

Joachim Krauth

# Testkonstruktion und Testtheorie

**BELTZ**

---

PsychologieVerlagsUnion

# Inhaltsverzeichnis

<b>Vorwort</b>	<b>13</b>
<b>A Konstruktion psychologischer Tests</b>	<b>15</b>
<b>1 Definition und Aufgaben psychologischer Tests</b>	<b>17</b>
1.1 Manifeste und latente Variablen . . . . .	17
1.2 Definition psychologischer Tests . . . . .	18
1.3 Entstehung des Testbegriffs . . . . .	18
1.4 Aufgaben psychologischer Tests . . . . .	19
1.5 Konstruktion psychologischer Tests . . . . .	20
<b>2 Itemkonstruktion</b>	<b>23</b>
2.1 Definition eines Items . . . . .	23
2.2 Auswahl von Items nach inhaltlichen Gesichtspunkten . . . . .	24
2.3 Lokale Unkorreliertheit von Items . . . . .	25
2.4 Skalenniveau von Items . . . . .	28
2.5 Formale Itemgestaltung . . . . .	35
2.5.1 Items mit zwei geordneten Antwortkategorien . . . . .	36
2.5.2 Items mit mehr als zwei geordneten Kategorien . . . . .	37
2.5.3 Items mit mindestens intervallskalierter Antwortskala . . . . .	38
2.6 Sprachliche Itemgestaltung . . . . .	40
2.7 Bewertung von Itembeantwortungen . . . . .	43
2.7.1 Antworttendenzen und Kontrollmethoden . . . . .	43
2.7.2 Fehlende Antworten . . . . .	48
2.7.3 Vom vorgegebenen Schema abweichende Antworten . . . . .	50
2.7.4 Antworten mit subjektivem Spielraum . . . . .	51
2.7.4.1 Erstellung eines Kategoriensystems . . . . .	52
2.7.4.2 Beurteilerschulung . . . . .	54
2.7.4.3 Bewertung der Beurteilerübereinstimmung . . . . .	54
2.7.5 Bewertung von Reaktionszeiten . . . . .	62
2.7.5.1 Zensierte Daten . . . . .	63
2.7.5.2 Reaktionszeiten für falsche Lösungen . . . . .	63
<b>3 Itemanalyse</b>	<b>67</b>
3.1 Eichstichprobe . . . . .	67
3.1.1 Zusammensetzung . . . . .	67

3.1.2	Umfang	68
3.2	Vorgabe der Items	69
3.2.1	Einmalige Vorgabe	69
3.2.2	Zweimalige Vorgabe	69
3.2.3	Zusätzliche Vorgabe eines Kriteriums	70
3.3	Items mit mindestens intervallskalierter Antwortskala	70
3.3.1	Schätzung der Itemschwierigkeiten	70
3.3.2	Schätzung der Itemvarianzen	72
3.3.3	Schätzung der Korrelationskoeffizienten	74
3.3.4	Schätzung der Itemreliabilitäten	75
3.3.4.1	Paralleltestreliabilität	75
3.3.4.2	Retestreliabilität	77
3.3.4.3	Itemreliabilität und Testreliabilität	78
3.3.5	Schätzung der Itemvaliditäten	78
3.3.5.1	Kriteriumsvalidität	78
3.3.5.2	Inhaltlich begründete Validität	82
3.3.5.3	Faktorenanalytisch begründete Validität	84
3.3.6	Schätzungen der Änderungssensitivität	90
3.4	Items mit zwei geordneten Kategorien	97
3.4.1	Schätzung der Itemschwierigkeiten	97
3.4.2	Schätzung der Itemvarianzen	100
3.4.3	Schätzung der Itemkorrelationen	100
3.4.4	Schätzung der Itemreliabilitäten	101
3.4.4.1	Paralleltestreliabilität	101
3.4.4.2	Retestreliabilität	103
3.4.5	Schätzung der Itemvaliditäten	103
3.4.5.1	Kriteriumsvalidität	103
3.4.5.2	Inhaltlich begründete Validität	105
3.4.5.3	Faktorenanalytisch begründete Validität	106
3.4.6	Schätzungen der Änderungssensitivität	108
3.4.7	Schätzung der Item- und Personenparameter aus dem Rasch-Modell	111
3.5	Items mit mehr als zwei geordneten Kategorien	117
3.5.1	Schätzung der Itemschwierigkeiten	118
3.5.2	Schätzung der Itemvarianzen	119
3.5.3	Schätzung der Itemkorrelationen	120
3.5.4	Schätzung der Itemreliabilitäten	120
3.5.4.1	Paralleltestreliabilität	120
3.5.4.2	Retestreliabilität	123
3.5.5	Schätzung der Itemvaliditäten	124
3.5.5.1	Kriteriumsvalidität	124
3.5.5.2	Inhaltlich begründete Validität	126
3.5.5.3	Faktorenanalytisch begründete Validität	128
3.5.6	Schätzungen der Änderungssensitivität	131
3.5.7	Schätzung der Item- und Personenparameter aus dem Rasch-Modell	135

