

Testtheorie und Testkonstruktion

von

Michael Eid

und Katharina Schmidt

HOGREFE



GÖTTINGEN · BERN · WIEN · PARIS · OXFORD · PRAG
TORONTO · BOSTON · AMSTERDAM · KOPENHAGEN
STOCKHOLM · FLORENZ · HELSINKI

Inhaltsverzeichnis

Vorwort	15
1 Grundfragen der Testtheorie und Testkonstruktion ...	19
1.1 Was sind psychologische Messungen?	21
1.2 Grundidee psychometrischer Modelle am Beispiel des Rasch- Modells	25
1.3 Was ist ein psychologischer Test?	29
1.4 Grundlegende Annahmen und Eigenschaften psychometrischer Modelle	30
1.5 Unterscheidungsmerkmale testtheoretischer Modelle	34
1.5.1 Klassifikation psychometrischer Modelle	38
1.5.2 Begriffliche Abgrenzungen	39
Zusammenfassung	40
Fragen	41
2 Wesentliche Schritte der Konstruktion psycholo- gischer Tests	43
2.1 Festlegung des zu erfassenden Konstrukts	46
2.1.1 Definition der Konstruktvalidität	49
2.1.2 Strategien zur Untersuchung der Konstruktvalidität	53
2.2 Erstellung eines Itempools: Testkonstruktionsprinzipien und Validitätsfacetten	56
2.2.1 Rationale Testkonstruktion	57
2.2.2 Inhaltsvalidität (Kontentvalidität): Kriteriumsorientierte und induktive Testkonstruktion	60
2.2.2.1 Kriteriumsorientierte Tests	61
2.2.2.2 Induktive Methode der Testkonstruktion	62
2.2.3 Kriteriumsvalidität: Externale Testkonstruktion	64
2.2.4 Augenscheinvalidität	64
2.3 Auswahl eines Antwortformats: Objektivität und Präzision	65
2.4 Gütekriterien der Itemauswahl	67
2.5 Testanalyse	70
2.6 Skalierung und Normierung	71

2.7	Testdokumentation	74
2.8	Weitere Untersuchungen zur Güte	74
	Zusammenfassung	75
	Fragen	76
3	Itemkonstruktion	77
3.1	Noch Fragen? – Konstruktion des Itemstamms	79
3.1.1	Itemgegenstände	79
3.1.2	Itemformulierung und -anordnung	86
3.1.2.1	Konversationsmaximen	87
3.1.2.2	Aufbau des Verfahrens: Effekte der Itemreihenfolge	88
3.1.2.3	Semantische Aspekte der Itemformulierung	92
3.2	„Ja“-„Nein“-„Äh ...“-„Kommt ganz drauf an“: Antwortformate	97
3.2.1	Freie Antwortformate	97
3.2.2	Gebundene Antwortformate	100
3.2.2.1	Ordnungsaufgaben	100
3.2.2.2	Auswahlaufgaben	104
3.2.3	Atypische Antwortformate	123
3.3	Von der Antwort zur Variablen: Itemkodierung	124
	Zusammenfassung	126
	Fragen	126
4	Eindimensionale Modelle für dichotome Antwort- variablen	129
4.1	Dichotome Variablen	130
4.2	Univariate Verteilung dichotomer Antwortvariablen	131
4.3	Unabhängigkeit dichotomer Antwortvariablen	133
4.4	Zusammenhang dichotomer Variablen	135
4.5	Das Rasch-Modell	144
4.5.1	Erste Modellannahme: Rasch-Homogenität	144
4.5.1.1	Skaleneigenschaften und Normierung	148
4.5.1.2	Logit-Transformation	152
4.5.1.3	Spezifische Objektivität	153
4.5.2	Zweite Modellannahme: Bedingte stochastische Unabhängigkeit	154
4.5.3	Methoden der Schätzung der Itemparameter und Personen- werte	156

