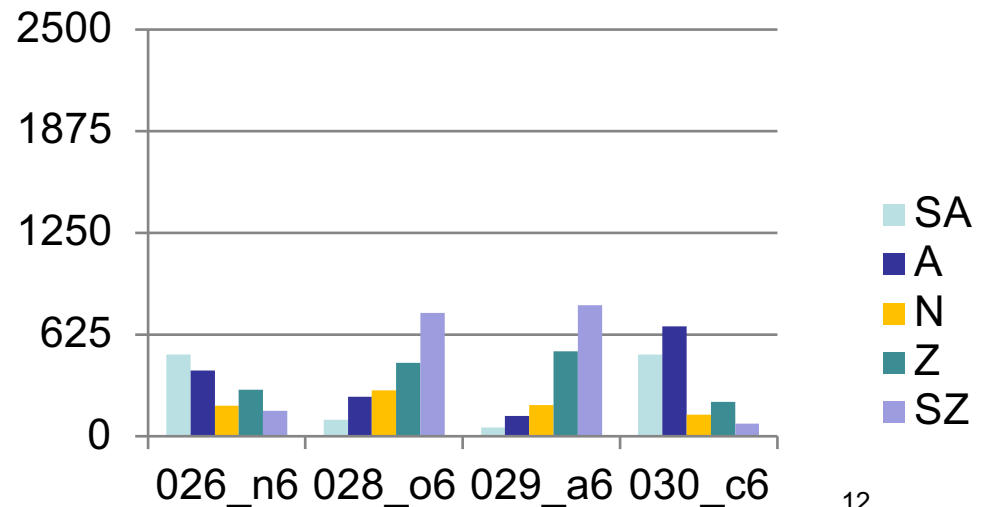
2.3.1.1 Antwortstile

Charakteristische Häufigkeitsverteilungen



Non-extreme Response Style

Extreme response style



2.3.1.1 Antwortstile

Zwei Ansätze zur Konzeptualisierung von Antwortstilen

1. **Kategorialer Ansatz**

Annahme: Antwortstile sind kategoriale Variablen

Zwischen Antwortstilen gibt es qualitative Unterschiede

2. **Dimensionaler Ansatz**

Annahme: Antwortstile sind kontinuierlich verteilt

2.3.1.1 Antwortstile

- Methoden zur Reduktion des Auftretens von Antwortstilen
 - Mittelkategorie weglassen
 - Extreme Bezeichnungen für die beiden Pole verwenden
 - Bezeichnungen für alle Antwortkategorien
 - Forced-choice Format
 - Regeln zu Itemformulierung beachten

2.3.1.2 Sozial erwünschtes Antworten

- Def.: Tendenz zu übermäßig positiven Selbstbeschreibungen
- „Positiv“: in Übereinstimmung mit den sozialen Normen und Werten der Gesellschaft
- Paulhus (2002) unterscheidet zwei Komponenten:
 - **Impression management**
Bewusste Selbstdarstellung, um die Meinung anderer zu steuern
 - **Self-deception**
Unbewusst: Selbstbeschreibungen reflektieren, was Proband*innen tatsächlich über sich denken
- Situationsabhängig: nur als relevant eingeschätzte Eigenschaften werden bewusst verfälscht

2.3.1.2 Sozial erwünschtes Antworten

Pauls & Crost (2005): Veränderungen in den Mittelwerten für fake good, Manager und Krankenpfleger Instruktionen verglichen mit der Standardinstruktion

