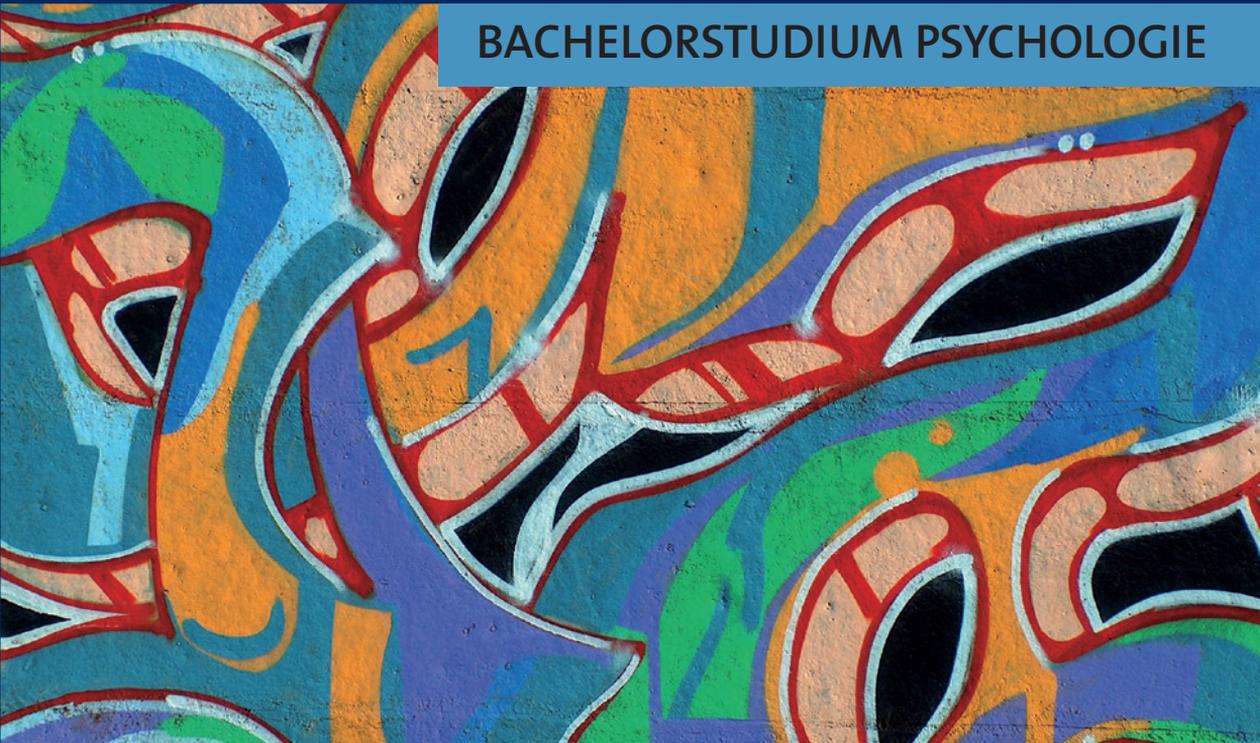


Michael Eid · Katharina Schmidt

Testtheorie und Testkonstruktion

BACHELORSTUDIUM PSYCHOLOGIE



Testtheorie und Testkonstruktion

Bachelorstudium Psychologie

Testtheorie und Testkonstruktion

von Prof. Dr. Michael Eid und Katharina Schmidt

Herausgeber der Reihe:

Prof. Dr. Eva Bamberg, Prof. Dr. Hans-Werner Bierhoff,
Prof. Dr. Alexander Grob, Prof. Dr. Franz Petermann

Testtheorie und Testkonstruktion

von

Michael Eid

und Katharina Schmidt

HOGREFE



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

Prof. Dr. Michael Eid, geb. 1963. Seit 2006 Professor für Methoden und Evaluation an der Freien Universität Berlin.

