



Quantitative Verfahren zur Erhebung von Schülerleistungen

Dr. Christof Nachtigall
Universität Jena



Inhalt

Einführung zum Thema ‚Messen‘

Gütekriterien bei der Testentwicklung

**Testentwicklung am Beispiel der
Thüringer Kompetenztests**

Auswertung von Testergebnissen



Einführung zum Thema ‚Messen‘: Qualitative vs. Quantitative Verfahren

- oft als Gegensätze gesehen
- Nutzen der Trennung fraglich
- Oberbegriff: Methoden der Datenerhebung / Messen

Beispiele:

- Erfassen des Geschlechts mittels Fragebogen
- Notenvergabe
- Vergleichsarbeiten (z.B. Thüringer Kompetenztests)
- psychometrische Tests (z.B. Intelligenztests)



Was heißt Messen?

Relationen in der erfahrbaren Welt (empirisches Relativ) derart in Zahlen (numerisches Relativ) abzubilden, dass die relevanten Relationen erhalten bleiben.





Verschiedene Messniveaus (Skalenniveaus):

- kategorial (Nominalskala, z.B. *Geschlecht*)
- ordinal (Ordinalskala, z.B. *Schulnoten*)
- metrisch (Intervallskala, z.B. *Körpergröße*)



Bedeutung des Messniveaus

Das Messniveau legt fest, welche Analysen mit meinen Daten sinnvoll oder sinnlos sind.

Beispiel: Mittelwertbildung erst bei metrischen Skalen sinnvoll

Konsequenz: Theoretische Vorüberlegungen vor der eigentlichen Messung wichtig



Wissenschaftliche Gütekriterien beim Messen

- **Objektivität**
(Messung ist unabhängig vom Messenden)
- **Reliabilität**
(Genauigkeit der Messung)
- **Validität**
(Es wird gemessen, was gemessen werden soll)



Messmodelle

- **Klassische Testtheorie**
Gemessener Wert = wahrer Wert + Messfehler
- **Item Response Theorie (IRT)**
Richtige Lösung einer Aufgabe (Item) ist probabilistische abhängig von der Fähigkeit der Person und der Schwierigkeit der Aufgabe
Lösungswahrscheinlichkeit = $f(\text{Fähigkeit, Schwierigkeit})$



Thüringer Kompetenztest als Messinstrument

Testentwicklung in 5 Schritten

Ziel: Optimierung der Gütekriterien



5 Schritte der Testentwicklung

1. Zieldefinition (*was wollen wir messen*)
2. Operationalisierung durch Aufgabenpool
(*wie sollen wir messen*)
3. Aufgabenoptimierung (*Eindeutigkeit der Fragestellung,
Korrekturhinweise, Praktikabilität*)
4. Vortest (Pilotierung)
5. Aufgabenselektion und Erstellung des finalen Tests



Beispiel zu 1: Textverstehen gemäß Pisa-Modell

Drei Teildimensionen:

- Information ermitteln
- Textbezogenes Interpretieren
- Reflektieren und bewerten

Zusätzlich: verschiedene Schwierigkeitsstufen



Beispiel zu 2: Standpunkt vertreten

Aufgabe 3

Hier folgt ein Ausschnitt aus einem Gespräch zwischen einem Katzenfreund und einer Hundeliebhaberin:

Klaus

Erika

Notiere in Stichworten, wie beide Personen ihren Standpunkt begründet vertreten könnten. Stütze dich dabei auf Informationen aus dem Text.



Erhöhung der Validität

- Theoretische Überlegungen
- Einschätzung von Experten
- Außenkriterien

Objektivität und Reliabilität sind Voraussetzung für Validität



Beispiel zur Erhöhung der Objektivität: Geschlossene Antwortformate

Aufgabe 1:

a) Welches Ziel verfolgt Immanuel Birmelin mit seiner Forschung? Kreuze die richtige Antwort an!

Er will ...

- A den Streit zwischen Hundeliebhabern und Katzenfreunden schlichten.
- B die Frage beantworten, ob Hunde zählen können.
- C zeigen, dass Katzen und Hunde beide schlau sind.
- D herausfinden, ob Hunde oder Katzen schlauer sind.
- E Experimente mit Hunden und Katzen machen.



**Beispiel zur Erhöhung der Objektivität:
Präzise Korrekturanweisungen**

b) Begründe deine Antwort mit Hilfe des Textes.

[Anspruch: eine These textbezogen vertreten]

Lösung:

Vgl. Z. 2/3: *Birmelin hat versucht, die Frage „Wer ist schlauer: Hunde oder Katzen?“ mit Experimenten zu beantworten.* **2 P**

Für diese Antwort erhält der Schüler 2 Punkte, andere Antworten sind falsch (0 Punkte).