Table 1
Mean increases from standard to faking conditions and effect sizes

| Domain | Condition | ΔM | SD | t | d |
|--------|-----------|------------|------|---------|-------|
| N- | Fake good | 0.70 | 0.59 | 14.96** | 1.27 |
| | Manager | 1.15 | 0.60 | 17.02** | 2.17 |
| | Nurse | 1.07 | 0.57 | 16.80** | 2.38 |
| E | Fake good | 0.30 | 0.34 | 11.22** | 0.78 |
| | Manager | 0.39 | 0.36 | 9.68** | 1.40 |
| | Nurse | 0.36 | 0.41 | 7.66** | 1.01 |
| O | Fake good | 0.13 | 0.30 | 5.27** | 0.38 |
| | Manager | -0.18 | 0.39 | 4.14** | -0.54 |
| | Nurse | -0.20 | 0.44 | 4.20** | -0.58 |
| A | Fake good | 0.36 | 0.39 | 11.76** | 1.05 |
| | Manager | -0.22 | 0.42 | 4.74** | -0.73 |
| | Nurse | 0.63 | 0.42 | 13.46** | 1.83 |
| C | Fake good | 0.50 | 0.45 | 14.01** | 1.20 |
| | Manager | 0.99 | 0.54 | 16.38** | 1.92 |
| | Nurse | 0.88 | 0.41 | 19.47** | 2.75 |

2.3.1.2 Sozial erwünschtes Antworten

- Methoden zur Detektion sozial erwünschten Antwortens
 - **Spezielle Fragebogen** zur Erfassung der Tendenz zum sozial erwünschten Antworten (z.B. Balanced Inventory of Desirable Responding; Paulhus, 1988)
Beispiel:

| | Trifft nicht zu 0 | 1 | 2 | 3 | Trifft zu 4 |
|--|-------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Ich fluche niemals. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Ich bleibe bei Rot immer an der Ampel stehen, auch wenn kein Auto kommt. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

2.3.1.2 Sozial erwünschtes Antworten

- Methoden zur Detektion sozial erwünschten Antwortens
 - **Spezielle Fragebögen** zur Erfassung der Tendenz zum sozial erwünschten Antworten (z.B. Balanced Inventory of Desirable Responding; Paulhus, 1988)
 - **Validitätsskalen** im Fragebogen (z.B. Lügenskala im MMPI): Verneinen von Items wie „Manchmal werde ich wütend.“, „Gelegentlich tratsche ich über andere.“ als Hinweis auf sozial erwünschtes Antworten

2.3.1.2 Sozial erwünschtes Antworten

- Methoden zur Detektion sozial erwünschten Antwortens
 - **Spezielle Fragebögen** zur Erfassung der Tendenz zum sozial erwünschten Antworten (z.B. Balanced Inventory of Desirable Responding; Paulhus, 1988)
 - **Validitätsskalen** im Fragebogen (z.B. Lügenskala im MMPI): Verneinen von Items wie „Manchmal werde ich wütend.“, „Gelegentlich tratsche ich über andere.“ als Hinweis auf sozial erwünschtes Antworten
 - **Overclaiming-Technique**
Angabe, dass man nicht-existierende Begriffe aus unterschiedlichen Bereichen (Naturwissenschaften, Literatur, Kunst, Sprache...) kennt, deutet auf sozial erwünschtes Antworten hin

2.3.1.2 Sozial erwünschtes Antworten

Wie vertraut sind Sie mit den folgenden Begriffen?

| | noch nie gehört | | | | sehr vertraut |
|-------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Boston Tea Party | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Wiener Kongress | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Sozialmendelismus | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Bäckerrevolution | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| | noch nie gehört | | | | sehr vertraut |
| Aminosäuren | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Floratilien | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Pasteurisation | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Zellulose | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

2.3.1.2 Sozial erwünschtes Antworten

- Probleme mit den Fragebögen zur Erfassung von sozial erwünschtem Antworten:
 - Können selbst verfälscht werden
 - Enthalten Traitvarianz (z.B. Korrelationen mit den Big Five)
 - Erfassen nach Uziel (2010) interpersonell orientierte Selbstkontrolle
 - Nutzen? (z.B. hoher Wert auf der Lügenskala → Sollten Daten dann nicht ausgewertet werden?)
- Reduktion von sozial erwünschtem Antworten
 - Aufklärung über Untersuchungsgegenstand
 - Zusicherung der Anonymität
 - Forced-choice Format?
 - Hinweis, dass intentionale Verfälschung detektiert werden kann?

