



Fragebogen- und Testkonstruktion in der Online-Forschung

Dr. Meinald T. Thielsch

13. Juni 2008
Globalpark Innovations 2008



Diagnostisches Problem der (Online-)Forschung



Es werden in vielen Arbeiten der (Online-)Forschung diagnostische Prinzipien nicht berücksichtigt und Gütekriterien nicht bestimmt. Besonders relevant:

- **Grundregeln der Itemformulierung**
- **Grundregeln der Fragebogen-/Testkonstruktion**
- **Objektivität**
- **Reliabilität**
- **Validität**



Grundlegende These:

Die Qualität von Online-Fragebögen und -
Testverfahren kann erheblich durch eine
stärkere Beachtung von Grundlagen aus der
psychologischen Diagnostik verbessert
werden.



10 Stufen der Testentwicklung



10 Stufen der Testentwicklung



1. Anforderungsanalyse
2. Planung und Literaturrecherche
3. Eingrenzung des Merkmals und Arbeitsdefinition
4. Testentwurf
5. Überprüfung des Testentwurfs
6. Verteilungsanalyse
7. Itemanalyse und Selektion
8. Bestimmung der Gütekriterien
9. Revision des Tests
10. Normierung/Festlegung Cut-off-Werte



Schritt 1: Anforderungsanalyse



Am Anfang steht eine Fragestellung oder ein Konstrukt was untersucht werden soll. Drei Möglichkeiten dieses zunächst enger einzugrenzen (nach Schuler, 2005):

- Die erfahrungsgelitet-intuitive Methode (Beschäftigung mit der Tätigkeit / dem Material)
- Die analytisch-empirische Methode (formalisierte Vorgehensweise z. B. mittels Fragebogen)
- Die personenbezogen-empirische Methode (Zusammenhang zwischen Merkmalen der tätigen Personen und Kriterien wie z. B. der Leistungshöhe)



Schritt 2: Planung und Literaturrecherche



1. Überblicksartikel und Lehrbücher
2. Befragung von Experten
→ Online zumeist leicht & schnell gemacht!
3. Anforderungsbezogene Vorgehensweise



Schritt 3: Eingrenzung des Merkmals und Arbeitsdefinition



Deduktive Methode (rationale Strategie)

- Ableitung aus einer expliziten Theorie!
- Gezielte Sammlung von geeigneten Aufgaben, die als Indikatoren für das zu erfassende Konstrukt gelten können!



Schritt 3: Eingrenzung des Merkmals und Arbeitsdefinition



Induktive Methode (internale Strategie)

- Sammlung von relevant erscheinenden Items und
- nachträgliche Ordnung/Theorienbildung z. B. mittels Faktorenanalyse



Schritt 4: Testentwurf



- Welche Zielgruppe will ich ansprechen?
- Welche Art von Information will ich erheben?
- Welchem Zweck soll der Test dienen?
- Wichtig:
 - a) Repräsentative Itemmenge
 - b) Verständliche Instruktionen



Schritt 5: Überprüfung des Testentwurfs



Spezifika der Stichprobe:

Repräsentativ für die spätere Zielpopulation

Untergrenze $N = 100$

Optimalerweise $N > 300$

- ➔ Fallzahlen, die online zumeist deutlich leichter als offline zu erreichen sind.
- ➔ Zielgruppen können online ggf. spezifischer erreicht werden.



Schritt 6: Verteilungsanalyse



- Boden- oder Deckeneffekte?
- Bimodale Items?
- Ziel: Normalverteilte Items



Schritt 7: Itemanalyse und -selektion



- Schwierigkeit
- Streuung
- Trennschärfe



Schritt 8: Bestimmung der Gütekriterien



- Objektivität
 - Reliabilität
 - Validität
- ➔ Online hohe Objektivität erreichbar.
 - ➔ Retest zur Bestimmung der Retest-Reliabilität online weitaus einfacher zu realisieren.
 - ➔ Parallele Erhebung konvergenter/divergenter Validitätskriterien online umsetzbar.



