



LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

# Vorlesung Testtheorien

Dr. Tobias Constantin Haupt, MBA

Sommersemester 2007



**Vorlesung Testtheorien:  
Inhalte im Überblick**

10.2.2003 - v16

Auf Wunsch zum  
Veranstaltungsende: Probeklausur  
😊 inkl. Besprechung

**Kap. 13:  
Arten psychologischer  
Tests**

- Überblick
- wichtige Testbereiche
  - Leistungstests
  - Persönlichkeitstests
- Beispiele für psych. Tests

**Kap. 12:  
Entscheidungen in der psych.  
Diagnostik**

- Überblick
- Selbstdarstellung/Impression Management
- Soziale Erwünschtheit
- Antworttendenzen
- Urteilsfehler bei Rating - Skalen

**Kap. 11:  
Testverfälschungen**

**Handout - Kap. 4:  
Tests als  
Datenerhebungsverfahren**

- Was ist eigentlich ein psychologischer Test?
- Grundvoraussetzungen für die Erfassung und Interpretation von interindivid. Unterschieden
- Arten von Tests

**Kap. 5:  
Testbestandteile, Testitems und  
Testgestaltung**

- Sprachliche Gestaltung von Items und Antwortmodi
- Itemanalyse
  - Itemschwierigkeit
  - Trennschärfe
  - Skalenhomogenität
  - Itemselktion

**Kap. 6 und 7:  
Die beiden großen  
Testtheorien**

- 1 Klassische Testtheorie (KTT)
  - Axiome der KTT
  - Ableitungen aus den Axiomen
  - Kritik an der KTT
- 2 Probabilistische Testtheorie (IRT)
  - Grundlagen, Grundkonzepte
  - Modelle der IRT
  - Kritik der IRT

**Kap. 8:  
Kriteriumsorientierte Tests**

- ? Frage: Konkretes Ziel erreicht oder nicht?
- Grundlagen
- Gütekriterien - Besonderheiten bei diesen Tests

**Kap. 10:  
Kriterien zur Bewertung  
von Tests: Gütekriterien**

- Überblick
- Durchführung
- Auswertung 3 Objektivität der...
  - Interpretation
- Retest reliabilität
- Paralleltest reliabilität
- Split - half - Reliabilität 2 Reliabilität
- Interne Konsistenz
- Inhaltsvalidität
- Kriteriumsvalidität 1 Validität - gesonderte Mind - Map beachten!
- Konstruktvalidität
- Beziehungen zwischen Objektivität, Reliabilität und Validität

**Kap. 9:  
Testkonstruktionsansätze**

- Grundlagen
- Rationale Konstruktion
- Externale Konstruktion
- Induktive Konstruktion
- Prototypische Konstruktion
- Vergleich der Konstruktionsstrategien

## nach Art des zu erfassenden Merkmals (Anwendungsbereich):

- ***Persönlichkeitstests:*** Tests in denen Persönlichkeitsmerkmale bzw. Charaktermerkmale wie Eigenschaften, Motive, Interessen, Einstellungen etc. erfasst werden. Dabei spielen objektive Beurteilungsmaßstäbe keine Rolle.

## nach Zeitbemessung (bei Leistungstests):

- ***Speed -Tests (Schnelligkeits- oder Geschwindigkeitstests)***: Enthalten relativ leichte Aufgaben in einer Anzahl, die von praktisch keinem Pb innerhalb der knapp bemessenen, vorgegebenen Testzeit alle beantwortet werden können. Die Differenzierung zwischen Versuchspersonen wird nur durch die Begrenzung der Bearbeitungszeit erreicht. Je schneller die Aufgaben (richtig) gelöst werden, desto höher der Testwert.



### ***Power-Tests (Niveautests):***

Enthalten Aufgaben, die in ihrem Schwierigkeitsgrad kontinuierlich so weit ansteigen, daß die letzten Aufgaben trotz fehlender oder großzügig bemessener Zeitbegrenzung nur von sehr wenigen Probanden (Pb) beantwortet werden können.

