

Grundlagen der Testtheorie

WS 2020/21

12. Normierung, Standards für
psychologisches Testen

08.02.2021

Prof. Dr. Eunike Wetzel

Semesterplan

Sitzung	Termin	Thema
1	02.11.	Grundlagen & Gütekriterien
2	09.11.	Schritte der Testkonstruktion: Übersicht Konstruktdefinition & Itemgenerierung
3	16.11.	Erstellung eines Testentwurfs
4	23.11.	Klassische Testtheorie
5	07.12.	Item Response Theorie
6	14.12.	Exploratorische Faktorenanalyse
7	04.01.	Itemanalyse 1
8	11.01.	Itemanalyse 2, Itemselektion & Testrevision
9	18.01.	Objektivität
10	25.01.	Reliabilität
11	01.02.	Validität
12	08.02.	Normierung, Standards für psychologisches Testen

Normierung

Normierung

1. Testwertermittlung
2. Normorientierte Testwertinterpretation
3. Kriteriumsorientierte Testwertinterpretation
4. Testnormierung

1. Testwertermittlung

- Der Testwert einer Person wird nach bestimmten Regeln aus ihren Itemantworten gebildet
- Vorgehen:
 - Kodierung jeder einzelnen Itemantwort nach einer bestimmten Regel (z. B. falsch = 0, richtig = 1)
 - Ermittlung des Testwerts über alle Items hinweg (z. B. Summenscore)
- Die Regeln können unterschiedlich komplex sein

1. Testwertermittlung

Beispiel: Testwertermittlung beim Konzentrationstest d2
(Brickenkamp, 2002)

	GZ	F ₁	F ₂	KL
1	d d p d d d p p d p d d d d p d p d d d p p d d d d d d p d p d d p p d d d d p p d p d d p	" "	" "	" "
2	p d p p d d d d p d d d p d d p d p d p d d d p d p d p d p d d d d p d p d d	" "	" "	" "
3	d d d d p p d p d p p p d d p d p d p d d p d p d d p d p d p d d p d p d d d d p d	" "	" "	" "
4	d d p d d d p p d p d d d d p d p d d d p p d d d d d d p d p d d p p d d d d p p d p d d p	" "	" "	" "

GZ = Gesamtzahl der bearbeiteten Zeichen
(Bearbeitungsmenge)

F1 = Anzahl Auslassungsfehler

F2 = Anzahl Verwechslungsfehler

KL = Anzahl richtig durchgestrichene Zeichen – F2

1. Testwertermittlung

- Der Testwert ist erst einmal ein Rohwert, der zwar das Antwortverhalten der Person widerspiegelt, aber alleine nicht sehr aussagekräftig ist
- Daher werden Testwerte anhand eines Vergleichsmaßstabs interpretiert
- 2 Möglichkeiten:
 - Normorientierte Testwertinterpretation: Bezugsgruppe
 - Kriteriumsorientierte Testwertinterpretation: psychologisch-inhaltliche Kriterien
- Beide Arten der Testwertinterpretation können auch integriert werden und sich ergänzen

Normierung

1. Testwertermittlung
- 2. Normorientierte Testwertinterpretation**
3. Kriteriumsorientierte Testwertinterpretation
4. Testnormierung

2. Normorientierte Testwertinterpretation

- Ausprägung der Person auf dem Konstrukt wird relativ zu einer relevanten Bezugsgruppe bestimmt
- Transformation des individuellen Testwerts in einen Normwert
- Arten der Transformation:
 - Nicht-linear: Prozentrangnormen
 - Linear: Standardisierte Normwerte (z. B. z-Werte, IQ-Werte, T-Werte)

2. Normorientierte Testwertinterpretation

Prozentrangnormen

- Der Prozentrang (PR) gibt an, wie viel Prozent der Normierungsstichprobe einen Testwert erzielten, der kleiner oder gleich dem Testwert der Person ist
- Die Transformation in PR ist eine nicht-lineare Transformation, da sie durch die Häufigkeitsverteilung der Normierungsstichprobe gebildet wird

2. Normorientierte Testwertinterpretation

Prozentrangnormen: Bestimmung

1. Testwerte in aufsteigende Rangordnung bringen
2. Häufigkeiten der einzelnen Testwerte bestimmen
3. Kumulierte Häufigkeiten $\text{freq}_{\text{cum}}(X_v)$ bis einschließlich des jeweiligen Testwertes berechnen
4. Prozentränge berechnen:

$$\text{PR}_v = 100 \cdot \frac{\text{freq}_{\text{cum}}(X_v)}{N}$$

Bsp. NARQ

- Summenscore von Admiration, $N = 430$

```
> summary(adm.sort)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  9.00  23.00  29.00  28.59  34.00  54.00
```

Bsp. NARQ

```
> cbind(adm.haeufig, adm.kum, adm.pr)
```

	adm.haeufig	adm.kum	adm.pr
9	1	1	0.23
10	2	3	0.70
11	2	5	1.16
12	2	7	1.63
13	2	9	2.09
14	6	15	3.49
15	4	19	4.42
16	7	26	6.05
17	4	30	6.98
18	11	41	9.53
19	15	56	13.02
20	14	70	16.28
21	14	84	19.53
22	16	100	23.26
23	12	112	26.05
24	21	133	30.93
25	13	146	33.95
26	17	163	37.91
27	19	182	42.33
28	28	210	48.84
29	23	233	54.19
30	26	259	60.23
31	25	284	66.05
32	17	301	70.00
33	17	318	73.95
34	14	332	77.21
35	16	348	80.93