<b>4</b>	<b>Konstruktion optimaler Tests</b>	<b>141</b>
4.1	Items mit mindestens intervallskalierter Antwortskala . . . . .	141
4.1.1	Reliabilitätsoptimale Tests . . . . .	141
4.1.1.1	Schätzung der Reliabilität bei bekannter Testzusammen- setzung . . . . .	141
4.1.1.2	Reliabilitätsoptimierung durch Itemauswahl . . . . .	145
4.1.1.3	Reliabilitätsoptimierung durch Gewichtung . . . . .	147
4.1.1.4	Testverkürzung . . . . .	148
4.1.2	Validitätsoptimale Tests . . . . .	148
4.1.2.1	Schätzung der Validität bei bekannter Testzusammen- setzung . . . . .	148
4.1.2.2	Validitätsoptimierung durch Itemauswahl . . . . .	152
4.1.2.3	Validitätsoptimierung durch Gewichtung . . . . .	153
4.1.2.4	Testverkürzung . . . . .	153
4.2	Items mit zwei geordneten Kategorien . . . . .	154
4.2.1	Reliabilitätsoptimale Tests . . . . .	154
4.2.1.1	Schätzung der Reliabilität bei bekannter Testzusammen- setzung . . . . .	154
4.2.1.2	Reliabilitätsoptimierung durch Itemauswahl . . . . .	158
4.2.1.3	Reliabilitätsoptimierung durch Gewichtung . . . . .	159
4.2.1.4	Testverkürzung . . . . .	160
4.2.2	Validitätsoptimale Tests . . . . .	160
4.2.2.1	Schätzung der Validität bei bekannter Testzusammen- setzung . . . . .	160
4.2.2.2	Validitätsoptimierung durch Itemauswahl . . . . .	163
4.2.2.3	Validitätsoptimierung durch Gewichtung . . . . .	164
4.2.2.4	Testverkürzung . . . . .	166
4.3	Items mit mehr als zwei geordneten Kategorien . . . . .	166
4.3.1	Reliabilitätsoptimale Tests . . . . .	166
4.3.1.1	Schätzung der Reliabilität bei bekannter Testzusammen- setzung . . . . .	166
4.3.1.2	Reliabilitätsoptimierung durch Itemauswahl . . . . .	171
4.3.1.3	Reliabilitätsoptimierung durch Gewichtung . . . . .	172
4.3.1.4	Testverkürzung . . . . .	174
4.3.2	Validitätsoptimale Tests . . . . .	174
4.3.2.1	Schätzung der Validität bei bekannter Testzusammen- setzung . . . . .	174
4.3.2.2	Validitätsoptimierung durch Itemauswahl . . . . .	178
4.3.2.3	Validitätsoptimierung durch Gewichtung . . . . .	178
4.3.2.4	Testverkürzung . . . . .	179
<b>5</b>	<b>Testkonstruktion nach Auswahl der Items</b>	<b>181</b>
5.1	Itemauswahl . . . . .	181
5.2	Itemanordnung . . . . .	182
5.3	Verteilungsanalyse . . . . .	183
5.4	Konstruktion eines Paralleltests zu einem vorhandenen Test . . . . .	189