Dipl.-Psych. Katharina Schmidt, geb. 1963. Seit 2006 wissenschaftliche Mitarbeiterin im Arbeitsbereich für psychologische Diagnostik und Differentielle und Persönlichkeitspsychologie an der Freien Universität Berlin.



Informationen und Zusatzmaterialien zu diesem Buch finden Sie unter www.hogrefe.de/buecher/lehrbuecher/psychlehrbuchplus

© 2014 Hogrefe Verlag GmbH & Co. KG
Göttingen • Bern • Wien • Paris • Oxford • Prag • Toronto • Boston
Amsterdam • Kopenhagen • Stockholm • Florenz • Helsinki
Merkelstraße 3, 37085 Göttingen

<http://www.hogrefe.de>

Aktuelle Informationen • Weitere Titel zum Thema • Ergänzende Materialien

Copyright-Hinweis:

Das E-Book einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlags unzulässig und strafbar.

Der Nutzer verpflichtet sich, die Urheberrechte anzuerkennen und einzuhalten.

Umschlagabbildung: © Digitalstock – L. Halbauer
Satz: ARThür Grafik-Design & Kunst, Weimar
Format: PDF

ISBN 978-3-8409-2161-2

Nutzungsbedingungen:

Der Erwerber erhält ein einfaches und nicht übertragbares Nutzungsrecht, das ihn zum privaten Gebrauch des E-Books und all der dazugehörigen Dateien berechtigt.

Der Inhalt dieses E-Books darf von dem Kunden vorbehaltlich abweichender zwingender gesetzlicher Regeln weder inhaltlich noch redaktionell verändert werden. Insbesondere darf er Urheberrechtsvermerke, Markenzeichen, digitale Wasserzeichen und andere Rechtsvorbehalte im abgerufenen Inhalt nicht entfernen.

Der Nutzer ist nicht berechtigt, das E-Book – auch nicht auszugsweise – anderen Personen zugänglich zu machen, insbesondere es weiterzuleiten, zu verleihen oder zu vermieten.

Das entgeltliche oder unentgeltliche Einstellen des E-Books ins Internet oder in andere Netzwerke, der Weiterverkauf und/oder jede Art der Nutzung zu kommerziellen Zwecken sind nicht zulässig.

Das Anfertigen von Vervielfältigungen, das Ausdrucken oder Speichern auf anderen Wiedergabegeräten ist nur für den persönlichen Gebrauch gestattet. Dritten darf dadurch kein Zugang ermöglicht werden.

Die Übernahme des gesamten E-Books in eine eigene Print- und/oder Online-Publikation ist nicht gestattet. Die Inhalte des E-Books dürfen nur zu privaten Zwecken und nur auszugsweise kopiert werden.

Diese Bestimmungen gelten gegebenenfalls auch für zum E-Book gehörende Audiodateien.

Anmerkung:

Sofern der Printausgabe eine CD-ROM beigelegt ist, sind die Materialien/Arbeitsblätter, die sich darauf befinden, bereits Bestandteil dieses E-Books.

*Unseren Kindern
Joshua und Johanna gewidmet*

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

Vorwort

That this text presents to the reader more problems than it solves is perhaps merely a sign of the youth and vitality of a movement which I believe is destined to revolutionize the human relationship problems of society.

Kelley (1927, S. iv)

Über achtzig Jahre sind vergangen, seit Truman Lee Kelley mit diesem Satz das Vorwort zu einem der ersten Lehrbücher der Testtheorie (*Interpretations of educational measurements*) beendet hat. Über die letzten achtzig Jahre hinweg hat sich die Testtheorie rasant entwickelt und ist zu einem zentralen Methodenfach der Psychologie geworden, das unser Verständnis von den messtheoretischen Grundlagen der empirischen Psychologie entscheidend gefördert hat. Ob sich dadurch die „human relationship problems of society“ revolutioniert haben, sei dahingestellt, aber ohne ein fundiertes testtheoretisches Wissen kann man viele Tätigkeiten in der Psychologie nicht adäquat ausfüllen, nicht nur in der Forschung, sondern vor allem auch in der psychodiagnostischen Praxis.