36	17	365	84.88
37	11	376	87.44
38	14	390	90.70
39	9	399	92.79
40	7	406	94.42
41	2	408	94.88
42	8	416	96.74
43	2	418	97.21
44	5	423	98.37
45	1	424	98.60
46	2	426	99.07
49	1	427	99.30
53	1	428	99.53
54	2	430	100.00

$$PR_v = 100 \cdot \frac{\text{freq}_{\text{cum}}(X_v)}{N} =$$

$$100 \cdot \frac{182}{430} = 42.33$$

2. Normorientierte Testwertinterpretation

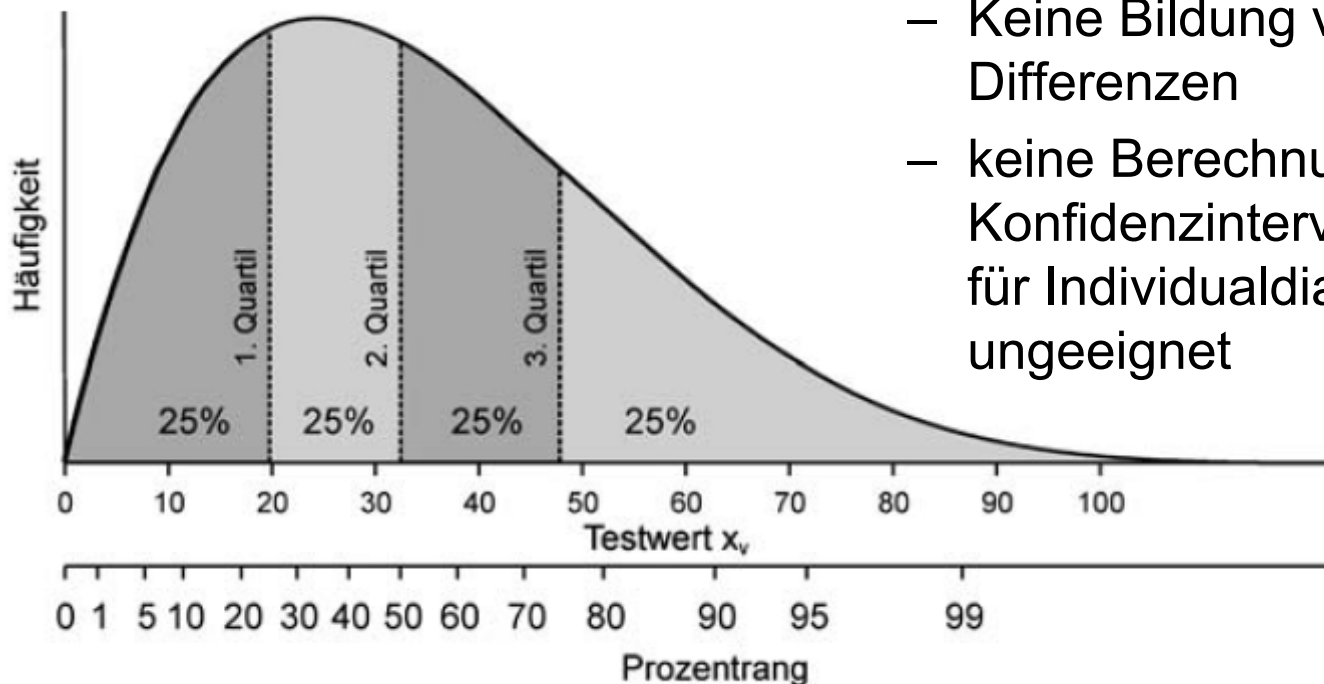
Beispiel:
Prozentrang-
normen im
d2

Altersgruppe 20;0 – 39;11 Jahre	PR	SW	GZ	GZ-F	KL
Eichstichprobe (N = 731)		–70–	254 – 260	239 – 245	66 – 69
		71	261 – 268	246 – 253	70 – 73
		72	269 – 276	254 – 260	74 – 76
		73	277 – 283	261 – 267	77 – 80
	<1	74	284 – 291	268 – 275	81 – 84
	1	75	292 – 298	276 – 282	85 – 88
	1	76	299 – 306	283 – 289	89 – 92
	1	77	307 – 314	290 – 297	93 – 96
	1	78	315 – 321	298 – 304	97 – 100
	2	79	322 – 329	305 – 311	101 – 103
	2	–80–	330 – 336	312 – 319	104 – 107
	3	81	337 – 344	320 – 326	108 – 111
	3	82	345 – 351	327 – 333	112 – 115
	4	83	352 – 359	334 – 341	116 – 119
	5	84	360 – 367	342 – 348	120 – 123
	7	85	368 – 374	349 – 355	124 – 126
	8	86	375 – 382	356 – 363	127 – 130
	10	87	383 – 389	364 – 370	131 – 134
	12	88	390 – 397	371 – 377	135 – 138
	13	89	398 – 405	378 – 385	139 – 142
	16	–90–	406 – 412	386 – 392	143 – 146
	18	91	413 – 420	393 – 400	147 – 150
	21	92	421 – 427	401 – 407	151 – 153
	24	93	428 – 435	408 – 414	154 – 157
	25	—	—	—	—