5.5	Korrekturen des Testwerts . . . . .	194
5.6	Normwerte . . . . .	196
5.6.1	Lineare Transformationen . . . . .	196
5.6.2	Flächen- und Normalisierungstransformationen . . . . .	202
5.7	Konfidenzintervalle für den wahren Testwert einer Person . . . . .	208
5.8	Testverkürzung . . . . .	210
<b>6</b>	<b>Erstellung eines Testmanuals</b>	<b>215</b>
<b>B</b>	<b>Theoretische Grundlagen psychologischer Tests</b>	<b>219</b>
<b>1</b>	<b>Statistische Verteilungsparameter und ihre Eigenschaften</b>	<b>221</b>
<b>2</b>	<b>Klassische Testtheorie</b>	<b>237</b>
2.1	Modell . . . . .	237
2.2	Gütekriterien . . . . .	249
2.3	Itemparameter . . . . .	265
2.4	Testbatterien . . . . .	278
<b>3</b>	<b>Berücksichtigung des Skalenniveaus in der Klassischen Testtheorie</b>	<b>289</b>
3.1	Spezialisierung auf Items mit zwei geordneten Kategorien . . . . .	292
3.2	Spezialisierung auf Items mit mehr als zwei geordneten Kategorien . . . . .	295
<b>4</b>	<b>Änderungssensitivität</b>	<b>297</b>
<b>5</b>	<b>Rate- und Auslassungskorrekturen</b>	<b>303</b>
5.1	Korrektur der Itemschwierigkeit . . . . .	304
5.2	Korrektur des Rohwerts . . . . .	307
<b>6</b>	<b>Konfidenzintervalle</b>	<b>311</b>
6.1	Konfidenzintervalle für Korrelationskoeffizienten . . . . .	312
6.1.1	Parametrische Konfidenzintervalle . . . . .	312
6.1.2	Nichtparametrische Konfidenzintervalle . . . . .	314
6.2	Konfidenzintervalle für den wahren Testwert einer Person . . . . .	318
<b>7</b>	<b>Rasch-Modell für Items mit zwei geordneten Kategorien</b>	<b>321</b>
<b>C</b>	<b>Programm zur Testkonstruktion</b>	<b>335</b>
<b>1</b>	<b>Dateneingabe</b>	<b>339</b>
1.1	Datei anlegen . . . . .	339
1.1.1	Personendatei . . . . .	340
1.1.2	Itemdatei zum Skalenniveau $S_I = 2$ . . . . .	341
1.1.3	Itemdatei zum Skalenniveau $S_I = 3$ . . . . .	341
1.1.4	Kriteriendatei zum Skalenniveau $S_C = 2$ . . . . .	341
1.1.5	Kriteriendatei zum Skalenniveau $S_C = 3$ . . . . .	342