Beispiel zu 2: verschiedene Wortfeldaufgaben

Aufgabe 5

Ein Wortfeld ist eine Gruppe von Wörtern mit ähnlicher Bedeutung, die der gleichen Wortart angehören.

a) **Unterstreiche im folgenden Kästchen alle Wörter, die dem Wort „fliegen“ inhaltlich ähnlich sind und zur gleichen Wortart gehören.**

segeln	Flugsaurier	schweben	Fliegerstaffel
düsen	Flugangst	ängstlich	flattern
fliehen	schwirren	hoch	Flügel
			aufsteigend

... verschiedene Wortfeldaufgaben

- b) Die Wörter, die du in Aufgabe a) herausgefunden hast, bilden ein Wortfeld. Welche der folgenden Wörter gehören nicht zum Wortfeld „sich fortbewegen“? Streiche diese Wörter deutlich durch.

laufen hasten schleichen verstärken eilig
trippeln Auto schreiten
Bewegung rennen schnell marschieren

...verschiedene Wortfeldaufgaben

- c) Bilde das Wortfeld „essen“ durch fünf weitere Wörter.

essen



Erhöhung der Reliabilität durch ‚Skalenbildung‘

Durch Messung mit mehreren ähnlichen Items kann die Messgenauigkeit deutlich erhöht werden

Beispiel: Summe der erreichten Punktzahl zu verschiedenen Aufgaben eines Aufgabentypus ermitteln

Bezeichnung: Summenscore



Zu 4.: Vortest (Pilotierung)

- Erproben an Stichprobe
- Zusätzlich Fragebogen /Interview mit allen Beteiligten
- Veränderung problematischer Items



Beispiel zu 5: Analyse der Aufgabenschwierigkeit

Schwierigkeit: Anteil der Schüler, welche die Aufgabe lösen

	N	Schwierigkeit
Ag_1N	180	,4294
Ag_2AN	180	,3708
Ag_2BN	180	,5889
Ag_2CN	179	,5112
Ag_2DN	180	,5028
Ag_2EN	180	,3778
Ag_3AN	180	,5972
Ag_3BN	167	,5763
Ag_4AN	180	,3462
Ag_4BN	180	,6278
Ag_51N	180	,7710
Ag_52N	180	,8083

Die Schwierigkeiten sollten den Fähigkeitsbereich der Schüler abdecken



Beispiel zu 5: Analyse der Reliabilität

- Überprüfung, ob Items, die ähnliches messen sollen, dies auch tun
- Ähnliche Items zu einer Skala zusammenfassen
- Reliabilität der Skalen bestimmen
(z.B. Cronbachs alpha)



Beispiel zu 5: Reliabilitätsanalyse

RELIABILITY ANALYSIS - SCALE (ALPHA)

Statistics for	Mean	Variance	Std Dev	Variables
SCALE	15,5783	35,1120	5,9255	10

Item-total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item- Total Correlation	Alpha if Item Deleted
Ag_1	13,4608	27,0030	,2399	,6250
Ag_3A	14,4036	31,2876	,3610	,5914
Ag_3B	14,4307	31,8027	,3276	,5975
Ag_2A	14,8012	32,2239	,2637	,6056
Ag_2B	14,3313	31,5441	,2931	,5997
Ag_2C	13,4819	24,8663	,4840	,5404
Ag_2D	13,4518	27,1886	,3670	,5769
Ag_2E	14,8223	32,3273	,3015	,6027

alpha = ,6217



Auswertung von Testergebnissen

- Analyse von Unterschieden und Zusammenhängen

Beispiele:

- *Geschlechtsunterschiede im Fach Deutsch*
- *Zusammenhang von Schülerleistung und Klassengröße*



Arten der Auswertung

- Deskriptive Analysen (Grafiken, Kennwerte bei den Stichprobendaten)
- Schließende Statistik (Inferenzstatistik) (Signifikanztests)

Beispiel:

Finden sich nicht nur bei den untersuchten Schülern sondern auch landesweit Geschlechtsunterschiede?



Literatur

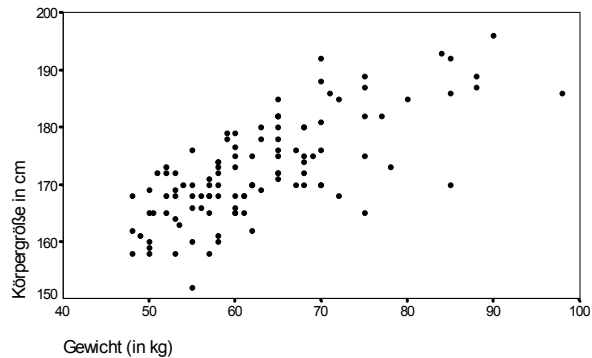
- Bortz, J. & Döring, N. (2001). Forschungsmethoden und Evaluation. Springer
- Bühner, M. (2004). Einführung in die Test- und Fragebogenkonstruktion. Pearson
- Nachtigall, Ch. & Wirtz (2004). Statistische Methoden für Psychologen. 2 Bände. Juventa



Anhang: Korrelation

Der Korrelationskoeffizient r ist eine Maßzahl für die Enge des Zusammenhanges zweier Merkmale

Beispiel: Zusammenhang von Gewicht und Größe einer Person



Anhang: Korrelation

Der Korrelationskoeffizient r liegt zwischen -1 und $+1$

$r = 0$ bedeutet: Kein (linearer) Zusammenhang

Zu beachten: Eine von Null verschiedene Korrelation muss keinen kausalen Zusammenhang bedeuten!

Die Korrelation im Beispiel (Körpergröße und Gewicht) liegt bei $0,7$.