4.5.3.1	Das Grundprinzip der Maximum-Likelihood-Schätzung	156
4.5.3.2	Die unbedingte ML-Schätzung	159
4.5.3.3	Schätzung der Itemparameter: Bedingte ML-Schätzung	161
4.5.3.4	Schätzung der Itemparameter: Marginale ML-Schätzung	169
4.5.3.5	Weitere Methoden der Itemparameterschätzung	172
4.5.3.6	Schätzung der Personenwerte: Unbedingte und gewichtete ML-Schätzung	173
4.5.3.7	Schätzung der Personenwerte: Weitere Ansätze	180
4.5.3.8	Schätzung der Personenwerte: Reliabilität	181
4.5.4	Methoden der Überprüfung der Modellgültigkeit	182
4.5.4.1	Gleichheit der Itemparameter in Subpopulationen	183
4.5.4.1.1	Grafischer Modelltest	183
4.5.4.1.2	Bedingter Likelihood-Quotienten-Test	185
4.5.4.1.3	Wald-Test	187
4.5.4.1.4	Mischverteilungs-Rasch-Analyse	189
4.5.4.2	Globale Modellgültigkeit: Wahrscheinlichkeitsverteilung der Antwortmuster	192
4.5.4.3	Globale Modellgültigkeit: Likelihood-Quotienten-Test	195
4.5.4.4	Gleichheit der Personenwerte in reduzierten Rasch-Modellen . . .	197
4.5.4.5	Weitere Tests	198
4.5.4.6	Identifikation von abweichenden Items	198
4.5.4.7	Identifikation von abweichenden Personen	200
4.5.4.8	Bewertung der Modellgüte: Empfehlungen	201
4.6	Weitere Modelle für dichotome Antwortvariablen	203
4.6.1	Das zweiparametrische logistische Modell	204
4.6.2	Das dreiparametrische logistische Modell	209
4.6.3	Power- vs. Speed-Tests	211
4.7	Benötigte Stichprobengröße	212
4.8	Computerprogramme	214
4.9	Weitere Ansätze	214
4.9.1	Schwierigkeitskoeffizienten	215
4.9.2	Trennschärfe	216
	Zusammenfassung	219
	Fragen	221
5	Eindimensionale Modelle für Antwortvariablen mit geordneten Antwortkategorien	223
5.1	Antwortvariablen mit geordneten Antwortkategorien	224
5.2	Univariate Verteilung kategorialer Variablen mit geordneten Antwortkategorien	224

5.3	Unabhängigkeit und Zusammenhang von Antwortvariablen mit geordneten Antwortkategorien	228
5.3.1	Unabhängigkeit	228
5.3.2	Zusammenhangsmaße	229
5.4	Das Partial-Credit-Modell	230
5.4.1	Erste Modellannahme: Item- und Kategorienhomogenität	231
5.4.1.1	Kategorienwahrscheinlichkeiten und -charakteristiken	234
5.4.1.2	Itemcharakteristik	239
5.4.1.3	Skaleneigenschaften und Normierung	240
5.4.2	Zweite Modellannahme: Bedingte (lokale) stochastische Unabhängigkeit	241
5.4.3	Methoden der Schätzung der Itemparameter und Personenwerte	242
5.4.4	Methoden der Überprüfung der Modellgültigkeit	243
5.4.4.1	Gleichheit der Itemparameter in Subpopulationen	243
5.4.4.2	Wahrscheinlichkeitsverteilung der Antwortmuster und globale Modellgültigkeit	244
5.4.4.3	Gleichheit der Personenwerte in reduzierten Rasch-Modellen	244
5.4.4.4	Identifikation von abweichenden Items und Personen	244
5.4.5	Anwendungsbeispiel	244
5.4.6	Spezialfälle des Partial-Credit-Modells	249
5.4.6.1	Ratingskalenmodell	249
5.4.6.2	Äquidistanzmodell	251
5.4.6.3	Dispersionsmodell	252
5.5	Weitere Modelle	253
5.6	Benötigte Stichprobengröße	254
5.7	Computerprogramme	254
5.8	Weitere Ansätze	254
	Zusammenfassung	255
	Fragen	256
6	Eindimensionale Modelle für metrische Antwortvariablen	257
6.1	Metrische Variablen	258
6.2	Univariate Verteilung metrischer Antwortvariablen	260
6.3	Unabhängigkeit und Zusammenhang von metrischen Antwortvariablen	261
6.3.1	Unabhängigkeit	261
6.3.2	Zusammenhangsmaße	262
6.4	Grundzüge der Klassischen Testtheorie	265