Mit derartigen Verfahren wird primär das intellektuelle Niveau oder die „Denkkraft“ (Power) ermittelt.

Die meisten Tests sind Mischformen (da die Aufgabenschwierigkeiten oft stark streuen und Zeitbegrenzung vorliegt).



## nach Anzahl der zu erfassenden Persönlichkeits- oder Leistungsmerkmale:

***eindimensionale Tests:*** Wenn ein Test nur ein Merkmal, bzw. ein Konstrukt erfasst.

***mehrdimensionale Tests:*** Wenn ein Test mehr als ein Merkmal erfaßt. Die KTT (klass. Testtheorie) befaßt sich nur mit eindimensionalen Tests, so daß die einzelnen Dimensionen mehrdimensionaler Tests separat (in Subtests) zu analysieren sind.

***Dimensionalitätsüberprüfung:*** Ob ein Test wirklich eine best. (die intendierte) Anzahl von Konstrukten erfasst, läßt sich mittels einer Faktorenanalyse überprüfen



**nach Ausmaß des erforderlichen subjektiven Urteils des Auswerters:**

- **objektive Tests:** Wenn das Testergebnis rein mechanisch gewonnen werden kann, d.h. wenn es praktisch keinen subjektiven Interpretationsspielraum für den Testauswerter gibt.
- **nicht-objektive Tests:** Wenn ein subjektives Urteil des Auswerters eine Rolle spielt, wie dies etwa bei projektiven Tests der Fall ist.



## nach Art der Durchführung:

**Individualtests:** Tests, die lediglich einzeln mit den Versuchspersonen durchgeführt werden können, z.B. HAWIE-R, alle projektiven Tests.

**Gruppentests:** Tests, die auch in Gruppen durchführbar sind, was besonders in institutioneller Diagnostik erheblich zeitökonomischer ist.

## Nach Art der Aufgabenbeantwortung:

- ***gebundene Aufgabenbeantwortung*** Wenn bei einer Aufgabe bereits Antwortalternativen vorgegeben sind.
- ***freie Aufgabenbeantwortung*** Wenn der Pb selbst eine Antwort auf eine Frage oder Aufgabenstellung finden muß.

## **Materialbestandteile:**

### **Testmanual:**

- Beschreibung, Entstehung und wissenschaftliche Grundlegung des Testes
- Durchführungs- und Interpretationsanweisungen samt Normen

### **Testmaterial:**

- Stimulus- und Reaktionsmaterial

### **Auswertungshilfen:**

- Kontrollblätter, Lochfolien, Schablonen, Messgeräte

### **Durchführungsbestandteile:**

- Testanweisung für Testleiter (Bedingungen, Durchführung und Auswertung des Tests)

## Instruktion für Probanden

→ ist die Anweisung an den Probanden zur Testsituation

- **Ziel:** Verständliche Einführung in das Material; auf die Testsituation einstimmen; Unterstützung der Motivation des Probanden
- **Inhalt:** Technische Anweisungen zur Aufgabenbeantwortung, z. B auch Hinweis, keine Daten auszulassen (vgl. missing - value Problematik)

**Hauptfunktion: „Werbung“;**  
erwecken von Teilnahmemotivation, *Aufbau eines Vertrauensverhältnisses*. z.B.:

- „Die Daten werden vertraulich behandelt.“
- „Es gibt keine richtigen oder falschen Antworten.“
- Sie werden auf den folgenden Seiten eine Reihe von Aussagen über best. Verhaltensweisen, Einstellungen o. ä. finden. Sie können entweder mit ... oder ... antworten. Überlegen Sie bitte nicht erst, welche Antwort den besten Eindruck machen könnte. Setzen Sie bitte immer ein Kreuz, ...
- Belohnung: Was haben die Probanden von der Teilnahme? Geld, Feedback (z.B. Persönlichkeitsprofil), Duplo?



Jedes Item sollte so beschaffen sein, daß Personen mit unterschiedlichen Standpunkten gegenüber dem jeweiligen Einstellungsobjekt unterschiedlich auf das Item reagieren.

