

F/F2 Diagnostische Methoden und Verfahren

Termin 2

Gesichtspunkte der Testkonstruktion

Sommersemester 2024

M.Sc. Leona Wahnschaffe

Ablauf

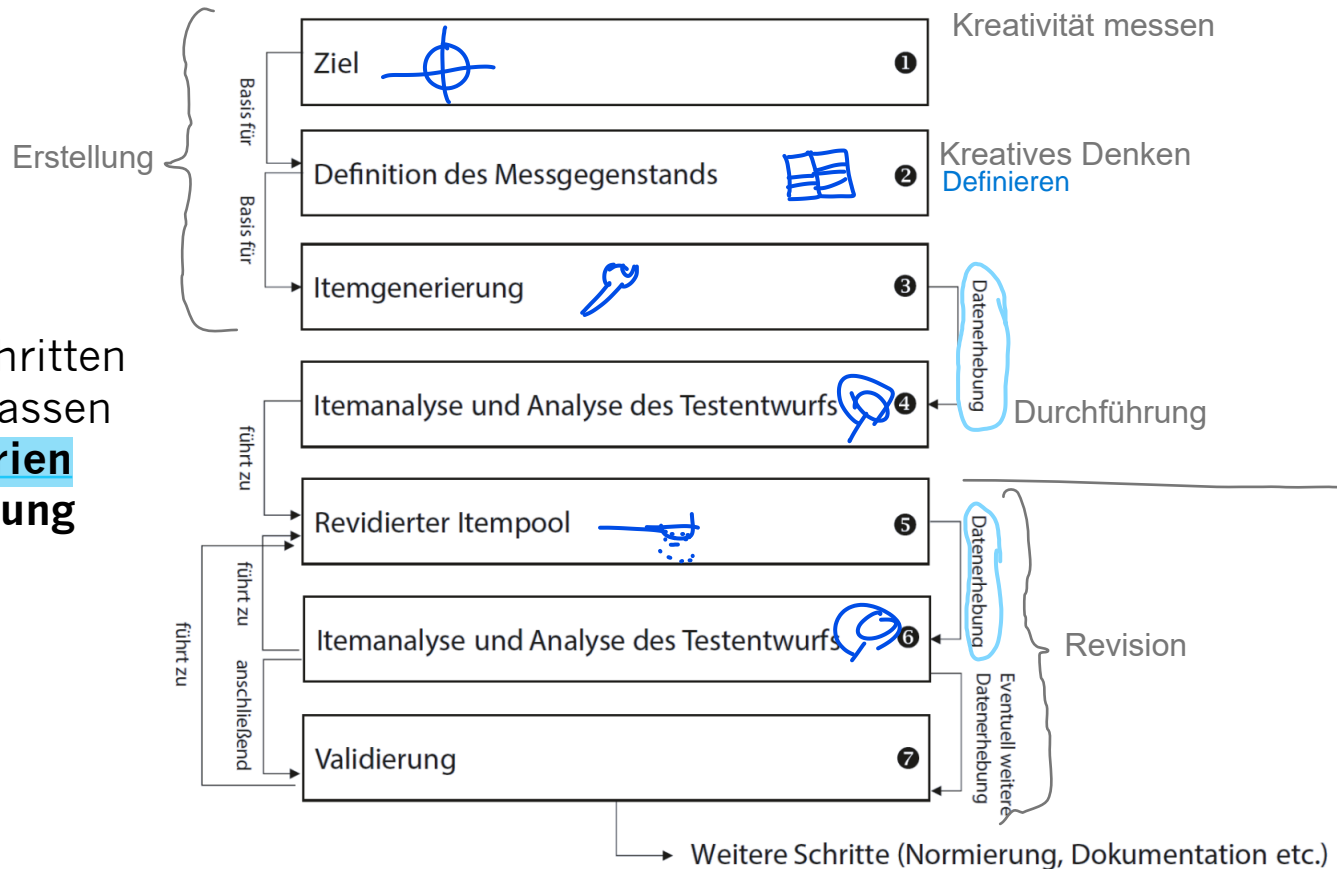
- › Konstruktionsprinzipien psychologischer Tests
- › Itemgenerierung
 - Vorgehensweisen der Itemgenerierung
 - Aufgaben- und Antworttypen
 - Gesichtspunkte der Itemformulierung



Konstruktionsprinzipien psychologischer Tests

Schritte der Testkonstruktion

Basierend auf den Schritten der Testkonstruktion lassen sich **Bewertungskriterien** für die **Güte** und **Passung** psychologischer Tests ableiten!



Ziele von psychologischen Tests

1. Messung eines psychologischen Merkmals

Indikatoren f. psych. Störungen

- Z.B. Ausprägung von Persönlichkeitsmerkmalen, Intelligenz, Lebenszufriedenheit

2. Prüfung von definierten Kriterien und Zuordnung von Personen zu Kategorien

'kann die Person sicher Auto fahren'

- Z.B. Prüfung bestanden vs. Nicht bestanden, Kompetenzstufen

3. Vorhersage von Gegebenheiten außerhalb der Testsituation

{Ist eine Person für bestimmtes Studienfach / Beruf geeignet}

- Z.B. Studienorientierung, Berufseignung

- › Mit diesen Zielen geht ebenso einher festzulegen, welchen Geltungsbereich der Test hat und welche Zielpopulation er anspricht

