

Markus Pospeschill

# **Testtheorie, Testkonstruktion, Testevaluation**

Mit 71 Abbildungen und 77 Fragen zur Wiederholung

Ernst Reinhardt Verlag München Basel

PD Dr. *Markus Pospeschill* lehrt und forscht als Akademischer Oberrat im Fach Psychologie an der Universität des Saarlandes u. a. in den Bereichen Methodenlehre, Forschungsmethoden und Psychodiagnostik. Vom Autor außerdem im Ernst Reinhardt Verlag erhältlich: Pospeschill/Spinath: „Psychologische Diagnostik“ (UTB basic).

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

UTB ISBN 978-3-8252-3431-7

ISBN 978-3-497-02161-1

© 2010 by Ernst Reinhardt, GmbH & Co KG, Verlag, München

Dieses Werk, einschließlich aller seiner Teile, ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne schriftliche Zustimmung der Ernst Reinhardt GmbH & Co KG, München, unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen in andere Sprachen, Mikroverfilmungen und für die Einspeicherung und Verarbeitung in elektronischen Systemen.

Printed in Germany

Einbandgestaltung: Atelier Reichert, Stuttgart

Covermotiv © appler – Fotolia.com

Satz: Arnold & Domnick, Leipzig

ISBN **978-3-8252-3431-7** (UTB-Bestellnummer)

Ernst Reinhardt Verlag, Kemnatenstr. 46, D-80639 München

Net: [www.reinhardt-verlag.de](http://www.reinhardt-verlag.de) E-Mail: [info@reinhardt-verlag.de](mailto:info@reinhardt-verlag.de)

# Inhalt

<b>Hinweise zur Benutzung dieses Lehrbuchs</b> . . . . .	9
<b>Vorwort</b> . . . . .	10
<b>Road (Concept) Map</b> . . . . .	12

## I Testtheoretische Grundlagen

<b>1 Kriterien wissenschaftlicher Testgüte</b> . . . . .	16
1.1 Objektivität . . . . .	18
1.2 Reliabilität . . . . .	21
1.3 Validität . . . . .	24
1.4 Normierung/Testeichung . . . . .	28
1.5 Skalierung . . . . .	29
1.6 Ökonomie/Wirtschaftlichkeit . . . . .	30
1.7 Nützlichkeit/Utilität . . . . .	30
1.8 Zumutbarkeit . . . . .	30
1.9 Unverfälschbarkeit . . . . .	31
1.10 Testfairness . . . . .	31
1.11 Vergleichbarkeit . . . . .	32
1.12 Transparenz . . . . .	32
<b>2 Testentwicklung</b> . . . . .	35
2.1 Testplanung . . . . .	35
2.1.1 Festlegung der Erhebungsmerkmale . . . . .	36
2.1.2 Festlegung der Testart . . . . .	37
2.1.3 Festlegung des Geltungsbereichs . . . . .	38
2.2 Strategien zur Testentwicklung . . . . .	40
2.2.1 Rationale Konstruktionsstrategie . . . . .	40
2.2.2 Externale Konstruktionsstrategie . . . . .	41
2.2.3 Induktive Konstruktionsstrategie . . . . .	42
2.2.4 Prototypenansatz . . . . .	43

2.2.5	Intuitive Konstruktionsstrategie . . . . .	43
2.3	Itemkonstruktion . . . . .	44
2.3.1	Items mit offenem Antwortformat . . . . .	45
2.3.2	Items mit gebundenem Antwortformat . . . . .	47
2.4	Fehlerquellen bei der Beantwortung von Items . . . . .	58
2.4.1	Soziale Erwünschtheit . . . . .	58
2.4.2	Antworttendenzen . . . . .	61
2.5	Prinzipien der Itemformulierung . . . . .	63
2.5.1	Itemarten . . . . .	63
2.5.2	Itemverständlichkeit . . . . .	64
2.5.3	Itemvariabilität . . . . .	66
2.6	Erstellung einer Test-Betaversion . . . . .	67
2.6.1	Testanweisung . . . . .	67
2.6.2	Anordnung des Itempools . . . . .	68
2.6.3	Erprobung der Test-Betaversion . . . . .	69
<b>3</b>	<b>Testevaluation</b> . . . . .	<b>72</b>
3.1	Datenstruktur der Messungen . . . . .	72
3.2	Itemschwierigkeit . . . . .	73
3.2.1	Itemschwierigkeit bei Leistungstests . . . . .	74
3.2.2	Itemschwierigkeit bei Persönlichkeitstests . . . . .	76
3.3	Itemvarianz . . . . .	77
3.4	Itemtrennschärfe . . . . .	80
3.4.1	Berechnungen der Trennschärfe . . . . .	80
3.4.2	Interpretation der Trennschärfe . . . . .	82
3.4.3	Part-Whole-Korrektur . . . . .	83
3.4.4	Selektionskennwert . . . . .	83
3.5	Kriterien der Itemselektion . . . . .	85
3.6	Testwertermittlung . . . . .	86
3.6.1	Testwertermittlung bei Leistungstests . . . . .	86
3.6.2	Testwertermittlung bei Persönlichkeitstests . . . . .	87
3.6.3	Statistische Maße . . . . .	88
3.6.4	Normalisierung schiefer Testwertverteilungen . . . . .	93
<b>4</b>	<b>Klassische Testtheorie</b> . . . . .	<b>97</b>
4.1	Axiome der Klassischen Testtheorie . . . . .	97
4.1.1	Existenzaxiom . . . . .	97