Im Laufe ihrer Entwicklung hat sich die Testtheorie zunehmend ausdifferenziert, es wurden aber auch vielfältige Bezüge zwischen einzelnen Testtheorien deutlich, die es heute ermöglichen, verschiedene testtheoretische Ansätze in einem integrativen Rahmen darzustellen. Dies hilft, Bezüge zu erkennen und „Mythen“ zu entlarven. Das vorliegende Lehrbuch basiert auf einer solch integrativen Sicht und stellt die verschiedenen testtheoretischen Ansätze gegliedert nach der Art der Antwortvariablen, die untersucht werden sollen, dar. Klassische Testtheorie und Item-Response-Theorie werden somit integrativ im Rahmen einer gemeinsamen Darstellungsform vermittelt, die an Mellenberghs (1994b) Konzept einer *Generalisierten Linearen Item-Response-Theorie* orientiert ist. Dies steht im Gegensatz zu der weithin getrennten Vermittlung der Klassischen Testtheorie und der Item-Response-Theorie, die eher daran orientiert war, die Unterschiede zwischen diesen Ansätzen herauszuarbeiten, ohne die vielen Gemeinsamkeiten zu sehen. Durch die integrative Darstellung soll das Lernen erleichtert werden und gleichsam die Klassische Testtheorie als eine moderne Testtheorie vermittelt werden, die der Item-Response-Theorie in nichts nachsteht.

Das Buch führt in die Grundlagen der Testtheorie und Testkonstruktion ein. Es wurde Wert darauf gelegt, dieses zentrale Fach der Psychologie verständlich und anwendungsorientiert zu vermitteln, ohne die notwendige fachliche Tiefe zu vernachlässigen. Es richtet sich an Studierende der Psychologie, die ein fundiertes

Verständnis der modernen testtheoretischen Ansätze und ihrer zentralen anwendungspraktischen Implikationen erwerben wollen. Hierzu wird auf reale Anwendungsbeispiele der Psychologie zurückgegriffen, die ausführlich dargestellt werden. Auf einer begleitenden Internetseite (www.hogrefe.de/buecher/lehrbuecher/psychlehrbuchplus) werden alle Datensätze zur Verfügung gestellt, sodass Studierende alle Beispiele nachrechnen können. Um dies einfach zu ermöglichen, werden kommentierte Inputs und Outputs zur Verfügung gestellt, die sich vor allem auf das im Internet frei verfügbare Computerprogramm R beziehen (<http://www.r-project.org/>). Dies soll es Studierenden ermöglichen, nicht nur Wissen, sondern auch anwendungspraktische Kompetenzen zu erwerben. Auf der Internetseite sind auch Antworten zu den Lernkontrollfragen des Buches zu finden. Das Buch richtet sich aber nicht nur an Studierende, sondern an all jene, die sich ein grundlegendes Verständnis der testtheoretischen Grundlagen und ihrer anwendungspraktischen Bezüge erwerben wollen.

Das Buch ist ein einführendes Buch und daher zwangsläufig in der Breite der behandelten Themen beschränkt. Im Mittelpunkt steht die Vermittlung der grundlegenden Prinzipien anhand der einfachsten Modelle der Testtheorie, die zur Erfassung eindimensionaler Merkmale (Konstrukte) entwickelt wurden. Es gliedert sich in acht Kapitel.

Kapitel 1 führt in die Grundfragen der Testtheorie ein: Warum ist dieses Gebiet für die Psychologie von grundlegender Bedeutung? Worin unterscheidet sich Messen in der Psychologie von der Messung in anderen Naturwissenschaften wie z. B. der Physik? Warum braucht die Psychologie eine eigene Testtheorie und wieso sollte man sich mit dieser Thematik überhaupt beschäftigen? Was versteht man unter einem psychologischen Test und was unter einem psychologischen Konstrukt?