Definition des Messgegenstandes

- › **Messgegenstand** = psychologisches Merkmal, dessen Ausprägung durch einen Test beschrieben werden soll (Konstrukt)
- › Definition des Messgegenstandes kann auf ^{Test der die Big-five misst} **bestehenden Theorien** und **Fachliteratur** aufbauen oder **eigenständig konstruiert** werden
_{Studienzufriedenheit, Definition selber festlegen}
 - Z.B. mittels Literaturrecherchen, Expertenbefragungen, Betroffenenbefragungen
- › Je nach Zielsetzung des Tests sind weitere Definitionen notwendig:
 - Kriteriumsorientierte Testkonstruktion: Festlegung von Kategorien und Identifikation relevanter Kriterien _{Was muss eine autofahrende Person alles Können}
 - Wenn es das Ziel ist, reale Gegebenheiten vorherzusagen, müssen relevante Bedingungen in der Realität abgeleitet werden → **Anforderungsanalyse**
 - Anforderungen an bestimmte Bedingungen in einer Situation _{Arbeitsplatz}

Empfehlungen des Diagnostik- und Testkuratoriums zur Bewertung von Testmanualen (2018):

- › Schließt der Test an eine **bestehende Theorie** an oder entwickeln die Testautor(inn)en eine **eigene Theorie**?
- › Wird diese **Theorie** ausreichend **beschrieben**? Wird das **Konstrukt** hinlänglich **beschrieben**?
- › Wird deutlich, was und was nicht zu dem zu **messenden Bereich** gerechnet wird?
- › Wird beschrieben, was die **Unterschiede und Gemeinsamkeiten** gegenüber **Tests mit überlappendem Geltungsanspruch** sind?
- › Wird angegeben, was auf theoretischer Ebene/auf der Ebene des Aufgabenmaterials der **Mehrwert** des neuen Instruments über bestehende Instrumente hinaus ist?
- › Wird deutlich, ob ein **beliebiges Item** zum Test gehören könnte oder nicht?
- › Werden das oder die zu messende(n) Konstrukt(e) auf solche Weise analysiert, sodass deutlich wird, welche **Aspekte innerhalb des Konstrukts** oder der Konstrukte unterschieden werden können?

Weitere Planungsaspekte der Testkonstruktion

Testlänge und Testzeit

- › **Testlänge:** Anzahl der Items eines Testes/Fragebogens
 - Angemessene Länge abhängig von **Geltungsbereich und Zielgruppe des Tests** (je breiter Konstrukt, desto länger der Test)
 - Generell: **Je mehr Items, desto höhere Messgenauigkeit** (**Reliabilität**)
 - **ABER:** Ab gewisser Testlänge **kein zusätzlicher Zuwachs** an Reliabilität + **Abnahme an Validität** durch wachsenden **Einfluss von testfremden Variablen** (z.B. Konzentration & Motivation)
 - Je besser die **Qualität** der verwendeten Items, desto **weniger Items** sind für reliable und valide Messung notwendig!
- › **Testzeit:** **Zeitdauer**, die für die Bearbeitung der Testaufgaben vorgesehen ist
 - Starker Zusammenhang mit Testlänge → ebenfalls v.a. abhängig von **Testart, Geltungsbereich und Zielgruppe**
 - **zielgruppenbedingte Einschränkungen** müssen berücksichtigt werden (z.B. bei Tests für Schulkinder)
 - Bei Leistungstests abhängig davon, ob es sich um **Niveau- oder Geschwindigkeitstest** handelt

Weitere Planungsaspekte der Testkonstruktion

Testadministration

"Wie soll der Test durchgeführt werden"

Durchführungsart	Paper-Pencil	Am weitesten verbreitet , nicht auf technische Geräte angewiesen
	Computerbasiert	Hohe Durchführungs- und Auswertungsobjektivität , ökonomischere Auswertung, neue Itemformate (Video, Audio) möglich, vereinfachtes adaptives Testen , Online-Erhebungen möglich
Form der Testung	Einzeltestung	Sorgfältigere Überwachung der Testung, aber sehr aufwändig
	Gruppentestung	Wesentlich ökonomischer , aber fehler- und verzerrungsanfälliger (insb. Bei Online-Testungen)
Bearbeitungsform	Selbstbericht	besonderer Zugang zu Merkmalen, die nur durch eine Innensicht erfasst werden können, also z. B. Gefühle, Motivationen, Einstellungen
	Fremdbericht	Dann sinnvoll, wenn davon ausgegangen wird, dass Testpersonen kein ausreichendes Einsichtsvermögen oder Verständnis verfügen oder wenn starke Antwortverzerrungen zu erwarten sind

Itemanalyse und Testrevidierung

- › Nach der **erstmaligen Itemgenerierung** werden die neuen Items hinsichtlich ihrer **Eigenschaften** und **Brauchbarkeit** für den Test/Fragebogen analysiert
- › Analyse möglichst **mit einer Stichprobe**, die für die **Zielstichprobe** des Tests **repräsentativ** ist
- › Analyseschwerpunkte (nach KTT):
 - **Itemschwierigkeit** (möglichst gemischte Schwierigkeiten)
 - **Itemvarianz** (möglichst hoch)
 - **Itemtrennschärfe** (idR sollte sie größer als .3 sein; Zielwerte abhängig davon, wie homogen das Merkmal ist)
 - **Itemladungen** auf Faktoren (Items soll nur das interessierte Merkmal/ entsprechenden Subskalen abbilden)
 - **Itemvalidität** (Items sollten reale Gegebenheiten abbilden, sofern dies der Messanspruch des Tests ist)
 - **Testgütekriterien**
- › Bei Tests, die nach **probabilistischen Testtheorien** entwickelt sind, idR Überprüfung, inwiefern die Items jeweils die Modellannahmen widerspiegeln (**Itemfitmaße**)
- › Entscheidungen bei der Itemselektion am besten aufgrund der **Kombination verschiedener Kennwerte**