4.1.2	Verknüpfungsaxiom . . . . .	98
4.1.3	Unabhängigkeitsaxiom . . . . .	98
4.1.4	Zusatzannahmen . . . . .	98
4.1.5	Erläuterungen zum Konzept des Messfehlers . . . . .	99
4.2	Bestimmung des wahren Testwertes . . . . .	101
4.3	Bestimmung der wahren Varianz und Fehlervarianz . . . . .	102
4.4	Bestimmung der Reliabilität . . . . .	103
4.5	Reliabilität und Testlänge . . . . .	104
4.6	Standardmessfehler . . . . .	105
4.7	Konfidenzintervall des wahren Testwertes . . . . .	107
4.8	Minderungskorrekturen . . . . .	107
4.9	Kritik an der Klassischen Testtheorie . . . . .	109
<b>5</b>	<b>Probabilistische Testtheorie . . . . .</b>	<b>114</b>
5.1	Manifeste und latente Variablen . . . . .	114
5.2	Lokale stochastische Unabhängigkeit . . . . .	116
5.3	Latent-Trait-Modelle (LTM) . . . . .	119
5.3.1	Einparameter-Logistisches Modell (1PL) . . . . .	122
5.3.2	Zweiparameter-Logistisches Modell (2PL) . . . . .	132
5.3.3	Dreiparameter-Logistisches Modell (3PL) . . . . .	133
5.4	Latent-Class-Modelle (LCA) . . . . .	135
5.4.1	Allgemeine Modellgleichung der LCA . . . . .	136
5.4.2	Parameterschätzung und Modelltests in der LCA . . . . .	140
5.5	Modellerweiterungen . . . . .	148
5.5.1	Mixed-Rasch-Modelle . . . . .	148
5.5.2	Ordinales Rasch-Modell . . . . .	150
5.6	Adaptives Testen . . . . .	153
5.6.1	Strategien der Itemauswahl . . . . .	154
5.6.2	Vor- und Nachteile adaptiven Testens . . . . .	156
<b>II</b>	<b>Methodische Überprüfung</b>	
<b>6</b>	<b>Testgütebestimmung und Interpretation von Testresultaten . . . . .</b>	<b>160</b>
6.1	Reliabilität . . . . .	160
6.1.1	Test-Retest-Reliabilität . . . . .	161

6.1.2	Paralleltest-Reliabilität . . . . .	163
6.1.3	Splithalf-Reliabilität . . . . .	165
6.1.4	Interne Konsistenz . . . . .	167
6.1.5	Zusammenfassende Würdigung. . . . .	170
6.2	Validität. . . . .	176
6.2.1	Inhaltsvalidität . . . . .	178
6.2.2	Kriteriumsvalidität . . . . .	180
6.2.3	Konstruktvalidität . . . . .	182
6.2.4	Zusammenfassende Würdigung. . . . .	185
6.3	Norm- vs. Kriteriumsorientierung. . . . .	187
6.3.1	Normorientierte Testwertinterpretation . . . . .	188
6.3.2	Kriteriumsorientierte Testwertinterpretation . . . . .	192
<b>7</b>	<b>Faktorenanalyse</b> . . . . .	<b>198</b>
7.1	Exploratorische Faktorenanalyse. . . . .	198
7.1.1	Extraktionsmethode. . . . .	199
7.1.2	Fundamentaltheorem . . . . .	200
7.1.3	Abbruchkriterien . . . . .	203
7.1.4	Rotationsmethoden . . . . .	205
7.1.5	Voraussetzungen zur Durchführung. . . . .	207
7.2	Konfirmatorische Faktorenanalyse . . . . .	210
7.2.1	Modellspezifikation . . . . .	212
7.2.2	Modellidentifikation . . . . .	214
7.2.3	Modellevaluation. . . . .	217
7.2.4	Voraussetzungen zur Durchführung. . . . .	223
7.2.5	Konfirmatorische MTMM-Analyse . . . . .	228
	<b>Anhang</b> . . . . .	<b>234</b>
	Notationstabelle lateinischer Buchstaben . . . . .	234
	Notationstabelle griechischer Buchstaben. . . . .	238
	Literatur. . . . .	239
	Sachregister. . . . .	244