In *Kapitel 2* werden die zentralen Gütekriterien psychologischer Tests und die einzelnen Schritte der Testkonstruktion behandelt. Woran erkennt man einen guten psychometrischen Test? Wie geht man bei der Testkonstruktion vor? Worin unterscheiden sich verschiedene Prinzipien der Testkonstruktion? Wie kann man die Qualität psychologischer Tests sicherstellen?

Kapitel 3 stellt verschiedene Item- und Antwortformate vor. Was muss man bei der Formulierung einer Frage eines Fragebogens beachten? Wie viele Antwortalternativen sollte man in einem Multiple-Choice-Test vorgeben? Wie sollte man Antwortformate benennen?

Die *Kapitel 4 bis 6* behandeln Methoden zur Auswertung von Items eines Tests. Die Kapitel unterscheiden sich in der Art der behandelten Variablen, folgen aber demselben Aufbau. Zentrale Fragen sind: Wie lassen sich Merkmalsunterschiede

zwischen Personen beschreiben? Wie kann man Zusammenhänge zwischen verschiedenen Items erfassen? Wie lassen sich die Ausprägungen von Personenmerkmalen, die nicht direkt beobachtet werden können, schätzen? Was versteht man unter psychometrischen Modellen? Auf welchen Annahmen basieren diese und wie lässt sich deren Gültigkeit überprüfen? *Kapitel 4* widmet sich dichotomen Antwortvariablen und stellt insbesondere das Rasch-Modell und seine Erweiterungen vor. In *Kapitel 5* werden Modelle für Items mit geordneten Antwortkategorien behandelt, insbesondere das Partial-Credit-Modell und seine Spezialfälle sowie Erweiterungen. *Kapitel 6* hat schließlich metrische Antwortvariablen zum Gegenstand, es werden die Grundprinzipien und die verschiedenen Modelle der Klassischen Testtheorie behandelt.

Kapitel 7 überträgt die Grundideen testtheoretischer Modelle auf mehrdimensionale Modelle und stellt deren Grundideen exemplarisch dar. Worin unterscheiden sich eindimensionale von mehrdimensionalen Modellen? Was sind die Grundideen der konfirmatorischen und der exploratorischen Faktorenanalyse?

In *Kapitel 8* wird schließlich mit der Normierung ein zentrales Problem der psychologischen Diagnostik behandelt. Was bedeutet ein Testwert? Wie lässt er sich interpretieren? Wie kann man feststellen, ob sich zwei Personen unterscheiden oder eine Person sich verändert hat?

Eine zusätzliche *Formelsammlung*, in der die Formeln zum schnellen Nachschlagen zusammengestellt sind, findet sich auf der Internetseite www.hogrefe.de/buecher/lehrbuecher/psychlehrbuchplus.

Das Buch wurde so konzipiert, dass die Themen im Rahmen einer einsemestrigen Lehrveranstaltung behandelt werden können. Die Anzahl der Kapitel ist deutlich geringer als die Anzahl von Lehrveranstaltungen in einem Semester. Dies liegt daran, dass sich die geschlossene Darstellung im Rahmen weniger Kapitel als didaktisch sinnvoller erwies als die einzelnen Themen zergliedert über mehrere Kapitel hinweg zu behandeln. Die einzelnen Kapitel weisen jedoch eine sehr strukturierte Binnengliederung auf, sodass insbesondere auch die umfangreicheren Kapitel 3, 4 und 6 in sinnvolle Lehreinheiten aufgeteilt werden können, um sie dem spezifischen Lehrplan anzupassen.