2.3.1.3 Unaufmerksames Antworten

- Engl. *Careless Responding*
- Def.: Unabhängig vom Iteminhalt werden unaufmerksam oder zufällig Antwortkategorien ausgewählt
- Kann sich äußern in
 - Wiederholungen einzelner Antwortkategorien
 - Wiederholungen von Sequenzen (z.B. Ablehnung – Zustimmung – Ablehnung – Zustimmung – Ablehnung...)
 - Scheinbar zufälligem Antworten

2.3.1.3 Unaufmerksames Antworten

- Möglichkeiten zur Detektion
 - Instruiertes Antworten (instructed response item)

| | starke Ablehnung | Ablehnung | neutral | Zustimmung | starke Zustimmung |
|--------------------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Wählen Sie hier "starke Zustimmung". | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

- Berechnung von Indizes
 - Maximale Anzahl von identischen Antworten nacheinander
 - Konsistenzindex: Korrelationen zwischen Items, die sich inhaltlich sehr ähnlich sind

2.3 Fehlerquellen bei der Itembeantwortung

Fehlerquellen

1. Verzerrungstendenzen
 1. Antwortstile
 2. Sozial erwünschtes Antworten
 3. Unaufmerksames Antworten
- 2. Motivation**
- 3. Reihenfolgeeffekte**
- 4. Negativ gepolte Items**

2.3.2 Motivation

- Testmotivation ist abhängig von
 - Der Einschätzung der Proband*innen, wie relevant die Studie ist (für sie persönlich und für die Gesellschaft)
 - Der Ausprägung der Proband*innen im Trait *need for cognition*
- Die Testmotivation ist umso geringer, je
 - Komplexer die Items formuliert sind
 - Schwerer Items zu beantworten sind
 - Länger der Test dauert

2.3.3 Reihenfolgeeffekte

- Itemantworten müssen unabhängig voneinander sein
- Leistungstests: vorausgegangene Items dürfen keine Lösungshinweise für darauffolgende Items geben
- Persönlichkeitstests: Vermeidung von Konsistenzeffekten („stimmiges“ Antworten) durch Pufferitems oder Vermischung von Items aus unterschiedlichen Subtests
z.B. NEO-FFI Reihenfolge NEOAC

2.3.4 Negativ gepolte Items

- Items, bei denen Zustimmung auf niedrige Traitausprägung hinweist
- Z.B. „Ich bin unglücklich.“ statt „Ich bin glücklich.“
- Werden häufig eingesetzt, um Akquieszenz (Ja-sage-Tendenz) entgegenzuwirken
- Probleme:
 - Sprachlich oft schwieriger zu verstehen als positiv formulierte Items
 - Können Faktorstruktur verfälschen (z.B. wenn positiv und negativ gepolte Items eines Konstrukts zwei verschiedene Faktoren bilden obwohl eindimensionales Konstrukt erfasst werden soll)

Schritte der Testkonstruktion

1. Konstruktdefinition
 1. Vorüberlegungen
 2. Eingrenzung des Merkmals
 3. Konstruktdefinition
2. Itemgenerierung & Erstellung eines Testentwurfs
 1. Konstruktionsstrategien
 2. Item- und Antwortformate
 3. Fehlerquellen bei der Itembeantwortung
 - 4. Itemformulierung**
 5. Erstellung eines Testentwurfs
3. Empirische Überprüfung des Testentwurfs & Testrevision
4. Validierung
5. Normierung

2.4 Itemformulierung

1. Kategorisierung von Itemarten
2. Sprachliche Verständlichkeit
3. Eindeutigkeit des Iteminhalts
4. Varianz des Antwortverhaltens
5. Weitere Aspekte

2.4.1 Kategorisierung von Itemarten

- Es gibt verschiedene Arten, wie in Items die Testperson angesprochen und der interessierende Inhalt erfragt werden kann
- Generell sollten innerhalb einer Skala Vermischungen von Items aus unterschiedlichen Kategorien vermieden werden, da dies zu methodischen Artefakten führen kann

