



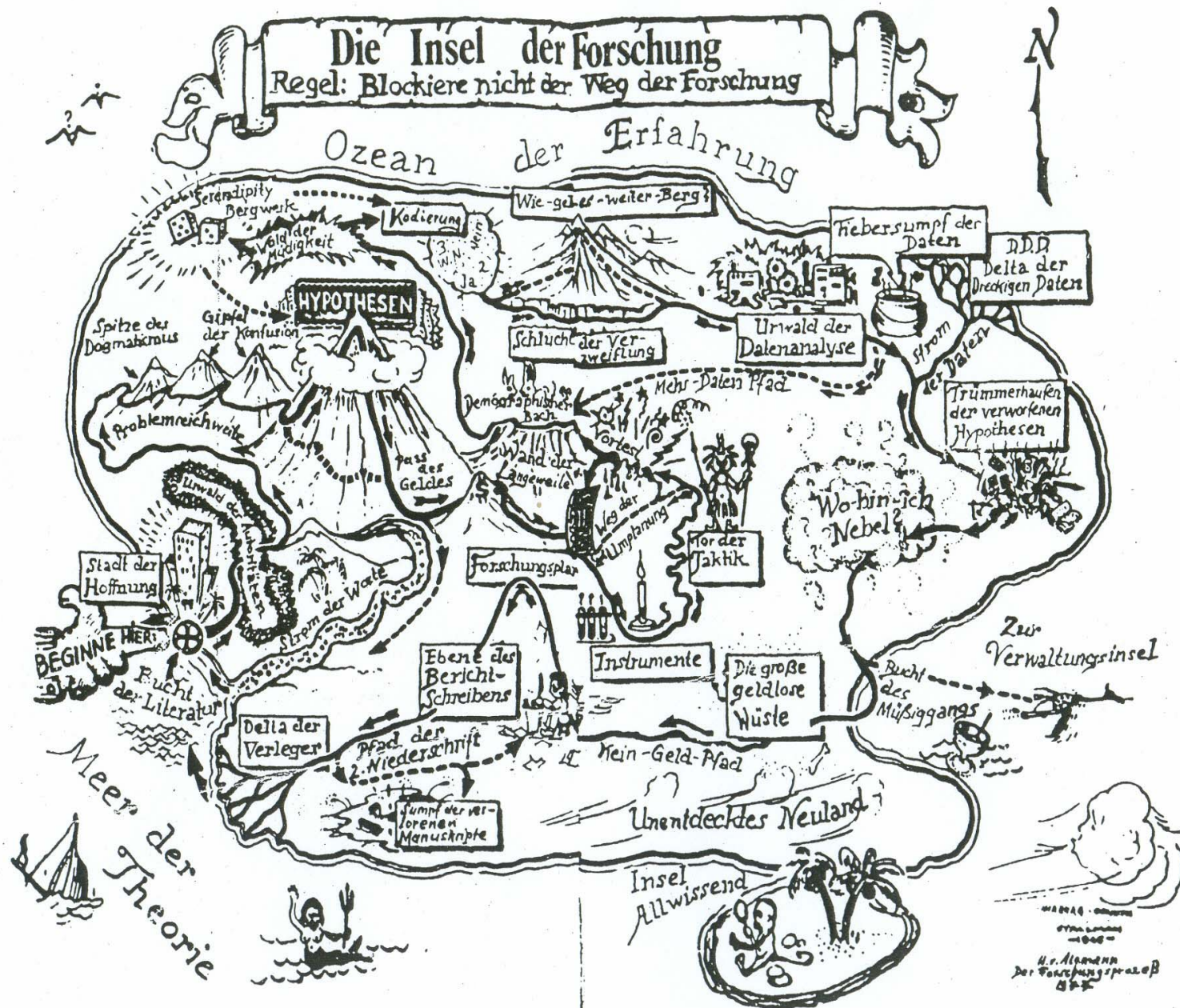
LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

# 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 interindiv. 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



## Die 7 P's des FBI:

- Proper prior preparation prevents piss poor performance

Tun Sie, was Sie befürchten, und die Angst stirbt einen sicheren Tod.

Wenn Sie zufrieden sind mit meiner Leistung, sagen Sie es bitte **ANDEREN**, wenn sie nicht zufrieden sind, sagen sie es bitte **MIR**. Danke!