1.1.6	Menu Dateneingabe . . . . .	342
1.2	Datei ändern / anschauen . . . . .	342
1.2.1	Personendatei ändern . . . . .	342
1.2.1.1	Gesamte Personendatei anschauen . . . . .	343
1.2.1.2	Gesamte Personendatei ändern . . . . .	343
1.2.1.2.1	Alle Personen drucken . . . . .	343
1.2.1.2.2	1. Person drucken . . . . .	343
1.2.1.2.3	1. Person ändern . . . . .	343
1.2.1.2.3.1	1. Person löschen . . . . .	343
1.2.1.2.3.2	Teilstichprobe der 1. Person ändern . . . . .	343
1.2.1.2.3.3	Alle Items der 1. Person ändern . . . . .	344
1.2.1.2.3.4	Ein Item der 1. Person ändern . . . . .	344
1.2.1.2.3.5	Alle Kriterien der 1. Person ändern . . . . .	344
1.2.1.2.3.6	Ein Kriterium der 1. Person ändern . . . . .	344
1.2.1.2.3.7	Menu Gesamte Personendatei ändern . . . . .	344
1.2.1.2.4	Weiter zur 2. Person . . . . .	344
1.2.1.2.5	Menu Personendatei ändern . . . . .	344
1.2.1.3	Eine weitere Person eingeben . . . . .	344
1.2.1.4	Eine Person löschen . . . . .	344
1.2.1.5	Daten einer Person ändern . . . . .	344
1.2.1.5.1	x. Person löschen . . . . .	345
1.2.1.5.2	Teilstichprobe der x. Person ändern . . . . .	345
1.2.1.5.3	Alle Items der x. Person ändern . . . . .	345
1.2.1.5.4	Ein Item der x. Person ändern . . . . .	345
1.2.1.5.5	Alle Kriterien der x. Person ändern . . . . .	345
1.2.1.5.6	Ein Kriterium der x. Person ändern . . . . .	345
1.2.1.5.7	Menu Personendatei ändern . . . . .	345
1.2.1.6	Zweitvorgabe wechseln . . . . .	345
1.2.1.7	Neue Zuordnung von Personen zur Behandlungs- oder Kontrollgruppe . . . . .	346
1.2.1.8	Anzahl der Teilstichproben ändern . . . . .	346
1.2.1.9	Neue Zuordnung der Personen zu den Teilstichproben . . . . .	346
1.2.1.10	Menu Datei ändern / anschauen . . . . .	346
1.2.2	Itemdatei zum Skalenniveau $S_I = 2$ . . . . .	346
1.2.2.1	Gesamte Itemdatei betrachten . . . . .	346
1.2.2.2	Geordnete Item-Kategorie hinzunehmen . . . . .	347
1.2.2.3	Geordnete Item-Kategorien löschen . . . . .	347
1.2.2.4	Benennungen von Item-Kategorien ändern . . . . .	347
1.2.2.5	Menu Datei ändern / anschauen . . . . .	347
1.2.3	Itemdatei zum Skalenniveau $S_I = 3$ . . . . .	347
1.2.3.1	Gesamte Itemdatei ändern . . . . .	347
1.2.3.2	Alle Kategorien eines Items ändern . . . . .	347
1.2.3.3	Nicht geordnete Item-Kategorie hinzunehmen . . . . .	348
1.2.3.4	Nicht geordnete Item-Kategorie löschen . . . . .	348
1.2.3.5	Nicht geordnete Item-Kategorie ändern . . . . .	348
1.2.3.6	Itemdatei anschauen . . . . .	348