6.4.1	Wahrer Wert und Messfehler	265
6.4.2	Bestimmung des wahren Wertes	269
6.4.2.1	Beobachteter Wert als Schätzwert für den wahren Wert	269
6.4.2.2	Regressionsanalytische Schätzung des wahren Wertes	270
6.5	Das Modell essenziell τ -äquivalenter Variablen	273
6.5.1	Erste Modellannahme: Essenzielle τ -Äquivalenz	273
6.5.2	Zweite Modellannahme: Unkorreliertheit der Fehler- variablen	278
6.5.3	Methoden der Schätzung der Itemparameter und Personen- werte	279
6.5.3.1	Leichtigkeitsparameter	281
6.5.3.2	Varianz von η und Fehlervarianzen	282
6.5.3.3	Reliabilität	284
6.5.3.4	Schätzung der latenten Personenwerte: Maximum-Likelihood- Schätzung	290
6.5.3.5	Schätzung der latenten Personenwerte: Weitere Schätz- methoden	293
6.5.4	Überprüfung der Modellgültigkeit des Modells essenziell τ -äquivalenter Variablen	295
6.5.4.1	Kovarianzstruktur	295
6.5.4.2	Gleichheit der Leichtigkeitsparameter in Subpopulationen	297
6.5.4.3	Weitere Möglichkeiten der Modellgeltungsüberprüfung	299
6.5.5	Spezialfälle des Modells essenziell τ -äquivalenter Variablen ...	300
6.5.5.1	Modell τ -äquivalenter Variablen	300
6.5.5.1.1	Überprüfung der Modellgültigkeit	301
6.5.5.1.2	Überprüfung der Modellgültigkeit: χ^2 -Differenztest	302
6.5.5.2	Modell essenziell τ -paralleler Variablen	304
6.5.5.2.1	Reliabilität der Summenvariablen	306
6.5.5.2.2	Schätzung der latenten Personenwerte: Maximum-Likelihood- Schätzung	307
6.5.5.2.3	Schätzung der latenten Personenwerte: Bayes-Modal- Schätzung	309
6.5.5.2.4	Überprüfung der Modellgültigkeit	310
6.5.5.3	Modell τ -paralleler Variablen	311
6.6	Modell τ -kongenerischer Variablen	312
6.6.1	Normierung	315
6.6.2	Methoden der Schätzung der Itemparameter und Personen- werte	316
6.6.2.1	Leichtigkeitsparameter	316
6.6.2.2	Diskriminationsparameter	317
6.6.2.3	Varianz von η und Fehlervarianzen	319
6.6.2.4	Konfidenzintervalle	319

6.6.2.5	Reliabilität der Summenvariablen	322
6.6.2.6	Schätzung der latenten Personenwerte	322
6.6.3	Überprüfung der Modellgültigkeit	323
6.7	Vergleich der verschiedenen Modelle	325
6.8	Benötigte Stichprobengröße	330
6.8.1	Anzahl der Items	330
6.8.2	Anzahl der Personen	330
6.9	Computerprogramme	332
6.10	Analyse von Reaktionszeiten	332
6.11	Weitere Ansätze	333
6.12	Klassische Testtheorie und Testkonstruktion	334
	Zusammenfassung	335
	Fragen	338
7	Einführung in mehrdimensionale Testmodelle	341
7.1	Mehrdimensionale Modelle und Testvalidierung	342
7.2	Multikomponentenmodelle	346
7.3	Faktorenanalytisches Modell	349
7.3.1	Konfirmatorische Faktorenanalyse	351
7.3.2	Exploratorische Faktorenanalyse	352
7.4	Faktorenanalyse für ordinale Variablen	353
	Zusammenfassung	357
	Fragen	358
8	Interpretation und Normierung von Testwerten	359
8.1	Vergleich der Testergebnisse mit den Ergebnissen anderer Personen einer Bezugsgruppe (Normpopulation)	360
8.1.1	Lineare Transformationen	361
8.1.1.1	z-Transformation	362
8.1.1.2	Transformationen der z-Werte	363
8.1.2	Nicht lineare Transformationen	366
8.1.2.1	Prozentrangwerte	366
8.1.2.2	Weitere nicht lineare Transformationen	368
8.1.3	Normalisierende Transformationen	371
8.1.4	Vergleich der verschiedenen Transformationen	374
8.1.5	Bestimmung der Anzahl der Skalenpunkte	375
8.1.6	Andere Normierungssysteme	377
8.1.7	Eichstichprobe	378

8.1.7.1	Arten von Normen und Stichproben	378
8.1.7.2	Bestimmung der Stichprobengröße	379
8.1.8	Vergleich zweier Personen	382
8.2	Vergleich der Testergebnisse derselben Person in mehreren Tests (Profilanalyse)	383
8.3	Vergleich der Testergebnisse derselben Person in demselben Test zu einer anderen Messgelegenheit	385
8.4	Kriteriumsorientierte Interpretation von Testwerten	386
	Zusammenfassung	392
	Fragen	393
	Anhang	395
	Literatur	397
	Griechisches Alphabet	411
	Glossar	412
	Sachregister	429