Itemgenerierung

Ziele und Aspekte der Itemkonstruktion

- › Um **interindividuelle Unterschiede** in **psychologischen Merkmalsausprägungen** messen zu können, braucht es **repräsentative Aufgabenstellungen**
- › Wichtig, dass die Aufgaben das **gesamte Merkmal** in **all seinen Facetten abbilden**
 - Z. B. Extraversion komplett
 - Geselligkeit
 - Dominanz
- › Aufgabenstellung aus **zwei Teilen** zusammengesetzt:
 - **Itemstamm/Item**: Die Testbestandteile, die Reaktionen hervorrufen sollen (Fragen, Aufgaben, Bilder, etc.)
 - **Antwortformat der Aufgabe**: Festlegung, in welcher Form die Testpersonen auf die Items reagieren sollen

Berücksichtigung der Zielgruppe bei der Itemgenerierung

- › **Itemstamm** muss so gewählt werden, dass Testpersonen mit **unterschiedlichen Merkmalsausprägungen** bei der Bearbeitung der Items auch **Unterschiede im Antwortverhalten** aufweisen
 - Daher **extrem leichte/schwere Aufgaben** bzw. Fragen, die **leicht bejaht/verneint** werden, **vermeiden!** "Fallschirmspringen ist mein Lieblingshobby"
 - **Ausnahme:** wenn **besonders starke/ schwache Merkmalsausprägungen** mit **sehr seltenen, extremen Verhaltensweisen** erfasst werden sollen → Itemschwierigkeit nach **Zielgruppe** richten
- › Bei **Leistungstests** je nach Testart unterschiedliches Realisieren der **Korrespondenz zwischen Itemschwierigkeit und Merkmalsausprägungen:**
 - **Powertest:** Items **aller Schwierigkeiten** müssen ausreichend vorhanden sein
 - **Speedtest:** Viele einfache Items mit **geringer Itemschwierigkeit**, die von (fast) allen Testpersonen gelöst werden können

Vorgehensweisen der Itemgenerierung

Man kann methoden kombinieren, zb erst deduktiv, dann induktiv

Deduktive Methode

Items werden auf Basis der dem Test zugrundeliegenden **Theorien** gewählt

= "rationale Methode"

Induktive Methode

Items werden aufgrund ihrer **Korrelationen/Ähnlichkeit** zueinander gewählt

sammelt aus der beispielsweise Literatur mögliche relevante Items

Kriteriumsorientierte Methode

Items werden danach gewählt, wie gut sie vorab definierte **Kriterien abfragen**

- Vorab definierte Kriterien die erfüllt werden sollen
- Symptome einer Erkrankung
- fühlerscheinkriterien, dass man den FS bekommt
- Items so wählen dass sie Kriterien abbilden

Externale Methode

Items werden anhand ihrer **Vorhersagekraft** für in der **Realität** bereits existierende Gruppen/ Merkmale gewählt

- hinsichtlich welcher Fragen unterscheiden sich Gruppen in ihrer Itembeantwortung
- was sind verschiedene Interessen von Psychologiestudenten vs. Mathestudenten
- daraus Schlussfolgern was pbn studieren sollte auf Grundlage seiner Interessen

Alle Methoden können natürlich auch miteinander **kombiniert** werden!

Deduktive Itemgenerierung

- › Basierend auf einer Theorie, die das zu erfassende Merkmal gut beschreibt, werden Items formuliert
- › Items sollten **alle (Sub-)Dimensionen** (sofern vorhanden) des Merkmals abbilden
- › Items werden dahingehend ausgewählt, inwiefern sie die **dimensionale Struktur** des Konstruktes **abbilden** zB. Extraversion: Dominanz vs. Geselligkeit
- › Items sollten möglichst nur das intendierte Konstrukt messen (hohe **Fehleranteile vermeiden**)
→ Überprüfung z.B. mittels **Faktorenanalyse**

Wichtig: Nur, weil deduktive Tests explizit auf Theorien basieren, heißt das nicht, dass die anderen Methoden nicht ebenfalls theoriefundiert vorgehen! Psychologische Tests sollten generell auf **wissenschaftlichen Befunden und Theorien** basieren!

Induktive Itemgenerierung

Grundgedanke: Items, die hoch miteinander **korrelieren**, erfassen mit hoher Wahrscheinlichkeit dasselbe **latente Konstrukt**

Typisches Vorgehen:

- › Items werden aus **bereits existierenden Messinstrumenten** oder **Forschungsarbeiten**, die dasselbe Konstrukt zu messen scheinen, gesammelt - {Items die Aggressivität messen}
- › Redundante Items werden aussortiert
- › Testfragebogen wird an neuer Stichprobe validiert
- › Bei mehrdimensionalen Merkmalen wird die **Faktorstruktur** geprüft

Prototypenansatz: Verhaltensweisen oder Adjektive werden dahingehend beurteilt, wie **typisch** sie für das zu messende Merkmal sind

Induktive Itemgenerierung

Beispiel: Freiburger Persönlichkeitsinventar (FPI/ FPI-R)

- › FPI-R von Fahrenberg et al. (2020), 1. Auflage (FPI) schon 1970
- › in der Praxis viel verwendeter **Persönlichkeitsfragebogen**, der die Persönlichkeit in eigenständigen Eigenschaftskonzepten, losgelöst von den Big 5, erfasst
- › Testentwicklung:
 - Zunächst **deduktive Skalenkonstruktion** (es wurden **anfangs bestimmte Konstrukte festgelegt** und für diese dazu passende Item **entwickelt**)
 - Anschließend weitere Itemauswahl teils nach **faktorenanalytischen**, teils nach **Trennschärfeprinzipien**
- › Testaufbau: 138 Items auf 10 Standardskalen mit jeweils 12 Items + 2 Zusatzskalen (Emotionalität, Extraversion)