2. Normorientierte Testwertinterpretation

Prozentrangnormen

- Vorteil von PR: Machen keine Verteilungsvoraussetzungen und eignen sich daher zur Beschreibung schief verteilter Testwerte
- Nachteil von PR: Kein Intervallskalenniveau
 - Keine Bildung von PR-Differenzen
 - keine Berechnung von Konfidenzintervallen möglich → für Individualdiagnostik ungeeignet



2. Normorientierte Testwertinterpretation

Standardisierte Normwerte

- Lineare Transformation: Bestimmung der Differenz des Testwerts einer Person vom Mittelwert der Verteilung der Normierungsstichprobe
- Zusätzlich Relativierung der Differenz an der SD der Testwerte → Vergleichbarkeit zwischen Tests mit verschiedenen Streuungen und Skalenbereichen

$$z_v = \frac{x_v - \bar{x}}{SD(x)}$$

x_v : Testwert von Person v

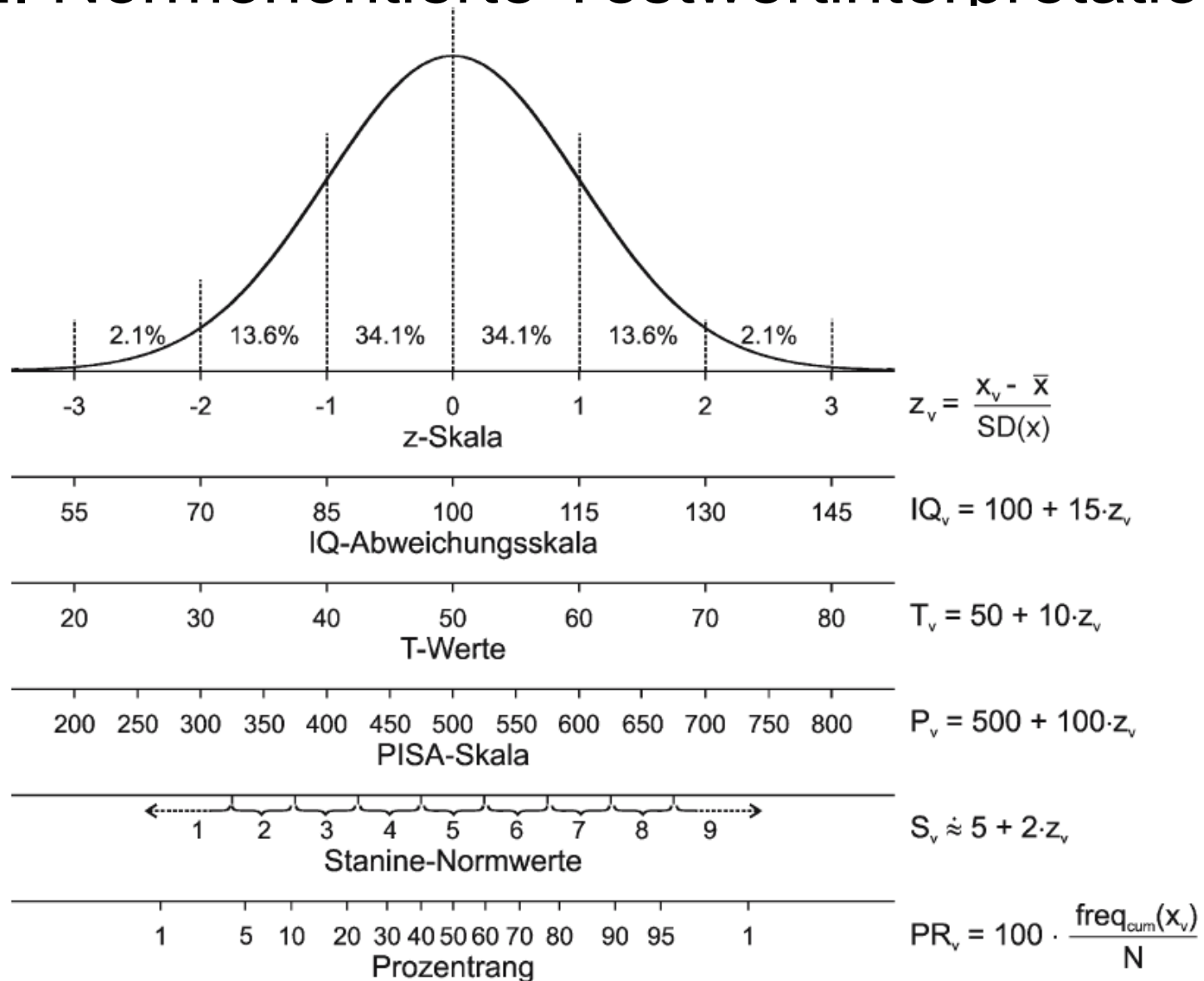
- Voraussetzung: Intervallskalenniveau der Testwertvariablen

2. Normorientierte Testwertinterpretation

Standardisierte Normwerte

- Ist die Testwertvariable normalverteilt, werden die transformierten z-Werte als Standardnorm bezeichnet
- Häufig werden z-Werte weiteren Lineartransformationen unterzogen, um Normwerte mit positivem Vorzeichen und ganzzahligen Abstufungen zu erhalten

2. Normorientierte Testwertinterpretation



Bsp. NARQ

Alexander hat einen Summenscore von 38 auf Admiration. In der Normierungsstichprobe waren $M = 30$ und $SD = 8$.

