

Zur Information über und Beurteilung von Tests

client
logo



Deutsche Gesellschaft
für Psychologie

Das Testbeurteilungssystem des Diagnostik- und Testkuratoriums (TBS-DTK)



Berufsverband
Deutscher
Psychologinnen
und Psychologen

Die Unterlagen sind ausschließlich zum persönlichen Gebrauch.

Bisheriges TK System: Kriterienkatalog f. d. Beurteilung von psychologischen Tests (Testkuratorium, 1986)

client
logo

- Testkuratorium der Föderation Deutscher Psychologenverbände (1986).

1. Testgrundlage

- 1.1 Diagnostische Zielsetzung
- 1.2 Theoretische Grundlagen
- 1.3 Nachvollziehbarkeit der Testkonstruktion

2. Testdurchführung

- 2.1 Durchführungsobjektivität
- 2.2 Transparenz
- 2.3 Zumutbarkeit
- 2.4 Verfälschbarkeit
- 2.5 Störanfälligkeit

3. Testverwertung

- 3.1 Auswertungsobjektivität
- 3.2 Zuverlässigkeit
- 3.3 Gültigkeit
- 3.4 Normierung
- 3.5 Bandbreite
- 3.6 Informationsausschöpfung
- 3.7 Änderungssensitivität

4. Testdurchführung

- 4.1 Ökonomie
- 4.2 Fairness
- 4.3 Akzeptanz
- 4.4 Vergleichbarkeit
- 4.5 Bewährung

5. Äußere Testgestaltung

Das TBS-DTK System ersetzt den von der Föderation Deutscher Psychologenverbände (1986) publizierten Kriterienkatalog für die Testbeurteilung!

z. B. 3.2 Zuverlässigkeit: Messgenauigkeit der beobachteten interindividuellen Unterschiede der Testergebnisse durch tatsächliche psychische Merkmalsunterschiede. Für die Bewertung ist die Angabe der verwendeten Berechnungsverfahren erforderlich.

Föderation Deutscher Psychologenverbände (1986). Beschreibung der einzelnen Kriterien für die Testbewertung. *Diagnostica*, 32, 358-360.

◆ Diagnostik- u. Testkuratorium ◆ Testbeurteilungssystem des Testkuratoriums (TBS-DTK) ◆ Stand: Sep. 2019 ◆

Gründe für die Einführung eines neuen Systems zur Informationen über und Beurteilung von Tests (1/2)

client
logo

- Die Anzahl der Testrezensionen ist quantitativ unbefriedigend (Initiative der DGPs kam nicht in Gang. Ein Grund u. a.: Testrezensionen gelten als “nachrangige” wissenschaftliche Leistung).
- Die vorhandenen Testrezensionen sind unsystematisch aufbereitet und verstreut publiziert, so dass ein zentraler Zugriff erschwert wird.
- Die Qualität der Testbeurteilungen ist heterogen, das Zustandekommen der Bewertung ist häufig nicht nachvollziehbar. Standards der Bewertung wissenschaftlicher Arbeiten (peer-review) werden nicht gewährleistet.
- Die vorhandenen Testrezensionen repräsentieren nicht die vorhandenen Tests.

Kersting, M. (2006). Zur Beurteilung der Qualität von Tests: Resümee und Neubeginn. *Psychologische Rundschau*, 57, 243-253

Gründe für die Einführung eines neuen Systems zur Informationen über und Beurteilung von Tests (2/2)

client
logo

- Die expliziten oder impliziten Urteile über Tests fallen in den deutschsprachigen Rezensionen häufig entweder unverbindlich oder aber extrem kritisch aus.
- In einem Themenheft der ZDDP aus dem Jahr 1997 wurden 25 Tests unter die Lupe genommen. In seinem Editorial nimmt Kubinger (1997, S. 3) das Fazit vorweg: Die meisten der 25 einschlägigen Tests seien nicht tauglich.
- Die Wirkung solcher negativen Beurteilungen ist auch vor dem Hintergrund der Tatsache zu würdigen, dass überwiegend nur die Tests rezensiert werden, die einen universitären Konstruktionshintergrund aufweisen, während die Tests von Anbieter ohne universitären Hintergrund der Rezension entgehen.