Kriteriumsorientierte Itemgenerierung

- › Nicht die Position einer Person in Relation zu einer Vergleichsnorm, sondern das **Erreichen** oder **Verfehlen** eines konkreten **Kriteriums** soll ermittelt werden
- › Items sollen so formuliert und ausgewählt werden, dass sie **alle für das Kriterium relevanten Bereiche** gleichmäßig abdecken
 - **Leistungs-/ Wissenstest**: damit eventuelle **Lücken** auffallen/ das Ergebnis nicht verzerren
 - **Klinische Diagnostik**: Alle **Symptombereiche** abfragen
- › z.B. PISA Tests, F-Klausur
 - Werden bestimmte Kriterien erreicht oder Verfehlt
 - Bei Leistungstests:
 - Alls Symptome einer Erkrankung {Screening für Agora-Phobie}
 - Fragebogen soll alle möglichen Symptome einer Agora-Phobie abdecken

Kriteriumsorientierte Itemgenerierung

Beispiel: Agoraphobie nach ICD-10 (F40.0)

- 1. Furcht oder Vermeidung von mind. 2 Situationen:**
 - Menschenmengen, Öffentliche Plätze, Alleine Reisen, Reisen mit weiter Entfernung von Zuhause
- 2. In der gefürchteten Situation mind. 2 Angstsymptome (davon mind. 1 vegetatives)**
 - **Vegetative Symptome:** Herzklopfen, Schweißausbrüche, Tremor, Mundtrockenheit
 - **Andere Symptome:** Atembeschwerden, Beklemmungsgefühl, Brustkorbschmerzen oder –missempfindungen, Gefühl von Schwindel, Schwäche, Benommenheit, etc.
- 3. Deutliche emotionale Belastung durch das Vermeidungsverhalten / Angstsymptome. Einsicht, dass diese übertrieben sind**
- 4. Symptome ausschließlich in der gefürchteten Situation oder bei Gedanken an diese**
- 5. Ausschluss: Furcht und Vermeidung nicht durch**
 - Organische psychische Störungen (F0)
 - Wahn oder Schizophrenie (F2)
 - affektive Störung (F3)
 - Zwangsstörung (F42)

Externale Itemgenerierung

- {Person mit hoher Studienzufriedenheit vs. Person mit niedriger Studienzufriedenheit}
- MMPI
- Erfassung psychopathologische Symptom 'Depression, Schizophrenie..'

Typisches Vorgehen:

1. Identifikation von Personen, die **relevanten Gruppen** angehören/ das zu untersuchende **Merkmal verkörpern**
2. **Gruppenangehörige und Nicht-Angehörige** bearbeiten möglichst großen und breitgefächerten Itempool
3. **Items**, die besonders gut zwischen Gruppenangehörigen und Nicht-Angehörigen **diskriminieren**, werden für vorläufige **Testversion** verwendet
4. **Kreuzvalidierung** der diskriminanten Items an einer neuen Stichprobe von Gruppenangehörigen und Nicht-Angehörigen

Ähnliches Vorgehen, wenn der Test statt einer Gruppenzugehörigkeit **kontinuierliche Werte** (z.B. Abschlussnoten) vorhersagen soll:

Prüfung, welche Items das Merkmal **statistisch signifikant vorhersagen**

Externale Itemgenerierung

Beispiel: MMPI-2

MMPI-2: Minnesota Multiphasic Personality Inventory-2 (Hathaway, S. R. & McKinley, J. C., 1998; deutsche Fassung von Engel, 2000)

- › Ursprüngliche Version in 1943 veröffentlicht (deutsche Übersetzung „MMPI-Saarbrücken in 1963) zur Erfassung **psychopathologischen Symptomen**
 - Anfangs Liste von 1.000 Items zu allgemeiner Gesundheit, familiären und ehelichen Beziehungen, sexuellen und religiösen Einstellungen sowie emotionalen Zuständen
 - Bearbeitung der Items durch **Patientinnen und Patienten** mit psychiatrischen Diagnosen und „unauffällig-normalen“ **Kontrollpersonen**
 - Auswahl von **566 Items** (567 in MMPI-2), wonach sich Patientinnen und Patienten **signifikant** von den Kontrollpersonen **unterscheiden**
 - Zuweisung der Items zu **Skalen** basierend auf Antwortwahrscheinlichkeiten (Items sind teilweise **mehreren Skalen** zugehörig)
- › **Revidierte Version:** Überarbeitung von Formulierungen, Streichungen, Hinzufügen neuer Items, Erweiterung der Skalen und Subskalen, ausgeprägtere Normierung

Externale Itemgenerierung

Beispiel: MMPI-2

Beispielitems:

Es fällt mir schwer, meine Gedanken bei einer Aufgabe oder Arbeit zu behalten.

Ich habe nie Blut erbrochen oder gehustet.

Ich höre seltsame Dinge, wenn ich alleine bin.

Ich war nie in jemanden verliebt.

Ich wünschte, ich wäre nicht so schüchtern.

Manchmal komme ich mir wirklich nutzlos vor.

Zu welchen dieser drei Skalen würdet ihr diese Items jeweils zuweisen?

Depression

Schizophrenie

Soziale Introversion

Externale Itemgenerierung

Beispiel: MMPI-2

Depression

- Manchmal komme ich mir wirklich nutzlos vor.
- Ich habe nie Blut erbrochen oder gehustet.

Schizophrenie

- Ich höre seltsame Dinge, wenn ich alleine bin.
- Ich war nie in jemanden verliebt.