$$z_v = \frac{x_v - \bar{x}}{SD(x)} = \frac{38 - 30}{8} = 1$$

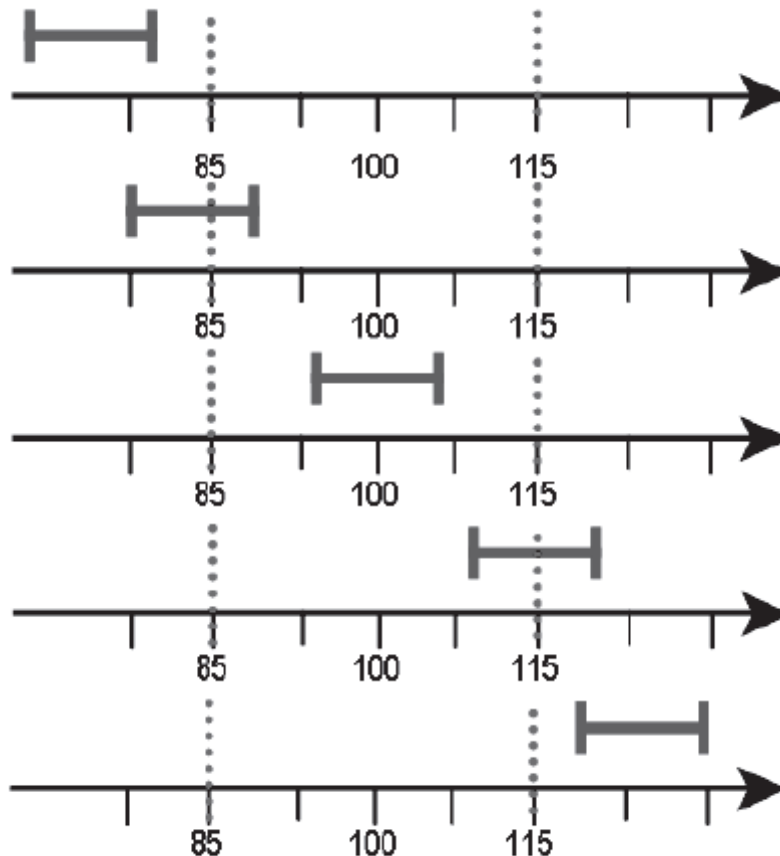
$$T_v = 50 + 10 \cdot z_v = 60$$

$$S_v \approx 5 + 2 \cdot z_v = 7$$

2. Normorientierte Testwertinterpretation

Klassifikation der standardisierten Normwerte mit KI

Beispiel: IQ-Werte



Unterdurchschnittlich

Unterdurchschnittlich bis
durchschnittlich

Durchschnittlich

Durchschnittlich bis
überdurchschnittlich

Überdurchschnittlich

2. Normorientierte Testwertinterpretation

Bsp.: Eine 20-jährige Abiturientin hat auf der Skala Offenheit für Ideen (O5) einen Testwert (Summenscore) von 19. Wie ist ihre Ausprägung einzuordnen? (KI: +/- 5)