Probleme können sein:

- zu allgemeine (vieldeutige) Formulierung (Proband muß Item vage interpretieren)
- Generalisierungen vermeiden
- Formulierung möglichst an konkretem Verhalten orientieren
- unklarer Antwortbezug
- Fremdwörter
- mehrere Gedanken in einem Satz (z. B. „Ich gehe regelmäßig in die Kirche, teile aber die offizielle Auffassung der katholischen Kirche zum Thema Sexualität nicht“)
- Negationen (Mehrfachverneinungen vermeiden); Extreme Formulierungen (provozieren einheitliche Antworten)



Die Items müssen also sinnvoll sein, gewisse ethische Grundprinzipien beachten und auch wohlformuliert sein.

Leichter gesagt als getan... !



- Vermeide mehrdeutige Begriffe (u. a. „oft“)
- Vermeide Begriffe, die nur einem Teil der Pbn geläufig sind
- Jede Aufgabe soll nur einen Aspekt oder Gedanken enthalten, d.h. keine durch „und“ verbundenen Aussagen
- Möglichst positive Aussagen, Fragen oder Formulierungen (→ doppelte Verneinung)
- Vermeide Verallgemeinerungen jeder Art, besonders solche, die nur zeitweise Gültigkeit haben
- Vermeide umständliche Länge oder telegrafische Kürze im Aufgabentext
- Prüfe die Eindeutigkeit der Frage an einer kleinen Teststichprobe („Fragentext ist eindeutig / nicht ganz / unklar“)

## Dichotom

- **Essen Sie gerne Spaghetti?**  
 ja  nein
- **Das Jahr 2000 ist ein Schaltjahr**  
 stimmt  stimmt nicht

## dichotome Antworten

Vorteil: Proband muß sich entscheiden; reliablere Ergebnisse, jedoch auf Kosten der Validität (wir werden noch sehen, warum).

Nachteil: durch fehlende mittlere Antwortmöglichkeiten muß die Antwort nicht der tatsächlichen Einstellung des Probanden entsprechen; Gefahr der Suggestivwirkung und Ja-Sage-Tendenz

## Geeignet wenn ...

- ...ein Test schnell konstruiert bzw. improvisiert werden soll
- ...die Testdurchführung wenig Zeit beanspruchen soll
- ...Auswertung durch Hilfspersonal geschieht
- ...Aufgabeninhalt leicht verständlich ist
- ...Informationen über ein Persönlichkeitsmerkmal auf möglichst breiter Basis durch möglichst viele Fragen und Aufgaben erzielt werden soll
- ...gedankliche Verarbeitung und Gedächtnisleistung keine Rolle spielt
- ...Zufallslösungen irrelevant sind oder berücksichtigt werden (Zufallskorrektur)

## Schätzskalen (Ratings)

0

lehne ab

1

2

3

4

stimme zu

Vorteil: Antworten können besser den tatsächlichen  
Einstellungen des Pb entsprechen

Nachteil: bei ungerader Zahl von Antwortmöglichkeiten ist die mittlere Kategorie oft nicht eindeutig interpretierbar; es bleiben auch noch weitere Interpretationsmöglichkeiten offen:

- Unentschiedenheit („Flucht“kategorie)
- mangelnde Kompetenz des PB zur Beantwortung des Items
- Gleichgültigkeit/Faulheit/Motivationsprobleme des Probanden!  
werden außerdem i.d.R. auch als Intervallskala behandelt (ist aber umstritten)

besser: Schätzskala mit **gerader** Anzahl von Ratingmöglichkeiten

## graphische Angaben



### Multiple-Choice-Fragen

Kreuzen Sie eine oder mehrere Alternativen an:

Deutschland war Fußballweltmeister ...

- a) 1968                      b) 1980                      c) 1974                      d) 1984

Ich komme gerne in dieses Seminar:

Immer            oft            manchmal            selten            nie

Vorteile und Nachteile:

- ökonomische Durchführung und Auswertung
- geringe Ratewahrscheinlichkeit, wenn Distraktoren gut sind
- Erfahrung zur Formulierung von Items nötig

## Offene Antworten

Wie beurteilen Sie dieses Seminar?