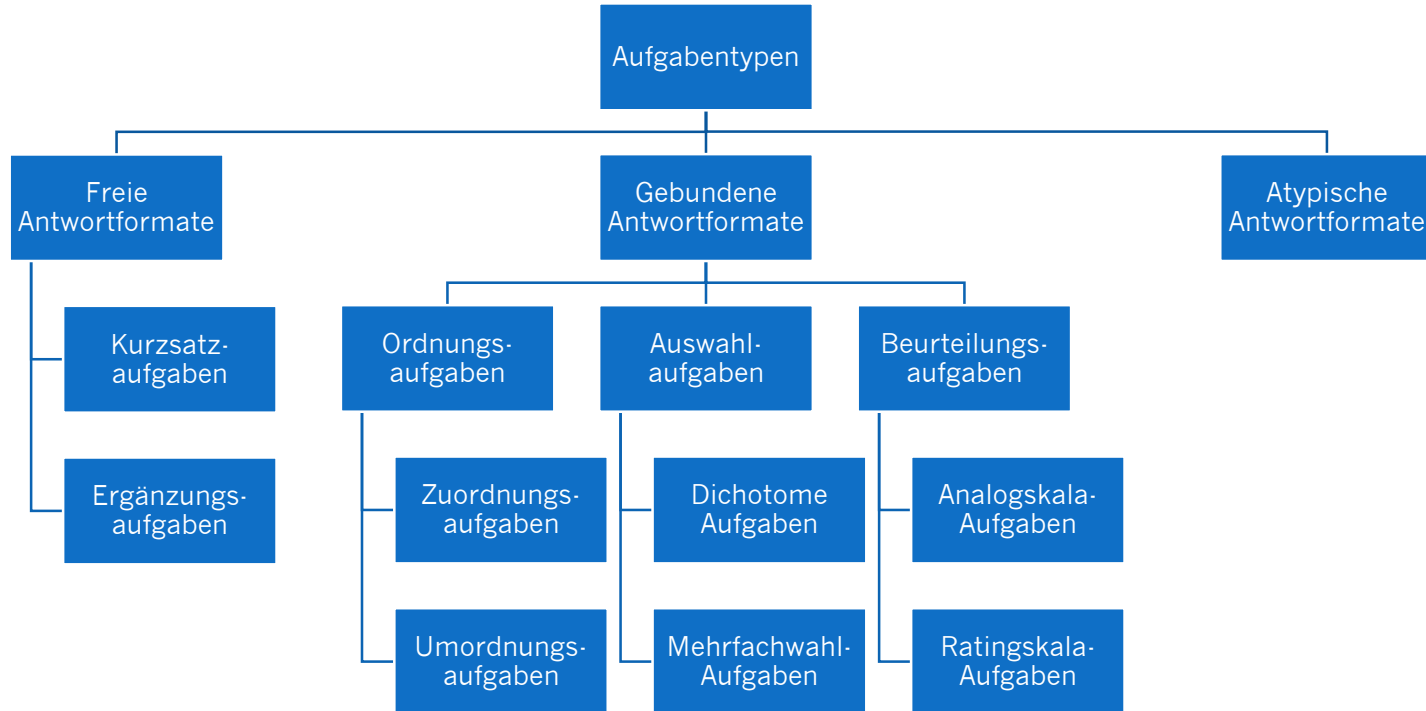
Soziale Introversion

- Ich wünschte, ich wäre nicht so schüchtern.
- Es fällt mir schwer, meine Gedanken bei einer Aufgabe oder Arbeit zu behalten.

Vor- und Nachteile von Itemgenerierungsverfahren

	Deduktiv	Induktiv	Kriteriumsorientiert	External
Vorteile	<p>Hohe wissenschaftliche Güte</p> <p>Gute Interpretierbarkeit von Testergebnissen aufgrund theoretischer Grundlagen</p>	<p>Meist hohe statistische Güte (hohe Konsistenzen)</p> <p>keine theoretischen Vorannahmen notwendig - Theorie wird aus Daten abgeleitet</p>	<p>Keine/ wenig Normierung notwendig</p>	<p>Hoher Realitätsbezug</p> <p>Bietet auch bei geringem theoretischem Vorwissen präzise Messung von Merkmalen</p>
Nachteile	<p>Setzt voraus, dass zum Messgegenstand bereits Theorien/ Forschung existieren</p> <p>Paradigmagebundenheit</p>	<p>Interpretation der extrahierten Dimensionen/ Skalen schwierig</p>	<p>Ergebnisse abhängig von der Qualität der ursprünglichen Klassifikationen und Kriterien</p>	<p>Interpretation der Skalenzugehörigkeit der Items schwierig</p> <p>Ergebnisse abhängig von der Qualität der ursprünglichen Klassifikationen und Gruppenzuweisungen (braucht große und repräsentative Stichproben)</p>

Aufgabentypen und Antwortformate



Aufgaben mit freiem Antwortformat

- › Keine vorgegebenen Antwortoptionen, Antworten werden **von Testpersonen selbst formuliert/produziert**
- › **Antwortformat** in der **Instruktion vorgegeben** (Aussage treffen, Text schreiben, Zeichnung erstellen, ...)
- › Antworten müssen bei der Auswertung nach einem vorher erstellten **Kategoriesystem kodiert** werden → **Auswertungsobjektivität** abhängig von der **Qualität** dieses Kategoriesystems
- › **Einsatzbereiche**: Schulisch-pädagogischer Bereich, Kreativitätstests, projektive Verfahren

Kurzaufsatzaufgaben

- › Auf Fragen soll in Form von **kurzen Aufsätzen oder Essays** geantwortet werden (kann aber auch aus **einzelnen Wörtern** bestehen)
- › **Vorteil:** Antwort von Testperson **selbst erzeugt**, richtige Antwort zu erraten daher nicht möglich
- › **Nachteile:**
 - Bearbeitung und Auswertung wesentlich **aufwändiger**
 - **Auswertungsobjektivität eingeschränkt**
 - stark abhängig von **verbalem Verständnis** und **Formulierungskompetenzen** der Testperson
 - **Mehrdeutigkeit** von Antworten → gut entwickeltes Kategorisierungssystem wichtig!