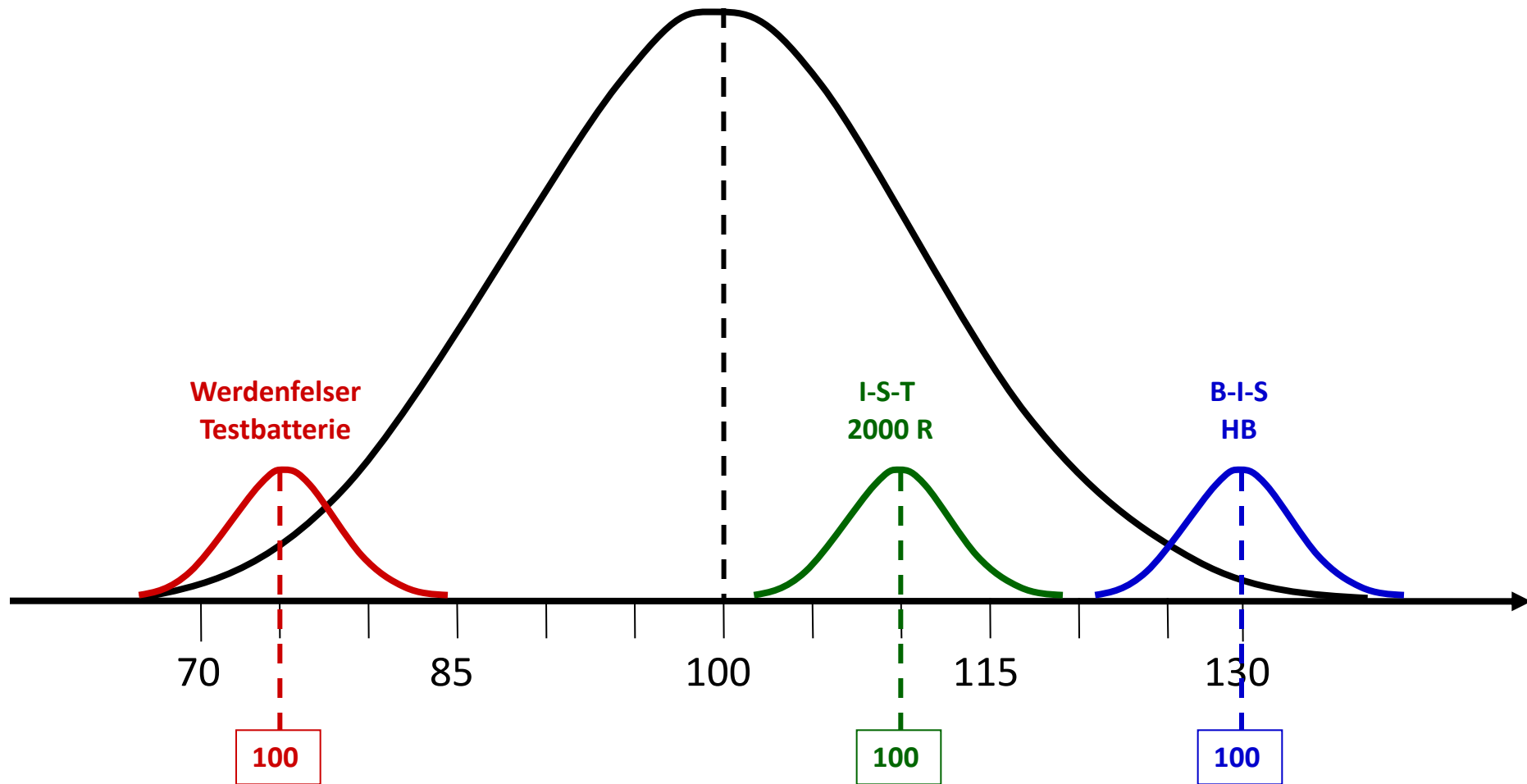
NEO-PI-R-Form S Normwerte

Tabelle E: Frauen im Alter von 16–29 Jahren mit Hochschul- bzw. Fachhochschulreife (N=3 075)

T	O	O1	O2	O3	O4	O5	O6	T	ST	PR
≥80	174–192				30–32		31–32	≥80	9	≥99.9
79	173					32		79	9	99.8
78								78	9	99.7
77	172							77	9	99.7
76	171	32			29			76	9	99.5
75	169–170		32	32		31	30	75	9	99.4
74	167–168							74	9	99.2
73	166				28			73	9	98.9
72	165						29	72	9	98.6
71	164	31			27	30		71	9	98.2

T	O	O1	O2	O3	O4	O5	O6	T	ST	PR
49	129	22			18	19		49	5	46.0
48	127–128		23	24			21	48	5	42.1
47	125–126	21				18		47	4	38.2
46	123–124				17			46	4	34.5
45	122	20	22	23		17	20	45	4	30.9
44	120–121				16			44	4	27.5
43	118–119	19	21	22		16		43	4	24.2
42	116–117	18	20		15		19	42	3	21.2
41	115					15		41	3	18.4
40	113–114	17	19	21				40	3	15.9

2. Normorientierte Testwertinterpretation



Hypothetische Verteilung der Intelligenz in der Population in IQ-Werten

2. Normorientierte Testwertinterpretation

Normorientierte Testwertinterpretation in der IRT

- Personenparameter lassen sich normorientiert interpretieren, wenn sie so geschätzt werden, dass ihre Summe 0 ist
- Vorzeichen und Betrag von θ geben an, wie weit sich eine Person über/unter dem Durchschnitt befindet

Normierung

1. Testwertermittlung
2. Normorientierte Testwertinterpretation
3. **Kriteriumsorientierte Testwertinterpretation**
4. Testnormierung

3. Kriteriumsorientierte Testwertinterpretation

- Interpretation des Testwerts in Bezug auf ein spezifisches inhaltliches Kriterium
- Bsp.: Mit einem Fragebogen soll festgestellt werden, ob bei einer Person die Symptome einer Major Depression vorliegen und eine genauere Diagnostik und evtl. Therapie angezeigt sind
 - Kriterium: Vorliegen von Symptomen einer Major Depression
 - Höhe des Testwerts soll angeben, ob bei der Testperson das Kriterium erfüllt ist
- In der Regel werden vorab Schwellenwerte definiert, ab denen ein Kriterium als erfüllt angenommen wird (z. B. ab einem Testwert von 20 liegen die Symptome der Major Depression vor)

3. Kriteriumsorientierte Testwertinterpretation

Bestimmung von Schwellenwerten

- Bezug des Testwertes auf ein externes Kriterium zur Unterscheidung von zwei Gruppen (Kriterium erfüllt/nicht erfüllt) → ROC-Analysen
- Bezug des Testwertes auf Aufgabeninhalte (z. B. Lernziele)

3. Kriteriumsorientierte Testwertinterpretation

Bsp.: Bildungsstandards Mathematik

Die Kompetenz „Mathematische Darstellungen verwenden“ (K4)

Diese Kompetenz umfasst das Auswählen geeigneter Darstellungsformen, das Erzeugen mathematischer Darstellungen und das Umgehen mit gegebenen Darstellungen. Hierzu zählen Diagramme, Graphen und Tabellen ebenso wie Formeln. Das Spektrum reicht von Standarddarstellungen – wie Wertetabellen – bis zu eigenen Darstellungen, die dem Strukturieren und Dokumentieren individueller Überlegungen dienen und die Argumentation und das Problemlösen unterstützen.