Am Entstehen des Buches haben viele mitgewirkt, denen wir ganz herzlich danken möchten. Angela Coenders hat viele Diktate zu Papier gebracht, Grafiken und Tabellen erstellt und Teile des Buches und insbesondere das Literaturverzeichnis korrigiert. Dr. Georg Hosoya hat nicht nur alle Kapitel des Buches korrekturgelesen und alle Beispiele zur Kontrolle nachgerechnet, sondern auch die kommentierten R-Inputs und -Outputs für die Internetseite des Buches erstellt. Claudia

Crayen hat R-Inputs für Grafiken erstellt. Sie hat darüber hinaus zusammen mit Dr. Luna Beck, Fenne große Deters, Nele Feuerbach, Frances Hoferichter, Louisa Hohmann, Dr. Tobias Koch, Dr. Irina Kumschick, Tanja Kutscher, Mario Lawes, Jana Mahlke, Martin Schultze und Linda Wulkau Teile des Manuskripts korrekturegelesen, auf Verständlichkeit geprüft sowie wesentliche Verbesserungsvorschläge vorgelegt. Linda Wulkau hat zentrale Teile der Formelsammlung erstellt. Ramzi Fatfouta, Henriette Hunold und Tanja Kutscher haben umfangreiche Literaturrecherchen durchgeführt und diese aufgearbeitet. Die beiden Mitherausgeber der Reihe, Prof. Dr. Franz Petermann und Prof. Dr. Hans-Werner Bierhoff, haben ebenfalls den Text sehr sorgfältig gelesen und wichtige Verbesserungshinweise gegeben. Ihnen allen möchten wir herzlich danken. Wir freuen uns auf weitere Hinweise und Verbesserungsvorschläge!

Berlin, im Januar 2014

Michael Eid
Katharina Schmidt

Kapitel 1

Grundfragen der Testtheorie und Testkonstruktion

Inhaltsübersicht

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

Aufgaben der psychologischen Diagnostik

Eine der wesentlichen Aufgaben der *psychologischen Diagnostik* besteht darin, die Merkmalsausprägungen von Menschen zu erfassen. Dies ist notwendig, um wichtige diagnostische Fragen beantworten und Entscheidungen treffen zu können (Eid & Petermann, 2006). Wenn z. B. ein Kind in einer schulpсихologischen Beratungsstelle vorgestellt wird, da es den Unterricht wiederholt stört, geht es im Wesentlichen zunächst darum, die Ursachen dieses Verhaltens festzustellen. Eine mögliche Ursache könnte eine unterdurchschnittliche Intelligenz sein, die zu einer Überforderung führt, oder aber eine überdurchschnittliche Intelligenz, aufgrund derer das Kind unterfordert ist und sich langweilt. In beiden Fällen wäre die Störung des Unterrichts ein Ausdruck der Bewältigung der emotionalen Belastung, die sich durch die Unter- bzw. Überforderung ergibt. Um geeignete Maßnahmen wie z. B. Nachhilfe oder die Empfehlung, eine Klasse zu überspringen, auswählen zu können, ist es entscheidend, die Ursache zu kennen. Deswegen würde die Schulpsychologin die Intelligenz des Kindes feststellen (diagnostizieren) wollen.

Gütekriterien

Da das Ergebnis der Intelligenzdiagnostik weitreichende Folgen für die Zukunft des Kindes haben kann, wird die Psychologin alles daran geben, das beste Verfahren zur Erfassung der Intelligenz auszuwählen. Doch wodurch zeichnet sich das beste Verfahren aus und wie kann die Psychologin in ihrem Gutachten belegen, dass sie auf das beste Verfahren zurückgegriffen hat und daher bei der Auswahl des Verfahrens keinen Fehler begangen hat? Hierzu muss sie auf Qualitätskriterien zurückgreifen, die ihr erlauben, aus der Vielzahl der verfügbaren Instrumente dasjenige auszuwählen, das für ihre Fragestellung am geeignetsten ist und die *Gütekriterien* erfüllt, die sich in der Wissenschaft und Praxis bewährt haben und allgemein anerkannt sind (Kersting, 2008; Kersting, Häcker & Hornke, 2011). So sollte dieses Instrument auch wirklich die Intelligenz erfassen und nicht ein anderes Merkmal wie z. B. die Testängstlichkeit. Ob ein Instrument wirklich das erfasst, was es erfassen soll, und ob Schlussfolgerungen, die wir aus den Ergebnissen in Bezug auf das zu erfassende Merkmal ziehen, auch wirklich zutreffend und damit gültig sind, betrifft die Frage der *Validität*, dem zentralen Gütekriterium der Diagnostik (Messick, 1989). Schlüsse, die aufgrund der Messungen mit einem Erhebungsinstrument getroffen werden, sollten daher über eine hohe Validität verfügen.