Ergänzungsaufgaben

- › **Aufgabenstamm** soll durch ein bestimmtes Wort („Schlüsselwort“) oder eine Darstellung (Symbol, Zeichnung) **erweitert** werden
- › **Offene Fragen:** Schlüsselwort am Ende eines Satzes („Atheist verhält sich zu Religion wie Pazifist zu _____.“)
- › **Lückentexte:** Schlüsselwort fehlt im laufenden Text
- › **Vorteile:**
 - Reproduktion von Wissen, nicht nur Wiedererkennung → geringere Ratewahrscheinlichkeit
 - Objektivität idR gegeben
- › **Nachteile:**
 - Höherer zeitlicher Aufwand als Aufgaben mit gebundenen Antwortformaten
 - Meist nur Prüfung von Faktenwissen
 - Eingeschränkte Auswertungsobjektivität, wenn nicht alle möglichen Antworten bei der Testentwicklung berücksichtigt wurden

Aufgaben mit **gebundenen Antwortformaten**

- › **Mehrere Alternativen** für mögliche Antworten vorgefertigt
- › Probanden können aus Antwortoptionen **passende Antworten auswählen**, sind in ihrer Antwort aber an diese Optionen „**gebunden**“
- › Bei der Itementwicklung muss darauf geachtet werden, dass Antwortoptionen das **Merkmal vollends ausschöpfen**
- › **Vorteile:**
 - **Wenig Zeitaufwand** für Testbearbeitung
 - **Ökonomische** Auswertung
 - Höhere **Auswertungsobjektivität** als von Aufgaben mit freiem Antwortformat
- › **Nachteile:**
 - Antwortformate können von Testpersonen als zu „**eng**“/ **zu einschränkend** empfunden werden
 - Möglich, dies mit „Forced-Choice-Aufgaben“ zu umgehen (Wählen, was am ehesten zutrifft)

Ordnungsaufgaben

- › Einzelne Bestandteile der Aufgabe sollen zu einer **inhaltlich passenden Ordnung umgeordnet** oder einander **zugeordnet** werden

Zuordnungsaufgaben:

- › richtige Zuordnung von jeweils zwei Elementen zueinander (z.B. Hauptstädte zu Ländern)
- › Mit jeder richtigen Zuordnung sinkt die Zahl möglicher Antwortalternativen → daher ratsam, **zusätzliche Distraktoren** einzubauen

Umordnungsaufgaben:

- › Umsortieren von Worten, Satzteilen, Zahlen, Bildern oder Gegenständen zu einer **logisch-sinnvollen und eindeutigen Reihenfolge**
- › **Schwierigkeit** u.a. abhängig von der **Anzahl** der umzuordnenden Teile
- › Je mehr Teile, desto höherer **Konstruktionsaufwand**, um **Eindeutigkeit der Reihenfolge** zu gewährleisten

Auswahlaufgaben

- › Testpersonen sollen aus **mehreren vorgegebenen Antwortalternativen** die **richtige/-n bzw. zutreffende/-n Antwort/-en** angeben
- › Zu beachtende Aspekte bei Auswahlaufgaben:
 - Leistungstests:
 - Geeignete **Distraktoren** verwenden, um Rekognition zu erschweren
 - Antwortalternativen müssen **disjunkt** sein, sodass **Richtigantwort/-en eindeutig** sind
 - Persönlichkeitstests:
 - Antwortoptionen müssen **exhaustiv** sein (alle möglichen Antworten ausschöpfen)

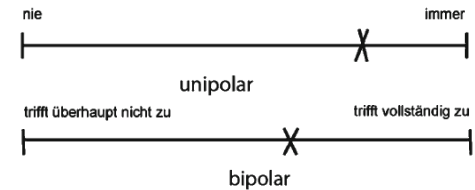
Auswahlaufgaben – Anzahl der Antwortalternativen

- › **Dichotom**: Zwei Antwortalternativen (richtig/falsch, ja/nein)
 - einfach und ökonomisch
 - dafür aber hohe Ratewahrscheinlichkeit, Rekognition und höhere Zustimmungstendenzen
- › **Mehrfachauswahl**: mehrere Antwortalternativen
 - klassisch mit nur einer richtigen Antwort oder „Multiple Mark Questions“ (beliebige Anzahl von richtigen Antworten)
 - Ebenfalls einfach und ökonomisch, geringere Ratewahrscheinlichkeit als bei dichotomen Items
 - dafür schwierigere Kodierung von Richtigantworten (bei Multiple Mark Questions) und höhere Gefahr von Verzerrungen bei schlechten Distraktoren

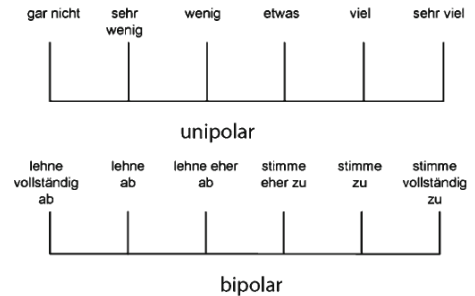
Beurteilungsaufgaben

- › Häufig eingesetzt bei **Persönlichkeitsfragebögen** → Grad der **Zustimmung oder Ablehnung** als Indikator für die Ausprägung des interessierenden Persönlichkeitsmerkmals
- › Zu beachtende Aspekte bei Beurteilungsaufgaben:
 - Kontinuierliche (Analogskalen) vs. diskrete gestufte Beurteilungsskalen (Ratingskalen)
 - Anzahl der Skalenstufen bei diskret gestuften Skalen (idR 5-7 optimal hinsichtlich Validität und Reliabilität)
 - Polarität der Antwortskala (unipolar vs. bipolar)
 - Bezeichnung der Skalenpunkte (numerisch, verbal, optisch, Symbole, Kombinationen)
 - Verwendung einer neutralen Mittelkategorie
 - „Weiß nicht“ als separate Antwortalternative