Kersting, M. (2006). Zur Beurteilung der Qualität von Tests: Resümee und Neubeginn. *Psychologische Rundschau*, 57, 243-253

Kubinger, K.D. (1997). Editorial zum Themenheft >Testrezensionen: 25 einschlägige Verfahren<. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 18, 1-3.

◆ Diagnostik- u. Testkuratorium ◆ Testbeurteilungssystem des Testkuratoriums (TBS-DTK) ◆ Stand: Sep. 2019 ◆

Das Testbeurteilungssystem des Diagnostik- und Testkuratoriums

client
logo

Literatur:



Deutsche Gesellschaft
für Psychologie



Berufsverband
Deutscher
Psychologinnen
und Psychologen

Das Beurteilungssystem:

Diagnostik- und Testkuratorium. (2018). TBS-DTK. Testbeurteilungssystem des Diagnostik- und Testkuratoriums der Föderation Deutscher Psychologinnenvereinigungen. Revidierte Fassung vom 03. Jan. 2018. *Psychologische Rundschau*, 69, 109-116.

→ Freier Download unter: <http://www.zpid.de/Testkuratorium>
http://kersting-internet.de/pdf/DTK_2018_PR_TBS-DTK-dritte_Fassung.pdf

Erläuterungen:

Kersting, M. (2006). Zur Beurteilung der Qualität von Tests: Resümee und Neubeginn. *Psychologische Rundschau*, 57, 243-253.

Moosbrugger, H., Stemmler, G. & Kersting, M. (2008). Qualitätssicherung und -optimierung im Aufbruch: Die ersten Testrezensionen nach dem neuen TBS-TK-System. *Report Psychologie*, 33, 299-300 sowie *Psychologische Rundschau*, 59, 182-184.

→ Freier Download unter: <http://www.zpid.de/Testkuratorium>

◆ Diagnostik- u. Testkuratorium ◆ Testbeurteilungssystem des Testkuratoriums (TBS-DTK) ◆ Stand: Sep. 2019 ◆

TBS-DTK System und andere Testbeurteilungssysteme: Gemeinsamkeiten und Unterschiede

client
logo

Das Diagnostik- und Testkuratorium (DTK) hat sich im Zuge der Entwicklung eines neuen Systems zur Information über Tests und deren Beurteilung zunächst ausführlich mit den in anderen Ländern etablierten Testinformations- und Testbeurteilungssystemen (z. B. dem niederländischen COTAN System) sowie vor allem mit der Initiative der European Federation of Psychologists Associations (EFPA) zur Entwicklung eines „common set of European criteria for test reviews“ auseinandergesetzt und zahlreiche Ideen dieser Systeme übernommen.

- Gegen eine vollständige Übernahme eines etablierten Systems wie COTAN oder EFPA sprach nach Ansicht des DTK vor allem die dort vorgesehene Bewertung der numerischen Ausprägung von Kennwerten nach vorgegebenen Regeln. So verlangt das COTAN System beispielsweise, dass eine Reliabilität kleiner als 0,80 für Tests, die für wichtige Entscheidungen auf individueller Ebene eingesetzt werden, als „unzureichend“ zu bewerten sei, usw. Derartig starre Beurteilungsvorschriften übersehen nach Ansicht des DTK, dass die Ausprägung der Kennwerte nicht nur von der Qualität des Tests, sondern auch von Merkmalen der jeweils herangezogenen Untersuchungsgruppe abhängt.
- Das TBS-DTK begnügt sich nicht mit der schematischen Klassifikation von Koeffizienten, sondern setzt für die Testbewertung auf die Urteilskraft ausgewiesener u. einschlägig qualifizierter Experten, die üb. ein tiefes Verständnis d. jeweiligen Konstrukte; Operationalisierungen; diagnostischen Zielsetzungen u. Rahmenbedingungen sowie üb. ein fundiertes psychometrisches Wissen verfügen.