---

wichtige Form zur Erfassung bestimmter Eigenschaften wie z.B. Ausdrucksfähigkeit, Wortschatz etc.

- Auswertungsobjektivität problematisch zu erreichen
- Auswertung zeitaufwendig

Blitz : Hören = Donner : ?

Sortiere die folgenden Brüche der Größe nach, beginne mit dem Größten

$4/9$  .....  $3/4$  .....  $2/3$  .....  $7/12$  .....  $5/6$

## Lückentext

Die menschliche Persönlichkeit und deren  
\_\_\_\_\_ ist eine sehr \_\_\_\_\_  
Angelegenheit.

- frei von Zufallseinflüssen
- Antwortformulierung nach freiem Ermessen ->
- Gefahr von Suggestivwirkung
- Antwortformulierung nach freiem Ermessen
- zeitaufwendig
- nicht immer eindeutig (Objektivität!)



## Zuordnungsaufgaben

Welches Verb gehört zu welchem Substantiv?

- |                    |             |
|--------------------|-------------|
| 1) ein Vortrag     | a) erzählen |
| 2) eine Geschichte | b) machen   |
| 3) eine Erklärung  | c) halten   |
| 4) ein Gespräch    | d) abgeben  |
| 5) einen Vorschlag | e) führen   |

- eignen sich gut zur Feststellung erworbenen Wissens
- große Anzahl von Aufgaben auf kleiner Fläche möglich
- Lösungsfindung erfolgt nicht selbstständig, da Antworten vorgegeben

## An welchen Flüssen liegen die folgenden Städte?

- Nürnberg: \_\_\_\_\_
- Frankfurt: \_\_\_\_\_
- Hamburg: \_\_\_\_\_
- Berlin: \_\_\_\_\_
- Bremen: \_\_\_\_\_

## Eindeutigkeit hinsichtlich des Konstruktes

### Konstrukt: „Abhängigkeit der Stimmung vom Wetter“

- Sonne hellt meine Stimmung auf.
- Bei miesem Wetter werde ich schneller ungeduldig als sonst.
- Trübes Wetter hat keinen Einfluß auf meine Stimmung.
- Bei gutem Wetter fällt es mir leichter, freundlich zu anderen zu sein.
- An sonnigen Tagen lache ich häufiger.

Eindeutigkeit hinsichtlich des Konstruktes

### **Konstrukt: „Einstellung zu Psychotherapie“**

- Psychotherapie sollte denselben Stellenwert haben wie medizinische Behandlung.
- Therapeuten haben selbst psychische Probleme.
- Wer sich in Therapie begibt, ist unfähig, seine Probleme selbst zu lösen.
- Ich kann mir nicht vorstellen, jemals eine Psychotherapie zu brauchen.

## Abstufung über die Antwortmöglichkeiten

Meine Arbeit ist eintönig und langweilt mich.

Meine Arbeit ist meistens eintönig und langweilt mich fast immer.

- ① trifft überhaupt nicht zu  
② trifft nicht zu  
③ trifft eher nicht zu  
④ trifft eher zu  
⑤ trifft zu  
⑥ trifft sehr zu

Wenn ich zum Zahnarzt muß, habe ich jedesmal extreme Angst.

Ich habe Angst, wenn ich zum Zahnarzt muß.

- ① trifft überhaupt nicht zu  
② trifft nicht zu  
③ trifft eher nicht zu  
④ trifft eher zu  
⑤ trifft zu  
⑥ trifft sehr zu

**Objektivität:** Sie muss von verschiedenen Auswertern gleichermaßen als „richtig“ oder „falsch“ bewertet werden

**Reliabilität:** Gleiche Antworten bei kurzzeitiger Testwiederholung muss erwartet werden

**Validität:** Eine Aufgabe ist dann valide, wenn sie von Pbn mit starker Merkmalsaufprägung häufiger richtig beantwortet wird als von Pbn mit geringer Merkmalsausprägung (vgl. Trennschärfe)