A. Analogskalen:



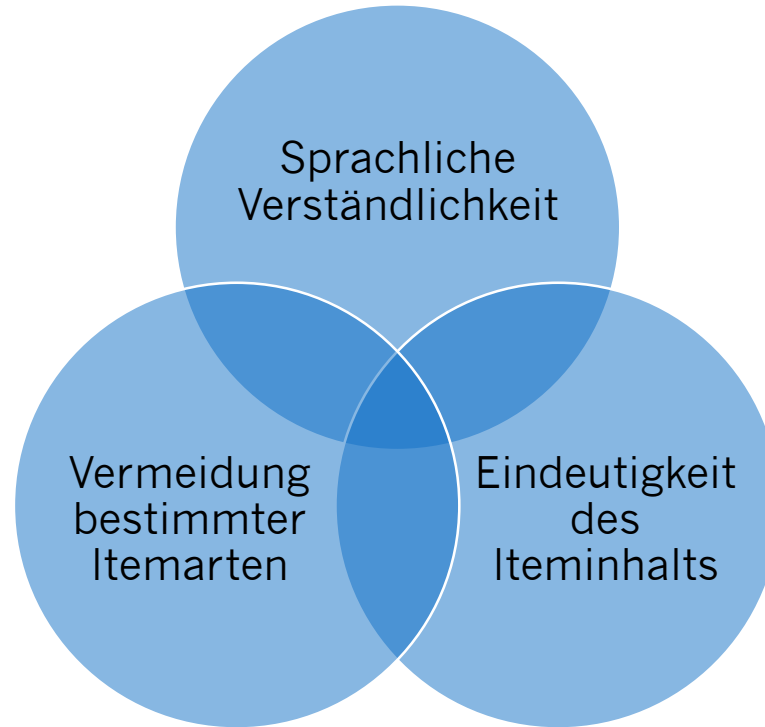
B. Diskret gestufte Ratingkalen:



Atypische Antwortformate

- › V.a. bei Leistungstests zur Überprüfung bestimmter Leistungen
- › Meist stärkere „**Eigenständigkeit**“ der Testpersonen bei der Generierung der Antworten
- › Oft **Kombinationen** mehrerer Antwortformate oder **sehr spezifische Reaktionen**, z.B.:
 - Abzeichnen einer Vorlage (z.B. geometrische Formen)
 - Nachahmen einer Handlung oder Bewegung
 - Gestisches Umsetzen von Aufforderungen (z.B. Blatt nehmen, falten, fallen lassen)
- › **Vorteile:**
 - Wenig Vorgaben, z.B. hinsichtlich des Formates → große **Konstruktionsfreiheit**
- › **Nachteile:**
 - Ergebnisse sehr **spezifisch** → schlechtere Übertragbarkeit auf andere Tests/ Bereiche
 - Hoher **Konstruktionsaufwand**

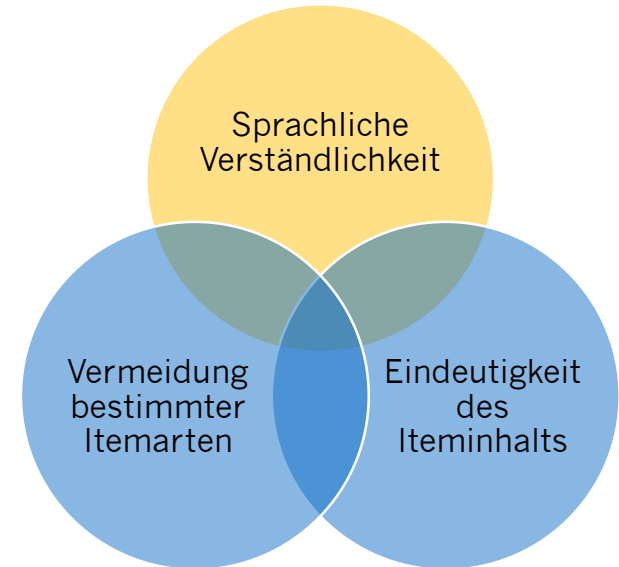
Gesichtspunkte der Itemformulierung



Gesichtspunkte der Itemformulierung

Sprachliche Verständlichkeit: Items sollten schon beim ersten Lesen verständlich sein

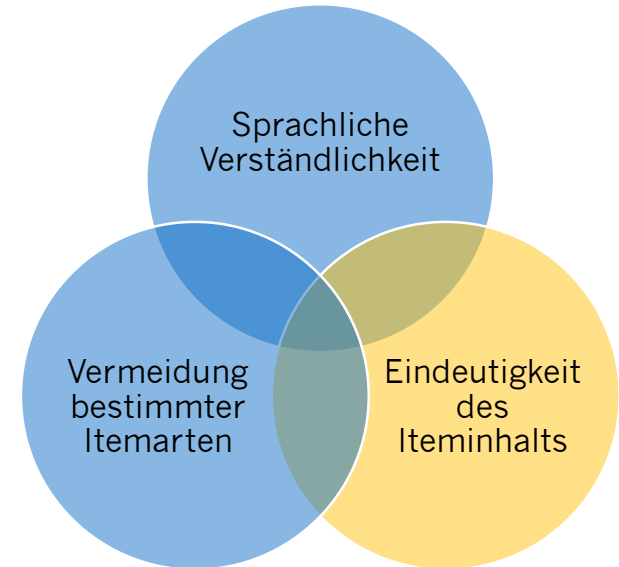
- › Items sollten **positiv formuliert** sein, Negationen sollten vermieden werden
- › Klare **Satzkonstruktionen** (so einfach, präzise und kurz wie möglich!)
- › **Fachbegriffe und Abkürzungen** vermeiden



