

Jürgen Rost

Lehrbuch Testtheorie – Testkonstruktion

Zweite, vollständig überarbeitete
und erweiterte Auflage

6.3
ROS
3



Verlag Hans Huber
Bern · Göttingen · Toronto · Seattle

Technische Universität Darmstadt
Fachbereich 03
Institut für Psychologie
Bibliothek
Alexanderstrasse 10; 64283 Darmstadt

Inv. Nr. 9/104274

Inhaltsverzeichnis

	Vorwort zur zweiten Auflage	9
	Über dieses Buch	11
1.	Was ist Testtheorie?	17
1.1	Der Gegenstand der Testtheorie	17
1.1.1	Was ist ein Test?	17
1.1.2	Warum eine Theorie über Tests?	21
1.2	Die wissenschaftstheoretischen Grundlagen der Testtheorie	23
1.2.1	Was sind Theorien?	23
1.2.2	Was ist ein formales Modell?	25
1.2.3	Was sind Testmodelle?	28
1.2.4	Was erklären Theorien über das Testverhalten?	29
2.	Testkonstruktion	32
2.1	Gütekriterien für Tests	33
2.1.1	Validität	34
2.1.2	Reliabilität und Messgenauigkeit	36
2.1.3	Objektivität	39
2.1.4	Norminierung	40
2.2	Arten von Tests	43
2.2.1	Leistungstests	43
2.2.2	Persönlichkeitsfragebögen	44
2.2.3	Objektive Persönlichkeitstests	45
2.2.4	Projektive Tests	46
2.2.5	Situationsfragebögen	48
2.2.6	Einstellungstests	49
2.2.7	Motivations- und Interessensfragebögen	51
2.2.8	Verhaltensfragebögen	52
2.3	Itemkonstruktion	55
2.3.1	Itemuniversum und Itemstichprobe	55
2.3.2	Die sprachliche Formulierung der Items	57
2.3.3	Arten von Antwortformaten	59
	2.3.3.1 Freie Antwortformate	60
	2.3.3.2 Gebundene Antwortformate	61
	2.3.3.3 Ratingformate	64
2.3.4	Die Zusammenstellung des Tests	69

2.4	Datenerhebung	72
2.4.1	Stichprobenprobleme	72
2.4.2	Durchführungsprobleme	75
2.5	Kodierung der Antworten	78
2.5.1	Die Signierung freier Antworten	79
2.5.2	Die Kodierung von Antwortkategorien	83
3	Testmodelle	89
3.1	Modelle für dichotome Itemantworten	90
3.1.1	Modelle mit quantitativer Personenvariable	96
3.1.1.1	Stufenförmige Itemfunktionen	100
3.1.1.1.1	Die Guttman-Skala: Der Sprung von Null auf Eins	100
3.1.1.1.2	Antwortfehlermodelle: Irrtum und Raten	105
3.1.1.2	Kontinuierlich ansteigende Itemfunktionen	107
3.1.1.2.1	Das Binomialmodell: Eine Gerade als Itemfunktion	108
3.1.1.2.2	Das Rasch-Modell: Parallele Itemfunktionen	115
3.1.1.2.3	Item Response Theorie (IRT): Rate- und Trennschärfeparameter	133
3.1.1.2.4	Die Mokken-Analyse: Überschneidungsfreie Itemfunktionen	136
3.1.1.3	Nichtmonotone eingipflige Itemfunktionen	139
3.1.1.3.1	Das Parallelogramm-Modell: Kastenförmige Itemfunktionen	140
3.1.1.3.2	Kontinuierliche, eingipflige Itemfunktionen	142
3.1.2	Modelle mit qualitativer Personenvariable	148
3.1.2.1	Deterministische Klassen: Verbotene Antwortmuster	152
3.1.2.2	Die Analyse latenter Klassen: Wahrscheinliche Antwortmuster	154
3.1.2.3	Das Fixieren und Gleichsetzen von Parametern	162
3.1.2.4	Lokalisierte Klassen: Punkte auf einem Kontinuum	167
3.1.3	Modelle mit quantitativer und qualitativer Personenvariable	171
3.1.3.1	Das mixed Rasch-Modell: Klassifizieren und quantifizieren zugleich	174
3.1.3.2	Hybrid-Modelle: In jeder Klasse ein anderes Modell	180

3.2	Modelle für nominale Itemantworten	183
3.2.1	Klassenanalyse nominaler Daten	187
3.2.2	Das mehrdimensionale Rasch-Modell	192
3.3	Modelle für ordinale Itemantworten	201
3.3.1	Das ordinale Rasch-Modell	203
3.3.2	Modelle für Ratingskalen	215
3.3.3	Klassenanalyse ordinaler Daten	226
3.3.4	Klassenmodelle für Ratingskalen	233
3.3.5	Mixed Rasch-Modelle für ordinale Daten	240
3.4	Itemkomponenten- und mehrdimensionale Modelle	250
3.4.1	Linear-logistische Testmodelle: Komponenten der Aufgabenschwierigkeit	253
3.4.2	Mehrdimensionale Modelle	259
3.4.3	Linear-logistische Klassenanalyse	268
3.5.	Modelle der Veränderungsmessung	272
3.5.1	Klassische Probleme der Veränderungsmessung	273
3.5.1.1	Die Reliabilität von Differenzwerten	274
3.5.1.2	Die Korrelation von Anfangswert und Differenzwert	276
3.5.1.3	Messen Vor- und Nachtest dasselbe?	280
3.5.2	Dreifaktorielle Testmodelle: Personen, Items und Zeitpunkte	281
3.5.3	Dynamische Modelle: Lernen während der Testbearbeitung	287
3.5.3.1	Personenspezifisches Lernen	288
3.5.3.2	Itemspezifisches Lernen	289
3.5.3.3	Globales reaktionskontingentes Lernen	291
3.5.4	Die Messung der Wirksamkeit von Maßnahmen	293
4.	Parameterschätzung	301
4.1	Die Maximum-Likelihood-Methode	303
4.2	Parameterschätzung bei Rasch-Modellen	309
4.3	Parameterschätzung für die latente Klassenanalyse	318
4.4	Parameterschätzung bei missing data	324

5.	Modellgeltungstests	330
5.1	Likelihoodquotienten und andere globale Fit-Statistiken	331
5.2	Informationstheoretische Maße für den Modellvergleich	339
5.3	Spezielle Tests für Rasch-Modelle	345
5.3.1	Prüfung der Personenhomogenität	347
5.3.2	Prüfung der Itemhomogenität	351
6.	Testoptimierung	356
6.1	Die Qualität der Personenmesswerte	356
6.1.1	Die Genauigkeit der Personenparameter	356
6.1.2	Modellkonforme und abweichende Antwortmuster	363
6.1.3	Die Zuordnungswahrscheinlichkeit bei Klassenmodellen	366
6.2	Die Qualität der Items	369
6.3	Die Reliabilität des Tests	376
6.3.1	Methoden der Reliabilitätsbestimmung	377
6.3.2	Reliabilitätssteigerung durch Testverlängerung	383
6.4	Die Validität des Tests	385
6.4.1	Die Berechnung der externen Validität	385
6.4.2	Maximal erreichbare Validitäten	389
6.4.3	Das Reliabilitäts-Validitäts-Dilemma	392
6.5	Die Normierung der Testergebnisse	395
	Literatur	399
	Chi-quadrat Tabelle	418
	Notationstabelle	419
	Stichwortverzeichnis	422