2.4.1 Kategorisierung von Itemarten

- **Direktes oder indirektes Ansprechen des Merkmals**
 - Bsp. direkt: „Sind Sie ängstlich?“
 - Bsp. indirekt: „Fühlen Sie sich unsicher, wenn Sie nachts allein auf der Straße sind?“
 - Direkt angesprochenes Merkmal kann interindividuell unterschiedlich interpretiert werden
 - Gut gewählte Indikatoren erleichtern eindeutige Interpretation
- **Hypothetische vs. biografiebezogene Itemformulierung**
 - Hypothetisch: „Stellen Sie sich vor...“
Problem: Situationen können unterschiedlich interpretiert werden
 - Biografiebezogen: „Wie haben Sie sich verhalten als...“
zuverlässiger, können aber nur für Situationen verwendet werden, mit denen die Testperson Erfahrungen hat

2.4.1 Kategorisierung von Itemarten

- **Konkreter vs. abstrakter Inhalt**
 - Konkret: Situationskomponente
 - Abstrakt: unterschiedliche Interpretationen möglich
- **Personalisierte vs. depersonalisierte Form**
 - Personalisiert: zuverlässiger bei ehrlicher Beantwortung, können aber als Verletzung der Privatsphäre empfunden werden
 - Depersonalisiert: evtl. nur allgemeine, nichtssagende Antworten
- **Stimulusqualität:** emotionale Intensität der Reaktion der Proband*innen
 - Bsp. neutral: „Halten Sie sich für einen ängstlichen Menschen?“
 - Bsp. hohe emotionale Intensität: „Bekommen Sie Herzklopfen, wenn Ihnen jemand nachts auf der Straße folgt?“

2.4.1 Kategorisierung von Itemarten

- **Abgefragte Inhalte**

- Selbstbeschreibung
„Ich lache oft.“
- Fremdbeschreibung
„Meine Freunde halten mich für eine tüchtige Person.“
- Biografische Fakten
„Ich habe mehrmals Abenteuerurlaube gemacht.“
- Trait-/Eigenschaftszuschreibungen
„Ich halte mich für spontan.“
- Motivationale Fragen
„Ich habe eine besondere Vorliebe für Aufgaben, die schwer zu lösen sind.“

2.4.1 Kategorisierung von Itemarten

- **Abgefragte Inhalte**

- Fragen zu Wünschen und Interessen
„Ich schaue gerne wissenschaftliche Sendungen an.“
- Fragen zu Einstellungen und Meinungen
„Es gibt im Leben Wichtigeres als beruflichen Erfolg.“

2.4.2 Sprachliche Verständlichkeit

- Die Klarheit des sprachlichen Ausdrucks hat oberste Priorität
- Items sollten ohne große Mühe bereits nach einmaligem Durchlesen verständlich sein
- Ziel: Iteminhalt sollte von allen Proband*innen in gleicher Weise verstanden werden
- **Regeln**
 - Items positiv formulieren und Negation vermeiden
v.a. doppelte Verneinungen vermeiden!
„Ich bin nicht oft unglücklich.“
 - Einfache Satzkonstruktionen
 - Keine Abkürzungen
„U.u. ist es m.E. legitim, gegen
Friedensbewegungsbefürworter mit Polizeigewalt
vorzugehen.“

2.4.3 Eindeutigkeit des Iteminhalts

- Anpassung der Formulierung an die Zielgruppe
keine (Fach-)Begriffe/Formulierungen, die nur einem kleinen Teil der Zielgruppe bekannt sein könnten
„Ich fühle mich depressiv.“
- Keine mehrdeutigen Begriffe
„Ich bin in Gesprächen angriffslustig.“
- Nur ein Sachverhalt/Gedanke
„Ich fahre sehr gerne und sehr schnell Auto.“
- Verallgemeinerungen vermeiden
„Alle Kinder machen Lärm.“
- Referenzzeitspannen eindeutig definieren
„In letzter Zeit war ich oft niedergeschlagen.“

2.4.4 Varianz des Antwortverhaltens

- Items sollten so formuliert sein, dass Personen mit unterschiedlichen Traitausprägungen unterschiedliche Lösungs- bzw. Zustimmungswahrscheinlichkeiten haben
→ Items mit Decken- oder Bodeneffekten vermeiden
- Ausnahmen:
 - Klinische Tests zur Unterscheidung zwischen klinischer und nicht-klinischer Population
z.B. „Ich denke oft daran, mich umzubringen.“
 - Leistungstests:
 - Items unterschiedlicher Schwierigkeitsgrade zur Differenzierung im unteren und oberen Merkmalsbereich
 - sehr leichtes Item als „Eisbrecher“