Anforderungsbereich I: Die Schülerinnen und Schüler können

- Standarddarstellungen von mathematischen Objekten und Situationen anfertigen und nutzen

Anforderungsbereich II: Die Schülerinnen und Schüler können

- gegebene Darstellungen verständig interpretieren oder verändern
- zwischen verschiedenen Darstellungen wechseln

Anforderungsbereich III: Die Schülerinnen und Schüler können

- mit unvertrauten Darstellungen und Darstellungsformen sachgerecht und verständig umgehen
- eigene Darstellungen problemadäquat entwickeln
- verschiedene Darstellungen und Darstellungsformen zweckgerichtet beurteilen

Normierung

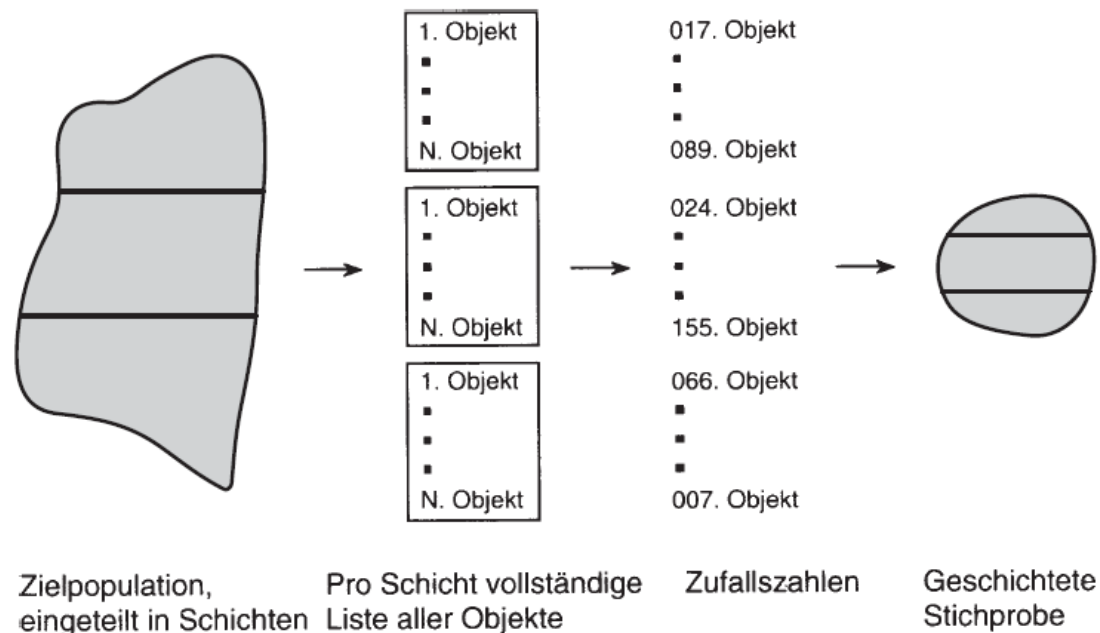
1. Testwertermittlung
2. Normorientierte Testwertinterpretation
3. Kriteriumsorientierte Testwertinterpretation
4. **Testnormierung**

4. Testnormierung

- Auch „Testeichung“
- Ziel: Gewinnung von Normwerten für die normorientierte Testwertinterpretation
- Durchführung des Tests an einer Normierungsstichprobe, die repräsentativ für die Zielpopulation ist
 - Globale Repräsentativität: Zufallsstichprobe
 - Spezifische Repräsentativität: Repräsentativ hinsichtlich der Faktoren, die mit dem untersuchten Konstrukt zusammenhängen (z. B. Geschlecht, Alter, Bildung)
- Normen sollten alle 8 Jahre überprüft werden (DIN 33430)

4. Testnormierung

- Liegen Kenntnisse über relevante Hintergrundfaktoren vor, kann dies bei der Stichprobenziehung berücksichtigt werden
- **Geschichtete Stichprobe:**



■ **Abb. 7.8.** Ziehung einer geschichteten Stichprobe

4. Testnormierung

- **Quotenstichprobe:** Vorgabe von Quoten, aber keine Zufallsauswahl
- Geschichtete Stichprobe und Quotenstichprobe erreichen, dass die prozentuale Verteilung der Ausprägungen auf relevanten Hintergrundfaktoren in der Normierungsstichprobe mit der Verteilung in der Population identisch ist
- Werden die Daten für die Normierungsstichprobe auf Basis einer **ad-hoc Stichprobe** gesammelt, ist die Stichprobe nicht repräsentativ, es kann aber nachträglich eine Quotenstichprobe gebildet werden
- Nach der Datenerhebung werden zunächst die Verteilungseigenschaften geprüft bevor Normwerte berechnet werden und ggf. eine Differenzierung der Normen vorgenommen wird

4. Testnormierung

Normdifferenzierung

- Differenzierung der Normen nach Hintergrundmerkmalen, die mit dem zu untersuchenden Konstrukt korrelieren
- Wird für relevante Ausprägungen auf dem Hintergrundfaktor jeweils eine eigene Norm gebildet, kann der Einfluss des Hintergrundfaktors auf die Testwertinterpretation kontrolliert werden
- Bsp.: Männer haben im Durchschnitt niedrigere Neurotizismus-Testwerte als Frauen ($M_{\text{Männer}} = 19.64$, $M_{\text{Frauen}} = 23.25$)

Normen NEO-FFI Neurotizismus

Bevölkerungs-
repräsentativ

Test- wert	cum %	Standardwert	
		T	ST
0	0.11	20	1
1	0.11	20	1
2	0.29	22	1
3	0.52	24	1
4	0.75	26	1
5	0.98	27	1
6	1.32	28	1
7	1.72	29	1
8	2.70	31	1
9	3.90	32	1
10	5.51	34	2
11	7.75	36	2
12	10.79	38	3
13	14.47	39	3
14	18.25	41	3
15	22.79	43	4
16	28.36	44	4
17	34.44	46	4
18	39.72	47	4
19	44.95	49	5
20	50.06	50	5
21	54.71	51	5
22	59.76	52	5
23	64.24	54	6
24	68.14	55	6
25	72.27	56	6
26	75.77	57	6