1.2.3.7	Itemdatei drucken . . . . .	348
1.2.3.8	Menu Datei ändern / anschauen . . . . .	348
1.2.4	Kriteriendatei zum Skalenniveau $S_C = 2$ . . . . .	348
1.2.4.1	Kriteriendatei anschauen . . . . .	349
1.2.4.2	Geordnete Kategorie hinzunehmen . . . . .	349
1.2.4.3	Geordnete Kategorien löschen . . . . .	349
1.2.4.4	Benennungen von Kategorien ändern . . . . .	349
1.2.4.5	Menu Datei ändern / anschauen . . . . .	349
1.2.5	Kriteriendatei zum Skalenniveau $S_C = 3$ . . . . .	349
1.2.5.1	Gesamte Kriteriendatei ändern . . . . .	350
1.2.5.2	Alle Kategorien eines Kriteriums ändern . . . . .	350
1.2.5.3	Nicht geordnete Kriterien-Kategorie hinzunehmen . . . . .	350
1.2.5.4	Nicht geordnete Kriterien-Kategorie löschen . . . . .	350
1.2.5.5	Nicht geordnete Kriterien-Kategorie ändern . . . . .	350
1.2.5.6	Kriteriendatei anschauen . . . . .	350
1.2.5.7	Kriteriendatei drucken . . . . .	350
1.2.5.8	Menu Datei ändern / anschauen . . . . .	350
1.2.6	Menu Dateneingabe . . . . .	351
1.3	Datei löschen . . . . .	351
1.4	Datei konvertieren . . . . .	351
1.4.1	Konvertierung einer Personendatei . . . . .	351
1.4.2	Konvertierung einer Itemdatei zum Skalenniveau $S_I = 2$ . . . . .	352
1.4.3	Konvertierung einer Itemdatei zum Skalenniveau $S_I = 3$ . . . . .	352
1.4.4	Konvertierung einer Kriteriendatei zum Skalenniveau $S_C = 2$ . . . . .	352
1.4.5	Konvertierung einer Kriteriendatei zum Skalenniveau $S_C = 3$ . . . . .	352
1.5	Vorhandene Dateien . . . . .	352
1.6	Hauptmenu . . . . .	353
<b>2</b>	<b>Auswahl einer Datei</b>	<b>355</b>
<b>3</b>	<b>Itemanalyse</b>	<b>357</b>
3.1	Schätzungen der Itemschwierigkeiten, Itemvarianzen und Itemkorrelationen . . . . .	358
3.2	Schätzungen der Paralleltestreliabilitäten von Items . . . . .	358
3.3	Schätzungen der Retestreliabilitäten von Items . . . . .	359
3.4	Schätzungen der Kriteriumsvaliditäten von Items . . . . .	359
3.5	Schätzungen der inhaltlich begründeten Itemvaliditäten . . . . .	359
3.6	Hauptkomponentenanalyse der Items mit VARIMAX-Rotation . . . . .	360
3.7	Schätzungen der Änderungssensitivitäten . . . . .	360
3.8	Rasch-Analyse . . . . .	360
3.9	Hauptmenu . . . . .	361
<b>4</b>	<b>Gütekriterien, Verteilungsanalyse und Normen</b>	<b>363</b>
4.1	Paralleltestreliabilität und Testhalbierungsreliabilität . . . . .	363
4.2	Retestreliabilität . . . . .	364
4.3	$\alpha$ -Koeffizient, $\lambda_2$ -Koeffizient und innere Konsistenz . . . . .	365
4.4	Kriteriumsvalidität . . . . .	366

4.5	Inhaltliche Validität . . . . .	366
4.6	Faktorenanalytische Validität . . . . .	367
4.7	Hauptkomponentenanalyse für Summenscore und Kriterien . . . .	367
4.8	Verteilungsanalyse und Normen . . . . .	368
4.9	Hauptmenu . . . . .	369
<b>5</b>	<b>Optimale Gewichte</b>	<b>371</b>
5.1	Optimale Gewichte zur Paralleltestreliabilität . . . . .	371
5.2	Optimale Gewichte zur Retestreliabilität . . . . .	372
5.3	Optimale Gewichte zum $\alpha$ -Koeffizienten . . . . .	372
5.4	Optimale Gewichte zur Kriteriumsvalidität . . . . .	372
5.5	Optimale Gewichte zur inhaltlichen Validität . . . . .	373
5.6	Optimale Gewichte zur faktorenanalytischen Validität . . . . .	373
5.7	Hauptkomponentenanalyse für den gewichteten Summenscore und die Kriterien . . . . .	374
5.8	Verteilungsanalyse und Normen . . . . .	375
5.9	Hauptmenu . . . . .	376
	<b>Literaturverzeichnis</b>	<b>377</b>
	<b>Tabelle der <math>P</math>-Werte und <math>Z</math>-Werte der Standardnormalverteilung</b>	<b>383</b>
	<b>Autorenregister</b>	<b>391</b>
	<b>Sachregister</b>	<b>393</b>