2.4.5 Weitere Aspekte

- Aktualität
 - Items sollten so formuliert sein, dass sie nicht schnell veralten
 - Z.B. „Wie heißt der Finanzminister von Deutschland?“
- Keine Wertungen
 - Z.B. „Warum ist es im Allgemeinen besser, einer Wohltätigkeitsorganisation Geld zu geben als einem Bettler?“
- Keine Suggestion
 - Z.B. „Sie stimmen doch zu, dass man mit cholerischen Menschen nichts zu tun haben möchte?“
- Festlegung der Antwortrichtung
 - Festlegen, ob zustimmende oder ablehnende Antwort im Sinne einer hohen oder niedrigen Ausprägung des interessierenden Konstrukts zu interpretieren ist

Itembeispiele

Was ist an den folgenden Items aus dem NEO-PI-R (Ostendorf & Angleitner, 2004) gut oder schlecht?

Ratingskala: *starke Ablehnung* bis *starke Zustimmung*

1. Ich bin nicht leicht beunruhigt.
2. Ich bin dominant, selbstsicher und durchsetzungsfähig.
3. Ich mag Partys mit vielen Leuten.
4. Ich benutze selten Worte wie etwa „phantastisch!“ oder „sensationell!“, um meine Erlebnisse zu beschreiben.
5. Ich bin sehr wissbegierig.
6. Wenn ich das Gefühl habe, dass meine Gedanken in Tagträumereien abschweifen, werde ich gewöhnlich geschäftig und beginne, mich stattdessen auf eine Arbeit oder Aktivität zu konzentrieren.

NARQ

Beurteile bitte, wie sehr die folgenden Aussagen auf dich zutreffen.

Dir steht dazu ein sechsstufiges Antwortformat zur Verfügung (von 1 = "trifft überhaupt nicht zu" bis 6 = "trifft vollkommen zu")

| | trifft über- haupt nicht zu | | | | | trifft voll- kommen zu |
|--|--------------------------------------|---|---|---|---|------------------------------|
| 1. Ich bin großartig. | 1 | 2 | 3 | 4 | 5 | 6 |
| 2. Ich werde einmal berühmt sein. | 1 | 2 | 3 | 4 | 5 | 6 |
| 3. Ich zeige anderen, was für ein besonderer Mensch ich bin. | 1 | 2 | 3 | 4 | 5 | 6 |
| 4. Ich reagiere genervt, wenn eine andere Person mir die Schau stiehlt. | 1 | 2 | 3 | 4 | 5 | 6 |
| 5. Ich genieße meine Erfolge sehr. | 1 | 2 | 3 | 4 | 5 | 6 |
| 6. Es freut mich insgeheim, wenn meine Gegner scheitern. | 1 | 2 | 3 | 4 | 5 | 6 |
| 7. In Gesprächen gelingt es mir meist, die Aufmerksamkeit der Anwesenden auf mich zu ziehen. | 1 | 2 | 3 | 4 | 5 | 6 |
| 8. Ich habe es verdient, als große Persönlichkeit angesehen zu werden. | 1 | 2 | 3 | 4 | 5 | 6 |

Schritte der Testkonstruktion

1. Konstruktdefinition
 1. Vorüberlegungen
 2. Eingrenzung des Merkmals
 3. Konstruktdefinition
2. Itemgenerierung & Erstellung eines Testentwurfs
 1. Konstruktionsstrategien
 2. Item- und Antwortformate
 3. Fehlerquellen bei der Itembeantwortung
 4. Itemformulierung
 - 5. Erstellung eines Testentwurfs**
3. Empirische Überprüfung des Testentwurfs & Testrevision
4. Validierung
5. Normierung