Validität

Das Erhebungsinstrument sollte darüber hinaus ermöglichen, die Intelligenz des Kindes zuverlässig zu erfassen. Wenn das Kind in kur-

zem Abstand mit dem Instrument erneut untersucht wird oder wenn ein anderes Instrument, das ebenfalls die Intelligenz valide misst, zur Feststellung der Intelligenz des Kindes herangezogen wird, sollten sich die Ergebnisse, z. B. der Intelligenzquotient des Kindes, zwischen den Messungen oder den Tests nur unwesentlich unterscheiden. Unterschiede zwischen den verschiedenen Messergebnissen zeigen an, dass die einzelne Messung mit einem Messfehler behaftet ist. Messfehler kann man bei psychologischen Erhebungen meist nicht ausschließen. Ihr Einfluss sollte aber möglichst gering sein. Ein Messinstrument, bei dem dies der Fall ist, ist ein zuverlässiges, man sagt auch reliables Messinstrument. Das dazugehörige Gütekriterium heißt *Reliabilität*. Das Instrument sollte zudem so geartet sein, dass das Ergebnis der Untersuchung – die diagnostizierte Intelligenz – nicht von der Schulpsychologin abhängt, die die Untersuchung durchgeführt hat. Eine andere Schulpsychologin sollte zu demselben Ergebnis kommen. Dieses Gütekriterium, das die Unabhängigkeit der Testergebnisse von der Person der Untersuchungsleiterin bzw. des Untersuchungsleiters fordert, nennt man *Objektivität*. Damit sind nur die wesentlichen Gütekriterien, die sogenannten Hauptgütekriterien, genannt. Man kann sich schon vorstellen, dass es nicht einfach und sehr aufwendig ist, ein objektives, reliables und valides Instrument zu entwickeln. Die praktisch tätige Schulpsychologin wird daher in den seltensten Fällen ein solches Instrument selbst erstellen, sondern aufgrund ihrer Expertise und ihrer wissenschaftlichen Ausbildung, die sie im Rahmen ihres Studiums in der Konstruktion und Bewertung psychologischer Erhebungsverfahren erhalten hat, ein adäquates Instrument auswählen. Ohne dieses theoretische Wissen wäre sie nicht in der Lage, ihren Beruf angemessen auszuüben.

Reliabilität

Objektivität

1.1 Was sind psychologische Messungen?

Wie konstruiert man nun ein psychologisches Erhebungsverfahren und wie kann man sicherstellen, dass es das erfasst, was es erfassen soll? Das vorliegende Buch vermittelt die wesentlichen theoretischen Grundlagen, die hierfür notwendig sind, und zeigt ihre praktischen Implikationen auf. Die Grundfragen, die sich hierbei stellen, sind ganz ähnlich zu den Fragen, die sich bei der Erfassung von Merkmalen auch in anderen wissenschaftlichen Disziplinen ergeben. So sind wir mit der Messung physikalischer Eigenschaften sehr vertraut. Die Messungen der Länge und des Gewichts sind uns geläufig und wir greifen fast täglich auf sie zurück. Sie sind uns so vertraut, dass wir

Vergleich zur physikalischen Messung