Schritt 9: Revision des Tests



- Legen Ergebnisse der vorherigen Schritte eine Überarbeitung nahe?
 - Evaluation anhand der Zielgruppe (Ziehung einer weiteren Stichprobe im Sinne einer Kreuzvalidierung)
- ➔ Ziehung weiterer Stichproben online meist weniger aufwendig als offline



Schritt 10: Normierung / Cut-off-Werte



- Normierung: Repräsentative Abbildung der Zielpopulation
- Festlegung von Cut-off-Werten



Stufen der Testentwicklung und Befragungsformen



Je nach Befragungsform können nicht stets alle Stufen durchlaufen werden:

Stufen der Testentwicklung	Ad hoc/einmalige Befragung	Standardisierter Fragebogen	Standardisierter Test
Anforderungsanalyse	X	X	X
Planung und Literaturrecherche	X	X	X
Eingrenzung des Merkmals und Arbeitsdefinition	zum Teil	X	X
Testentwurf	X	X	X
Überprüfung des Testentwurfs		X	X
Verteilungsanalyse	X	X	X
Itemanalyse und Selektion	zum Teil	X	X
Bestimmung der Gütekriterien	zum Teil	X	X
Revision	-	X	X
Normierung/Festlegung Cut-off-Werte	-	Bei Bedarf	X



Zentrale Gütekriterien



Die Objektivität betrifft den Grad, inwieweit das Ergebnis eines Tests unabhängig ist von der Person des Untersuchers.

- Durchführungsobjektivität
- Auswertungsobjektivität
- Interpretationsobjektivität

➔ Online-Verfahren können sehr objektiv sein! Objektivität ist aber nicht automatisch gegeben, nur weil etwas am Computer gemacht wird (Problem: Durchführungsobjektivität).



Reliabilität beschreibt die **Genauigkeit, mit der ein Testinstrument eine Merkmalsdimension erfasst!**

Die Betrachtung erfolgt unter Vernachlässigung der Frage, ob es sich dabei auch tatsächlich um diejenige Merkmalsdimension handelt, deren Erfassung intendiert ist!

➔ Instrumentelle Genauigkeit

In welchem Maße sind die Testwerte durch **Störeinflüsse** und **Fehler** belastet?

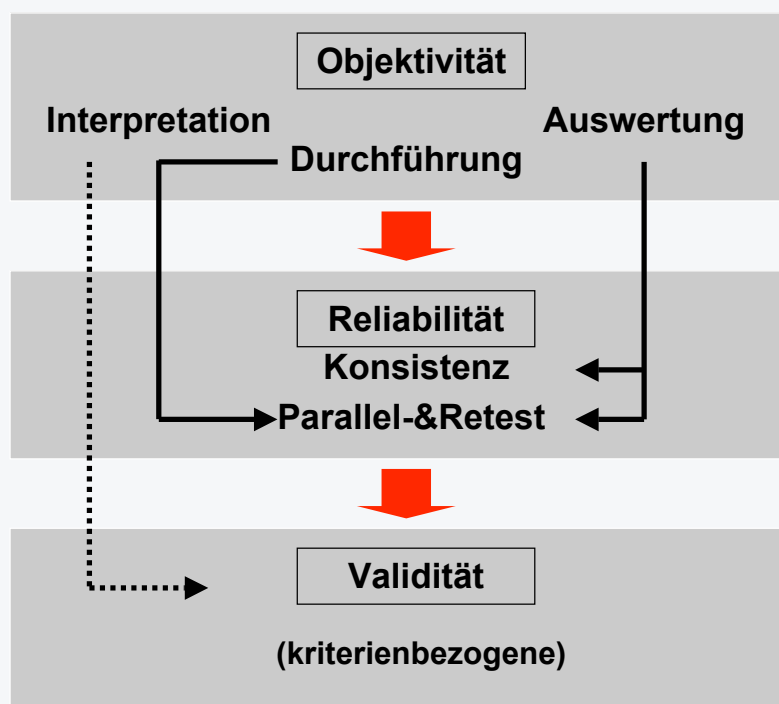
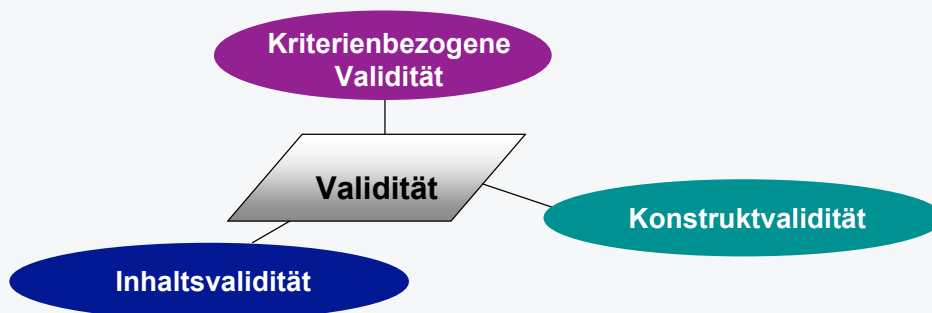
Achtung:

Das hier oft angegebene Cronbachs Alpha ist abhängig von der Menge der Items! Durch Hinzunahme von Items kann dieses Maß leicht erhöht werden. Vorsicht bei der Interpretation ist geboten!



Die Validität eines Testes gibt die Genauigkeit an, mit der dieser Test dasjenige Merkmal, das er zu messen vorgibt, auch tatsächlich misst.

→ Schluss vom Verhalten in der Testsituation auf Verhalten außerhalb der Testsituation.





Kennwert	Niedrig	Mittel	Hoch
Schwierigkeit	> .80	.80 - .20	< .20
Trennschärfe (korrigiert)	< .30	.30 - .50	> .50
Objektivität (Auswerter)	< .60	.60 - .90	> .90
Reliabilität	< .80	.80 - .90	> .90
Validität (unkorrigiert)	< .40	.40 - .60	> .60
Größe der Eichstichprobe	< 150	150 - 300	> 300

Nach Fissini, 1997; nach Weise, 1975



Praktische Hinweise



1. Online-Befragungen sind schnell und einfach durchführbar
→ trotzdem intensive Itemkonstruktion, Vortests, Itemrevision!
2. Augenmerk auf die technische Umsetzung der Fragebögen
→ Usability, Gestaltung/Ästhetik, Fehlermeldungen, ...
3. Nutzung von Eingabekontrollen (Plausi-Checks, DAC's), Erhebung der expliziten Ablehnung von Items.
4. Nutzung non-reaktiver Daten
→ Drop-Out-Statistiken zur Fragebogenverbesserung heranziehen
5. Stichwort: Lesen am Bildschirm?
6. Stichwort: Kontrolle über die Testpersonen?



7. Äquivalenz zwischen offline und online Versionen eines Verfahrens?
8. Nutzung angemessener statistischer Methoden, insbesondere bei der faktorenanalytischen Auswertung von Fragebögen!
9. Zusätzlich zur regulären Itemanalyse: Nutzung des Makros ALPHAMAX (Hayes, 2005).
- 10. Trennung von Test-/Fragebogenentwicklung und der eigentlichen Forschungsfrage!**



Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion* (2. Aufl.). München: Pearson Studium.

Hayes, A. F. (2005). *A computational tool for survey shortening applicable to composite attitude, opinion, and personality measurement scales*. Paper presented at the *Midwestern Association for Public Opinion Research*, Chicago, IL.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and velicer's map test. *Behavior Research Methods Instruments & Computers*, 32(3), 396-402.



Vielen Dank für Ihre Aufmerksamkeit!

Kontakt:

Dr. Meinald T. Thielsch, Dipl.-Psych.
Westfälische Wilhelms-Universität
Psychologisches Institut I
Flüednerstr. 21
48149 Münster
thielsch@uni-muenster.de

<http://www.meinald.de>