2.5 Erstellung eines Testentwurfs

- **Reihenfolge der Items**
 - Bei Leistungstests oft von leicht bis schwierig ansteigend
 - Bei Persönlichkeitstests Items aus unterschiedlichen Facetten mischen oder (bei Onlinetestung) randomisierte Darbietung der Items
- **Instruktion**
 - Soll Proband*innen zur Mitarbeit anregen
 - Wichtige Bestandteile
 - Klare Handlungsanweisung und Erläuterung des Antwortformats
 - Je nach Test ein Beispielitem und eine Beispielantwort
 - Anweisung spontan und wahrheitsgetreu zu antworten und keine Items auszulassen
 - Hinweis auf Anonymität
 - Evtl. Pilotierung mit unterschiedlichen Instruktionen, um zu überprüfen, welche am besten verstanden wird

2.5 Erstellung eines Testentwurfs

- **Demografische Angaben**

- Können am Anfang oder am Ende des Tests erhoben werden
- Sind auf notwendige Auskünfte zu beschränken

- **Layout**

- Sprachlich und optisch ansprechend
- Auf Zielgruppe angepasst
- Übersichtlichkeit
- Sollte Bearbeitung erleichtern

Itembeispiele Onlinetestung

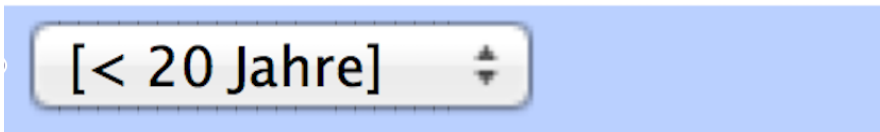
5. Where would you recommend your students to purchase the text?*

- ☒ Amazon
- ☐ Local Campus Bookshop
(Please provide address)
- ☐ Local High Street Bookshop
(Please provide address)
- ☐ Other
(Please provide details)



Itembeispiele Onlinetestung

1. Wie alt sind Sie?



[< 20 Jahre]

1. Wie alt sind Sie?



✓ [< 20 Jahre]
< 20 Jahre
20 - 25 Jahre
26 - 30 Jahre
>31 Jahre

Itembeispiele Onlinetestung



There was an error on your page. Please correct any required fields and submit again.

[Go to the first error](#)



This question is required

On the scale below, how would you roughly rate your mood at this present moment. *

Very Sad
(Negative)

0

☐

1

☐

2

☐

3

☐

4

☐

5

☐

6

☐

7

☐

8

☐

9

☐

Very
Happy
(Positive)

10

☐

2.5 Erstellung eines Testentwurfs

- Pilotstudie mit einer kleinen Stichprobe zur ersten Erprobung des Testentwurfs
- Methoden
 - **Retrospektive Befragung**
Bei welchen Items gab es Probleme?
Wie haben Sie die Instruktion verstanden?
 - **Interview des Testleiters (Debriefing)**
Bei welchen Items hat Testleiter*in Probleme beobachtet?
 - **Verhaltenskodierung**
Dritte Person beobachtet Testleiter*in und Proband*in und notiert instruktionsverletzendes und nicht erwartungskonformes Verhalten
 - **Kognitives Vortesten/Technik des lauten Denkens**
 - Proband*innen sollen Gedanken, die sie während der Bearbeitung haben, formulieren und laut äußern
 - Ermöglicht es, Verständnis- und Interpretationsschwierigkeiten sowie Probleme bei der Anwendung des Item- und/oder Antwortformats zu entdecken

Schritte der Testkonstruktion

1. Konstruktdefinition
 1. Vorüberlegungen
 2. Eingrenzung des Merkmals
 3. Konstruktdefinition
2. Itemgenerierung & Erstellung eines Testentwurfs
 1. Konstruktionsstrategien
 2. Item- und Antwortformate
 3. Fehlerquellen bei der Itembeantwortung
 4. Itemformulierung
 5. Erstellung eines Testentwurfs
3. Empirische Überprüfung des Testentwurfs & Testrevision
4. Validierung
5. Normierung

Literatur zu dieser Sitzung

- Prüfungsrelevant:
 - Moosbrugger & Kelava (2012): Kapitel 3.4 bis 3.7