Gesichtspunkte der Itemformulierung

Eindeutigkeit des Iteminhalts: Alle Testpersonen müssen den Iteminhalt gleich auffassen

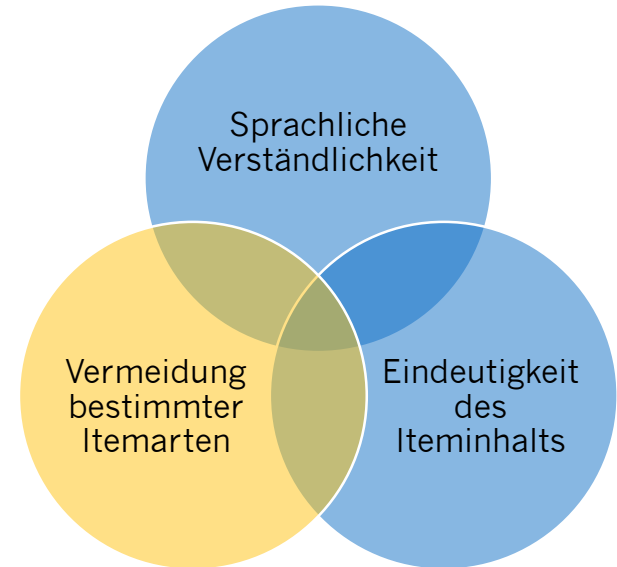
- › **Keine Universal ausdrücke** (z.B. „Immer“, „nie“, „alle“ → ein Gegenbeispiel reicht zur Falsifizierung)
- › **Häufigkeits- oder Intensitätsangaben** geeignet spezifizieren
- › **Definitionen** für unbekannte/ schwierige Frageninhalte geben (z.B. „Soziale Intelligenz“)
- › **Mehrdeutigkeit** vermeiden
- › Konzentration auf **einzelne Aussagen**, keine Verknüpfung mehrerer Aussagen
- › Vorwissen und Sprachniveau an **Zielgruppe** anpassen



Gesichtspunkte der Itemformulierung

Vermeidung bestimmter Itemarten: Bestimmte Itemarten können zu Beantwortungsproblemen führen

- › **Kurzlebige** Inhalte (Inhalte, die auf aktuellem Wissen basieren, z.B. politischem Wissen)
- › Implizite **Wertungen** („Warum ist X besser als Y?“)
- › **Suggestionen** („Würden Sie nicht sagen, dass...“)
- › **Merkmalskonfundierungen** (Antwortverhalten soll nur vom **interessierenden Merkmal** selbst abhängen; z.B. Befangenheit bei heiklen Fragen)
- › **Heikle Fragen** (z.B. zu sexuellen Inhalten) in der Instruktion **ankündigen**, **Anonymität** zusichern



Übung – Was ist an diesen Items problematisch?

Am ehesten
trifft auf mich zu...

1.

Im Umgang mit anderen bin ich verständnisvoll	
Ich neige dazu, Schwierigkeiten zu erwarten	
Ich bin entscheidungsfreudig	
Ich reagiere schnell und lebhaft	

- › Antwortoptionen sind nicht disjunkt (schließen sich nicht gegenseitig aus)

Übung – Was ist an diesen Items problematisch?

„Was glauben Sie: In welchem Alter beginnt bei Männern normalerweise die Andropause?“

- › Nicht definierter Fachbegriff – kann man davon ausgehen, dass Personen den Begriff „Andropause“ kennen?

Übung – Was ist an diesen Items problematisch?

„Es ist nicht gut, wenn Kinder ihren Eltern widersprechen.“

Stimme überhaupt nicht zu Stimme vollkommen zu

- › Sprachlich negative Formulierungen können zu Verwirrungen führen
Besser: „Es ist gut, wenn...“

Entscheidungshilfen für die Aufgabenauswahl

- › Leichte Verständlichkeit
- › Einfache Durchführbarkeit
- › Kurze Bearbeitungszeit
- › Geringer Material- bzw. Papierverbrauch
- › Leichte Auswertbarkeit
- › Geringe Häufigkeit von Zufallslösungen

Danke für eure Aufmerksamkeit!

Nächste Woche geht es dann weiter mit der Testauswahl und Testanwendung!

Quellen

- › Diagnostik- und Testkuratorium (2018). TBS-DTK. Testbeurteilungssystem des Diagnostik und Testkuratoriums der Föderation Deutscher Psychologinnenvereinigungen. Revidierte Fassung vom 03. Jan. 2018. Psychologische Rundschau 18, 109–116.
- › Eid, M. & Schmidt, K. (2014). Testtheorie und Testkonstruktion (Bachelorstudium Psychologie). Göttingen: Hogrefe.
- › Krumm S., Schmidt-Atzert L., Amelang M. (2021) Grundlagen diagnostischer Verfahren. In: Schmidt-Atzert L., Krumm S., Amelang M. (eds) Psychologische Diagnostik. Springer, Berlin, Heidelberg.
- › Kapitel 4 und 5 in: Moosbrugger, H. & Kelava, A. (Hrsg.). (2020). Testtheorie und Fragebogenkonstruktion (Lehrbuch, 3., vollständig neu bearbeitete, erweiterte und aktualisierte Auflage). Berlin: Springer.