Ein Test ist ein wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer empirisch abgrenzbarer Persönlichkeitsmerkmale mit dem Ziel einer möglichst quantitativen Aussage über den relativen Grad der individuellen Merkmalsausprägung (G. A. Lienert)

Ein Test ist ein „(nicht-apparatives  
Routine-) Verfahren zur Messung menschlicher  
Verhaltensweisen mit wissenschaftlichem Anspruch“  
(Brickenkamp, 1975)



„Ein Test ist ein **systematisches Verfahren**, mit dem einem Probanden konstruierte **Stimuli** vorgegeben werden, auf die er/sie **reagiert**.

Aufgrund dieser Reaktion kann der Testanwender dem Probanden **Zahlen zuordnen**, von denen er auf das **Vorhandensein** und die relative **Merkmalsausprägung** *schließt*.“

(Kerlinger, 1964, S. 480)



## Bitte Vorsicht vor Verwirrung durch diverse andere Bedeutungen des Begriffes „Test“:

- Vorgang der Durchführung der Untersuchung selbst (testen = Testanwendung und Testauswertung)
- Mathematisch - statistisches Prüfverfahren: z.B. „t-Test“
- Die Gesamtheit der zur Durchführung eines Testverfahrens notwendigen Requisiten
- Hypothesentest





### ***Zur Abgrenzung von Tests gegenüber Experimenten:***

Hauptunterschied ist, daß bei Experimenten mindestens eine unabhängige Variable manipuliert wird, was bei Tests i. d. R. nicht geschieht. Testitems sind insofern keine unabhängige Variablen.

Allerdings sieht Rost (1996) Tests als ein spezielles psych. Experiment mit dem Ziel, vergleichende Aussagen über Personen abzuleiten.

Ein *Test* ist ein Verfahren, das spezifische Informationen über einzelne Testobjekte (z. B. Personen) liefert.

*Testobjekte* sind

meistens Einzelpersonen,

können aber auch Organisationen, Gruppen oder ähnliches sein,

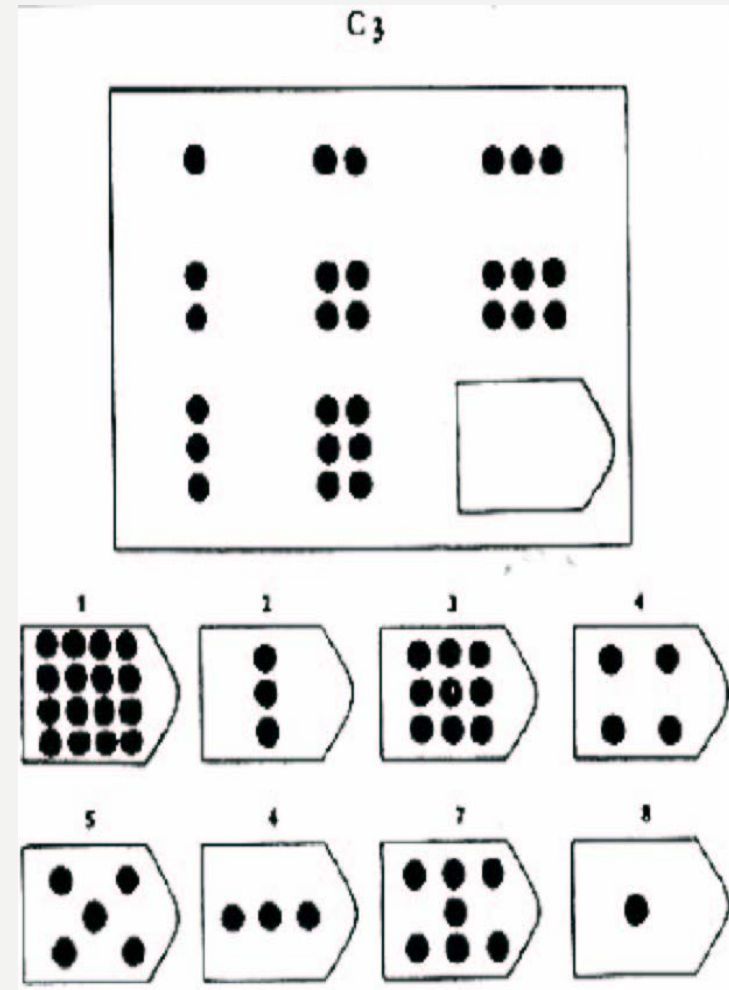
gegebenenfalls sogar Kombinationen aus Personen und

Situationen (Situationsspezifisches Testen) oder Zeitbereichen (Veränderungsmessung)

**Testitems:** Die einzelnen, zu bearbeitenden Aufgaben eines Tests.

Abb. rechts: Item aus Ravens Progressive Matrices (1956)

Wissen Sie die Lösung?