Männer

Test- wert	cum %	Standardwert	
		T	ST
0	0.12	20	1
1	0.12	20	1
2	0.35	23	1
3	0.71	25	1
4	1.06	27	1
5	1.30	28	1
6	1.77	29	1
7	2.48	30	1
8	3.90	32	1
9	5.67	34	2
10	7.57	36	2
11	10.05	37	2
12	13.36	39	3
13	17.73	41	3
14	22.81	43	4
15	28.25	44	4
16	34.04	46	4
17	40.66	48	5
18	46.69	49	5
19	51.89	50	5
20	56.74	52	5
21	61.47	53	6
22	66.19	54	6
23	70.21	55	6
24	74.23	57	6
25	77.90	58	7
26	80.73	59	7

Frauen

Test- wert	cum %	Standardwert	
		T	ST
0	0.11	19	1
1	0.11	19	1
2	0.33	23	1
3	0.33	23	1
4	0.67	25	1
5	0.67	25	1
6	1.00	27	1
7	1.00	27	1
8	1.56	28	1
9	2.23	30	1
10	3.57	32	1
11	5.58	34	2
12	8.37	36	2
13	11.38	38	3
14	13.95	39	3
15	17.63	41	3
16	22.99	43	4
17	28.57	44	4
18	33.15	46	4
19	38.39	47	4
20	43.75	48	5
21	48.33	50	5
22	53.68	51	5
23	58.59	52	5
24	62.39	53	6
25	66.96	54	6
26	71.09	56	6

4. Testnormierung

Normdifferenzierung

- Normdifferenzierung kann zu sinnvollen diagnostischen Entscheidungen beitragen
- Normdifferenzierung birgt auch Gefahren
 - Wettbewerbssituation: Z. B. Auswahl von Bewerber*innen mit Referenzgruppe von Personen, die ihnen ähnlich ist vs. mit Referenzgruppe der späteren Kolleg*innen/Konkurrent*innen
 - Überanpassung der Normen (Overadjustment):
 - Durch zu starke Anpassung der Normen an Teilpopulationen kann das Testresultat an Aussagekraft verlieren
 - Mögliche Folge: Fehleinschätzungen wie Nivellierung tatsächlich vorhandener Unterschiede zwischen Personen aus unterschiedlichen Teilpopulationen

4. Testnormierung

Dokumentation der Normen

- Geltungsbereich: Definition der Zielpopulation
- Art der Stichprobe: Grad der Repräsentativität
- Stichprobenumfang und -zusammensetzung
- Deskriptive Statistiken
- Jahr der Datenerhebung

Standards für psychologisches Testen

Standards für psychologisches Testen

1. Definition und Ziele von Standards
2. Standards für die Entwicklung und Evaluation psychologischer Tests
 1. Standards for Educational and Psychological Testing
 2. DIN 33430
3. Bsp.: Standards zur Validität
4. Standards zur Qualitätsbeurteilung psychologischer Tests:
TBS-TK

1. Definition und Ziele von Standards

- Standards sind vereinheitlichte Leitlinien, die sich auf verschiedene Bereiche des Testens beziehen:
 - Testkonstruktion
 - Testadaptation (z.B. Übersetzung)
 - Testanwendung
 - Qualitätsbeurteilung
- Ziele:
 - Optimierung der Prozesse in den verschiedenen Bereichen
 - Unterstützung des Ziehens valider Schlussfolgerungen aus den Testergebnissen
- „The intent of the *Standards* is to promote the sound and ethical use of tests and to provide a basis for evaluating the quality of testing practices.“ (AERA, APA & NCME, 1999, S.1)

2. Standards für die Entwicklung und Evaluation psychologischer Tests

- Verschiedene Organisationen haben Teststandards entwickelt
- Für die Entwicklung und Evaluation psychologischer Tests sind v.a. zwei Teststandard-Kompendien wichtig
 1. Standards for Educational and Psychological Testing (SEPT)
 2. DIN 33430
- Außerdem veröffentlicht die International Test Commission (<https://www.intestcom.org>) regelmäßig Richtlinien zur Testentwicklung und -anwendung in bestimmten Kontexten
 - Z.B. The ITC Guidelines for Translating and Adapting Tests

2.1 SEPT

- Standards for Educational and Psychological Testing (SEPT) werden herausgegeben von drei Organisationen: American Educational Research Association, American Psychological Association und National Council on Measurement in Education
- Teile:
 1. Test construction, evaluation, and documentation
 - Validität, Reliabilität, Normierung, Testauswertung, Dokumentation
 2. Fairness in testing
 - Rechte von Testpersonen
 - Testung von Personen mit verschiedenen sprachlichen Hintergründen, Personen mit Behinderungen
 3. Testing applications
 - Verantwortlichkeiten von Testanwender*innen
 - Anwendung von Tests in unterschiedlichen Bereichen (Psychologie, Bildung, Arbeitswelt, ...)