Kersting, M. (2006). Zur Beurteilung der Qualität von Tests: Resümee und Neubeginn. *Psychologische Rundschau*, 57, 243-253

Moosbrugger, H., Stemmler, G. & Kersting, M. (2008). Qualitätssicherung und Optimierung im Aufbruch: Die ersten Testrezensionen nach d. neuen TBSTKSystem. *Report Psychologie*, 33, 299-300 und *Psychologische Rundschau*, 59, 182-184

◆ Diagnostik- u. Testkuratorium ◆ Testbeurteilungssystem des Testkuratoriums (TBS-DTK) ◆ Stand: Sep. 2019 ◆

Testbeurteilungen und die DIN 33430

- Auf den ersten Blick ist die DIN 33430 nicht anwendbar, denn sie
 - (1) bezieht sich nur auf berufsbezogene Eignungsbeurteilungen
 - (2) ist eine Prozess- und keine Produktnorm
 - (3) bezieht sich nicht explizit auf Tests, sondern auf alle Verfahren
- Kersting (2008) hat eine Checkliste zur DIN 33430 erstellt und dafür den Text der DIN 33430 in 318 Einzelaussagen zerlegt sowie thematisch geordnet. Dabei stellte sich heraus: 140 Aussagen (44% aller Aussagen) der DIN 33430 formulieren Anforderungen an Verfahrenshinweise.
- Diese Teilmenge der DIN 33430 wurde daher zu einer eigenständigen "Checkliste 1" „Anforderungen an Verfahrenshinweise“ zusammengefasst → freier Download unter:

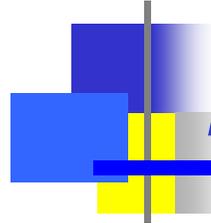
http://kersting-internet.de/pdf/Kersting_2017_in_DTK_DIN-SCREEN_Checkliste_1.pdf

Die Checkliste 1 gilt offiziell als „Standard zur Information und Dokumentation von Instrumenten zur Erfassung menschlichen Erlebens und Verhaltens“ des Diagnostik- und Testkuratoriums der Föderation Deutscher Psychologenvereinigungen.
- Während sich zahlreiche Aussagen der DIN 33430 explizit nur auf die Eignungsbeurteilung beziehen, kann die hier in Frage stehende Teilmenge verfahrensbezogenen Aussagen der DIN 33430 auf Tests aus allen Anwendungsbereichen angewendet werden.

Kersting, M. (2018). Zur Information über und Dokumentation von Instrumenten zur Erfassung menschlichen Erlebens und Verhaltens – Die DIN SCREEN Checkliste 1, Version 3. In: Diagnostik- und Testkuratorium (Hrsg.) Personalauswahl kompetent gestalten: Grundlagen und Praxis der Eignungsdiagnostik nach DIN 33430 (S. 223-244). Berlin: Springer. (DOI

TBS-DTK System: Zwei Schritte

- Der erste Schritt umfasst eine Prüfung der Verfahrenshinweise (i.d.R. des Testmanuals) auf die Erfüllung der in der DIN 33430 formulierten Anforderungen an die Vollständigkeit der Information in Verfahrenshinweisen (Testmanual). Die Operationalisierung dieser Anforderungen erfolgt mit der „Checkliste 1“ der Publikation „DIN Screen“ (Download: http://www.kersting-internet.de/DIN-Buch/din-buch_downloads.html).
- Der zweite Schritt beinhaltet die eigentliche Bewertung des Tests anhand von vorgegebenen Besprechungs- und Beurteilungskategorien des TBS-DTK durch zwei unabhängig voneinander arbeitende Rezensent(inn)en. Für einige Kategorien erfolgt darüber hinaus eine formalisierte Bewertung auf einer vierstufigen Skala. Die Gesamtlänge der Bewertung darf 12000 Zeichen nicht überschreiten. Bei der Bewertung sollen die im Anhang des TBS-DTK aufgeführten Gesichtspunkte Berücksichtigung finden.