## Beispielitem aus dem TAT (Revers & Taeber, 1968)





Beispielitem aus dem  
Rorschach - Test  
(Rorschach, 1954)

Was stellt der Klecks für  
Sie persönlich dar?



1	Es würde ihm/ihr mehr Spaß machen, ein Berater bzw. eine Beraterin zu sein als ein Architekt bzw. eine Architektin.	[a]	stimmt	1
		[b]	?	2
		[c]	stimmt nicht	3
2	Wenn andere etwas tun, das ihn/sie stört, macht er/sie meistens folgendes:	[a]	er/sie läßt es gut sein	1
		[b]	?	2
		[c]	er/sie bringt es ihnen gegenüber zur Sprache	3

Der 16 PF – R (Schneewind & Graf, 1998) in der modifizierten Version von Haupt et al. (1999) zur Einschätzung anderer, bekannter Personen.

Items bestehen somit – so unterschiedlich sie auch  
aussehen mögen - immer aus zwei Komponenten:

- einem Stimulus- und
- einem Reaktionsteil

**Itemstamm/Stimulusteil:**

Frage, Aussage, Bild, Geschichte,  
Zeichnung, Rechenaufgabe,  
Testsituation

**Antwortformat/Reaktionsteil:**

Ankreuzalternativen, leere  
Zeile, Antwortskala, weißes  
Blatt

Eine Testskala ist ein Satz von Items (also mehrere Items, die meist Bestandteil eines Tests sind), die spezifischen, mit der jeweiligen Testskala verbundenen Skalierungseigenschaften genügen.





LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

Was ist eine Testbatterie?



Kombination von Einzeltests,  
mit dem Ziel, die Validität des mit einem einzelnen  
Tests gemessenen Merkmals zu erhöhen.

*Welche empirischen Sachverhalte können durch welche numerischen Strukturen angemessen abgebildet werden?*

Abbildung eines empirischen Relativs in ein numerisches Relativ

Die Beziehung zwischen emp. Relativ und numer. Relativ soll homomorph („gleich- förmig“) sein



Eigentlich: Theorie über verschiedene Arten von Tests, ihren Aufbau und ihre Konstruktionsprinzipien, weil:

- Testauswertung basiert auf dem Antwortverhalten  
⇒ Schluss auf ein Merkmal

Die Testtheorien beschäftigen sich mit dem Zusammenhang von Testverhalten und dem zu erfassenden (Person-)Merkmal



Anders ausgedrückt:

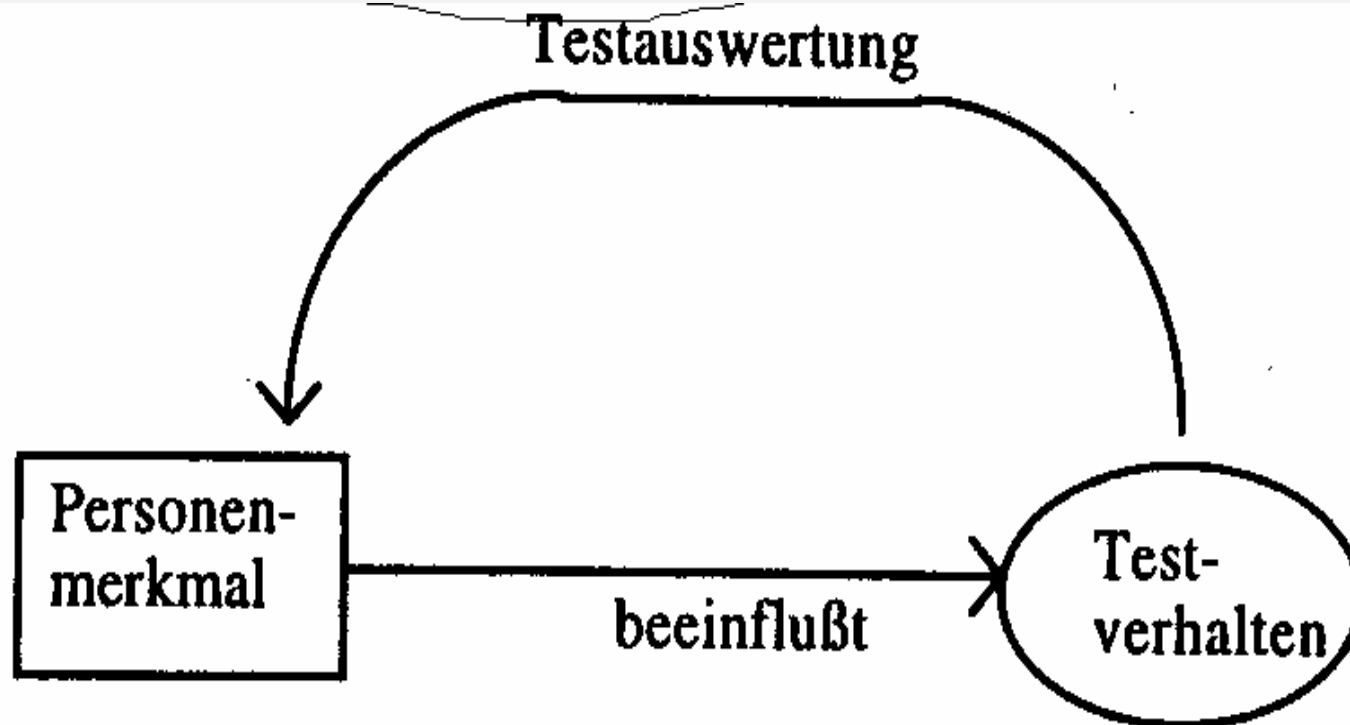
Testtheorien befassen sich mit der Frage, wie die empirischen Testwerte und die zu messenden (tatsächlichen) Merkmalsausprägungen zusammenhängen.



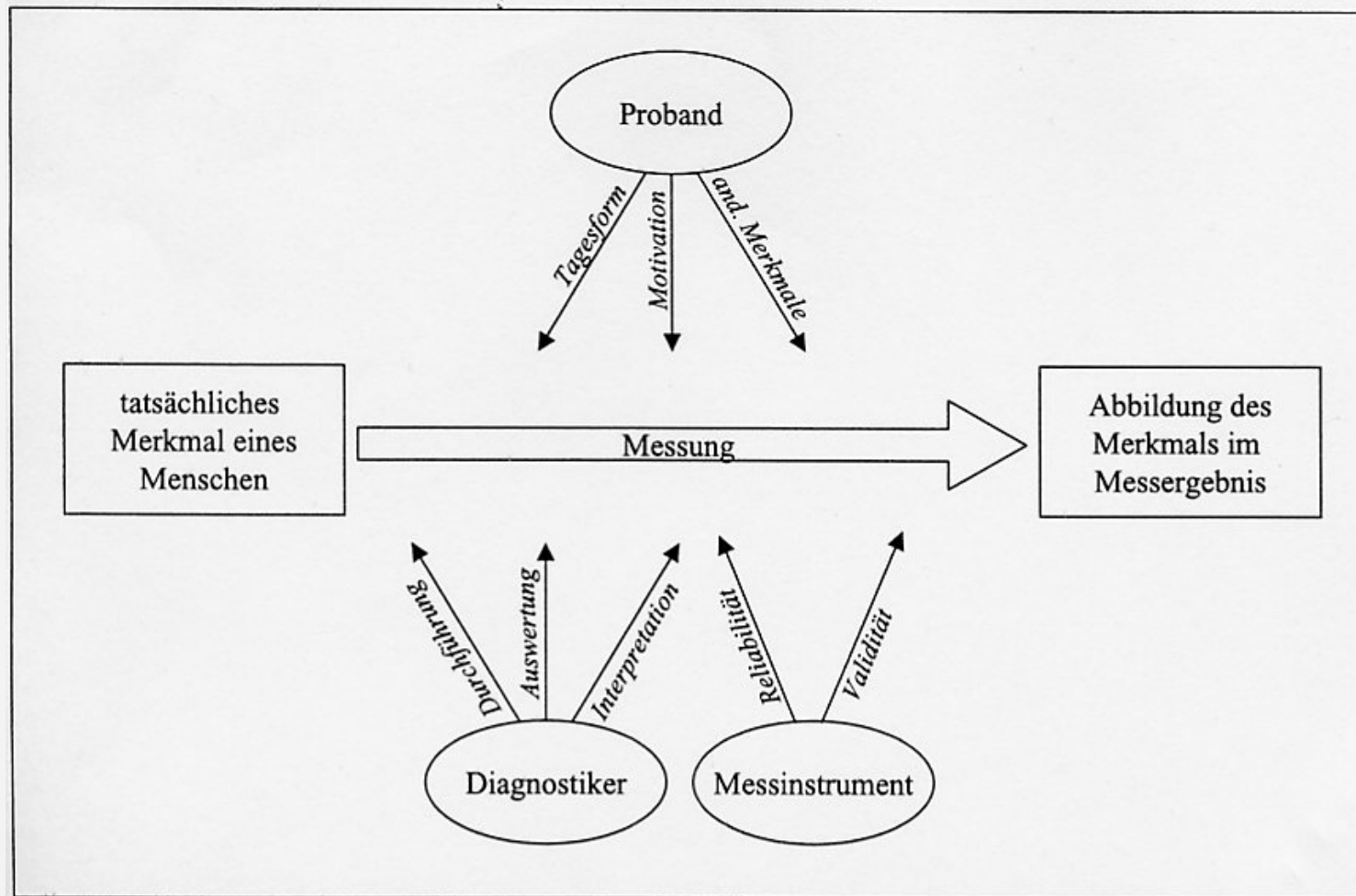
Oder auch:

Die Frage nach den Anforderungen, denen ein Test genügen muß, um aufgrund eines Testergebnisses auf die tatsächliche Ausprägung des getesteten Merkmales schließen zu können, ist Gegenstand der Testtheorien.

Aus den Vorgaben der Testtheorien können die Gütekriterien und deren Berechnung abgeleitet werden (Bortz, 1995).



**Abbildung 1:** Der Gegenstandsbereich der Testtheorie



## ***Standardisierung***

Messwerte aus Tests können inter- und intraindividuell nur dann sinnvoll miteinander verglichen werden, wenn die jeweiligen Untersuchungsbedingungen (z.B. Material, Instruktionen, Situationsbedingungen) konstant gehalten werden, bzw. identisch sind.





## *Differenzierung*

Test und Durchführungsbedingungen sollten so gestaltet sein, daß sich die Merkmalsträger im interessierenden Verhalten maximal voneinander unterscheiden (und gleichzeitig geringe Unterschiede hinsichtlich irrelevanter Variablen bestehen).



## ***Objektivität***

Die zwischen den Merkmalsträgern auftretende Varianz muß unabhängig von der Person des Versuchsleiters registrierbar sein, etwa indem es feste Durchführungs- und Auswertungsvorschriften gibt.



## ***Reliabilität:***

Messpräzision

Merkmalsausprägungen müssen z.B. zeitlich stabil sein (Retestreliaibilität), d.h., bei einer Messungswiederholung bei gleichen Personen und unter gleichen Bedingungen zu ähnlichen Ergebnissen führen (zumindest dort, wo dies inhaltlich sinnvoll ist; ansonsten gibt es alternative Varianten der Reliabilität).



### ***Validität: Misst der Test, was er messen soll?***

Es muß belegt sein, daß der Test auch wirklich das misst, was er messen soll.

Validiert werden kann ein Test entweder mit einem Außenkriterium oder einem Indikator für ein solches Außenkriterium (z.B. ein anderer Test).

Lienert & Raatz (1998) führen folgende  
**Klassifikationsmöglichkeiten von Tests** an:

- **nach Allgemeinheitsgrad ihrer Anwendung:**
  - ***standardisierte (geeichte) Tests:*** Wenn Tests wissenschaftlich entwickelt worden sind, hinsichtlich der wichtigsten Gütekriterien untersucht worden sind, unter Standardbedingungen durchführbar sind und normiert sind
  - ***nichtstandardisierte (informelle) Tests:*** Wenn mindestens eine der o. g. Bedingungen nicht gegeben ist.



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

***Vorbemerkung:*** Hier handelt es sich um eine grobe Zweiteilung nach Bortz (1995).

- ***Leistungstests:*** liegen immer dann vor, wenn Aufgaben entweder richtig oder falsch zu beantworten sind, also ein verbindlicher Beurteilungsmaßstab vorliegt.

***Arten von Leistungstests:*** Intelligenztests, Entwicklungstests, Schultests, allgemeine Leistungstests, spezielle Funktions- und Eignungstests.

## 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.