2.1 SEPT

- Beispiel: Standards zu Validität
- Standard 1.2
„The test developer should set forth clearly how test scores are intended to be interpreted and used. The population(s) for which a test is appropriate should be clearly delimited, and the construct that the test is intended to assess should be clearly described.“
(AERA, APA & NCME, 1999, S. 17)

2.1 SEPT

- Beispiel: Standards zu Fairness
- Standard 7.4
„Test developers should strive to identify and eliminate language, symbols, words, phrases, and content that are generally regarded as offensive by members of racial, ethnic, gender, or other groups except when judged to be necessary for adequate representation of the domain.“
(AERA, APA & NCME, 1999, S. 82)

2.2 DIN 33430

- „Anforderungen an berufsbezogene Eignungsdiagnostik“
- Die DIN 33430 beschreibt
 - Qualitätskriterien und -standards für berufsbezogene Eignungsbeurteilungen (Planung, Auswahl von Verfahren, Durchführung, Auswertung, Interpretation und Urteilsbildung)
 - Qualifikationsanforderungen an die an der Eignungsbeurteilung beteiligten Personen
- DIN 33430 wendet sich an Auftraggeber (z.B. Unternehmen), Auftragnehmer (Diagnostiker mit DIN-Lizenz) und Mitwirkende (z.B. bei Verhaltensbeobachtung)

2.2 DIN 33430

Beispiele für Anforderungen nach DIN 33430

- Arbeits- und Anforderungsanalyse als Grundlage
- Ausführliche Verfahrenshinweise (Manuale) für alle Verfahren
- *Vorab* Festlegung aller Regeln (z.B. Interpretation)
- Gültigkeit der Reliabilitäts-, Validitäts- und Normwerte ist spätestens alle acht Jahre zu überprüfen
- Streubreite der Urteile mehrerer Assessoren ist festzuhalten
- Nachvollziehbare Dokumentation des Prozesses

Standards für psychologisches Testen

1. Definition und Ziele von Standards
2. Standards für die Entwicklung und Evaluation psychologischer Tests
 1. Standards for Educational and Psychological Testing
 2. DIN 33430
3. **Bsp.: Standards zur Validität**
4. Standards zur Qualitätsbeurteilung psychologischer Tests:
TBS-TK

3. Bsp.: Standards zur Validität

- Für die verschiedenen Aspekte der Validität müssen aktuelle (< 8 Jahre) empirische Belege vorliegen
- **Inhaltsvalidität**
 - Definition des vom Test abgebildeten Inhaltsbereichs und seiner Relevanz für die vorgesehene Testanwendung
 - Bei Expertenurteilen Darlegung der Qualifikation der Expert*innen
- **Kriteriumsvalidität**
 - Genaue Beschreibung der Kriteriumsmaße und deren Erfassung
- **Konstruktvalidität**
 - Interessierendes Konstrukt muss von anderen Konstrukten klar abgegrenzt werden
 - Darlegen, wie sich das Konstrukt zu ähnlichen und unähnlichen Konstrukten verhält
 - Interpretation der Testwerte darstellen

Standards für psychologisches Testen

1. Definition und Ziele von Standards
2. Standards für die Entwicklung und Evaluation psychologischer Tests
 1. Standards for Educational and Psychological Testing
 2. DIN 33430
3. Bsp.: Standards zur Validität
4. **Standards zur Qualitätsbeurteilung psychologischer Tests: TBS-TK**

4. Standards zur Qualitätsbeurteilung psychologischer Tests: TBS-TK

- TBS-TK: Testbeurteilungssystem – Testkuratorium der Föderation deutscher Psychologenvereinigungen
- Mit dem TBS-TK werden Bewertungen von Tests vorgenommen
- Darüber hinaus soll das TBS-TK aber auch als Richtlinie bei der Testentwicklung sowie bei der Erstellung des Testmanuals dienen
- Ziel: Qualitätssicherung von Tests
- Kriterien und Testrezensionen unter <https://www.psyndex.de/tests/testkuratorium/>

4. TBS-TK

Tabelle 1. Besprechungs- und Beurteilungskategorien

	Bewertung
1. Allgemeine Informationen über den Test, Beschreibung des Tests und seiner diagnostischen Zielsetzung	frei und formalisiert*
2. Theoretische Grundlagen als Ausgangspunkt der Testkonstruktion	frei
3. Objektivität	frei und formalisiert*
4. Normierung (Eichung)	frei
5. Zuverlässigkeit (Reliabilität, Messgenauigkeit)	frei und formalisiert*
6. Gültigkeit (Validität)	frei und formalisiert*, auch unter Berücksichtigung der Fairness (soweit in Anspruch genommen)
7. Weitere Gütekriterien (Störanfälligkeit, Unverfälschbarkeit und Skalierung)	frei
8. Abschlussbewertung/Empfehlung	frei

4. TBS-TK

Bsp.: Zusammenfassung der TBS-TK Rezension für den I-S-T 2000 R

T B S - T K Rezension	Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R). 2., erweiterte und überarbeitete Auflage	Die TBS-TK-Anforderungen sind erfüllt			
		voll	weitgehend	teilweise	nicht
Testbeurteilungssystem - Testkuratorium der Föderation deutscher Psychologenvereinigungen	Allgemeine Informationen, Beschreibung und diagnostische Zielsetzung	●			
	Objektivität	●			
	Zuverlässigkeit		●		
	Validität		●		

Literatur zu dieser Sitzung

Moosbrugger & Kelava (2020). Kapitel 9 (ohne 9.3.1).

Moosbrugger & Kelava (2020). Kapitel 10.1, 10.2 und 10.5.