DIN - Checkliste 1: Beispiel



Tab. A.1 Beispiel für die Antwortkategorien der „DIN SCREEN V3/Checkliste 1/Anforderungen an Handhabungs- und Verfahrenshinweise“ sowie für eine Bewertungsverknüpfung

		SOLL(S) oder MUSS (M)	ja	nein	nicht zu bewerten	Anmerkungen	Quelle (Seite)
B17	Die Angemessenheit der Normwerte wurde in den letzten acht Jahren überprüft	S	<input type="radio"/>	<input type="radio"/>			
V7	<i>Verzweigungsfrage V7: Wurde die Angemessenheit der Normwerte in den letzten acht Jahren überprüft? Falls „ja“-> Bitte weiter bei Verzweigungsfrage V8</i>		<input type="checkbox"/>	<input type="radio"/>			
B18	Es wird begründet, warum und unter welchen Umständen das Verfahren für einen Anwendungsfall ausgewählt werden kann, obwohl die Angemessenheit der Normwerte nicht in den letzten acht Jahren überprüft wurde	M	<input type="radio"/>	<input type="radio"/>			
V8	<i>Verzweigungsfrage V8: Zielt das Verfahren auf die Erfassung eines Eignungsmerkmals ab, dessen Ausprägung in der Referenzgruppe möglicherweise relativ kurzfristigen Veränderungen unterliegt? Falls „nein“-> Bitte weiter bei Verzweigungsfrage V9</i>		<input type="radio"/>	<input type="checkbox"/>			
B19	Die Angemessenheit der Normwerte wurde bereits vor Ablauf der Acht-Jahres-Frist empirisch gezeigt	M	<input type="radio"/>	<input type="radio"/>			

Die Unterlagen sind ausschließlich zum persönlichen Gebrauch.

Kersting, M. (2018). Zur Information über und Dokumentation von Instrumenten zur Erfassung menschlichen Erlebens und Verhaltens – Die DIN SCREEN Checkliste 1, Version 3. In: Diagnostik- und Testkuratorium (Hrsg.) Personalauswahl kompetent gestalten: Grundlagen und Praxis der Eignungsdiagnostik nach DIN 33430 (S. 223-244). Berlin: Springer. (DOI

DIN 33430 Anforderungen an Verfahrenshinweise

- Mit dem TBS-DTK System werden Bewertungen von Tests vorgenommen. Über diese Bewertung hinaus soll das System aber vor allem zur Verbesserung der Testentwicklung beitragen, indem die Informationsanforderungen und Bewertungskriterien Testautoren als „Richtlinie“ bei der Testentwicklung sowie bei der Erstellung der Verfahrenshinweise (Testmanuale) dienen. In diesem Sinne ist das TBS-DTK System kulturprägend.
- Dies kommt etwas darin zum Ausdruck, dass in den Verfahrenshinweisen (Testmanualen) von Neupublikationen von Testverfahren mit einem hohen Qualitätsanspruch standardmäßig bereits Übersichten darüber vorgehalten werden, auf welcher Seite des Testmanuals die nach dem „Standard zur Information und Dokumentation von Instrumenten zur Erfassung menschlichen Erlebens und Verhaltens“ des Diagnostik- und Testkuratoriums der Föderation Deutscher Psychologenvereinigungen geforderten Informationen zu finden sind (Beispiele hierfür sind die Verfahrenshinweise zum BIP-6F, d2-R, FAKT, START-P und WIT-2).

Besprechungs- und Beurteilungskategorien des TBS-DTK Systems

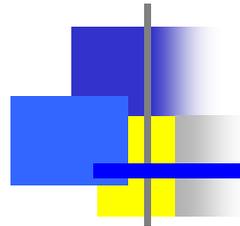


- Seit der Version 2018 sieht das TBS-DTK die explizite Forderung vor, dass die Verfahrenshinweise (Handbücher) eine tabellarische Übersicht beinhalten, in der verzeichnet ist, auf welchen Seiten der Verfahrenshinweise sich die nach DIN Checkliste 1 als relevant ausgewiesenen Informationen zum Test befinden.
- Dem Vorhandensein oder Fehlen dieser Tabelle ist eine eigenständige Beurteilungskategorie in der tabellarischen Bewertung der Tests gewidmet.

In den Verfahrenshinweisen ist verzeichnet, wo die nach dem DTK-Testinformationsstandard notwendigen Informationen zu finden sind.	ja	nein x
--	----	-----------

Der Test erfüllt die Anforderungen	voll	weit- gehend	teilweise	nicht
Informationsgehalt der Verfahrenshinweise	x			
Objektivität	x			
Zuverlässigkeit (Reliabilität, Messgenauigkeit)		x		
Einigkeit/Validität			x	

Diagnostik- und Testkuratorium der Föderation Deutscher Psychologinnenvereinigungen (2018). TBS-DTK. Testbeurteilungssystem des Diagnostik- und Testkuratoriums der Föderation Deutscher Psychologinnenvereinigungen. Revidierte Fassung vom 3. Januar 2018. Psychologische Rundschau (2018), 69, 109-148.



DIN – Screen im Testhanbuch, Beispiel BIP 6-F



Tabelle 79: Zuordnung der 140 Anforderungen an Verfahrenshinweise der DIN 33430 (gemäß Checkliste 1 der DIN Screen V2; Kersting, 2008) zu den Kapiteln der vorliegenden Handanweisung, die die nachgefragten Informationen enthalten

Nr.	Beleg in Kapitel ...	Nr.	Beleg in Kapitel ...	Nr.	Beleg in Kapitel ...
1.	konkludent ¹	48.	4.8.1	95.	4.3.3, 7.1
2.	konkludent	49.	4.8.1	96.	4.3.3
3.	konkludent	50.	konkludent	97.	nicht relevant
4.	konkludent ²	51.	konkludent	98.	nicht relevant
5.	konkludent ²	52.	konkludent	99.	nicht relevant
6.	nicht relevant ²	53.	2, 3.1.1	100.	4.8, 5.4.2.1
7.	nicht relevant ²	54.	2, 3	101.	4.7
8.	5.1.3, 5.1.4	55.	nicht relevant	102.	4.7
9.	nicht relevant ²	56.	4.8.1, 4.8.7	103.	5.2.1
10.	nicht relevant ²	57.	4.8.1	104.	4.8, 5.4.2.1
11.	konkludent	58.	4.8.1	105.	4.8
12.	konkludent	59.	nicht relevant	106.	3.1.1, 4.3.2, 5.4.2.3
13.	konkludent ³	60.	5.5.1	107.	3.1.1, 4.3.2
14.	5.2.1	61.	4.1	108.	3.1.1, 5.4.2.3
15.	5.1, 5.2.1	62.	4.1	109.	4.2, 4.3, 4.6, 4.8
16.	1, 5.1	63.	5.2	110.	2.1.1, 3.1.1, 4.3.2
17.	1, 5.1	64.	5.3	111.	nicht relevant
18.	5.4.1	65.	5.4	112.	nicht relevant

Die Unterlagen sind ausschließlich zum persönlichen Gebrauch.

Datei: ...Kersting\TBS_TK_Präsentation_Stand_Sep_2019.ppt

Überblickstabelle: Wo in den Verfahrenshinweisen finde ich die qualitätsrelevanten Informationen ?

client
logo

Anhang E

Tabelle 109: Zuordnung der Informationen des Manuals zu den „Anforderungen an Verfahrenshinweise“ der DIN 33430 (laut Checkliste Nr. 1 von Kersting, 2008)

DIN Nr.	Beleg	DIN Nr.	Beleg	DIN Nr.	Beleg
1.	konkludent	48.	S. 55	95.	S. 67, 69, 126
2.	konkludent	49.	S. 55	96.	S. 52–54, 67 ff., 126
3.	konkludent	50.	konkludent	97.	S. 62 ff.
4.	konkludent ¹	51.	konkludent	98.	S. 62 ff.
5.	konkludent ¹	52.	konkludent	99.	S. 62 ff.
6.	S. 63 f., 76, 78, 79 f.	53.	S. 23–29	100.	nicht relevant
7.	S. 14, 15	54.	S. 23–29	101.	S. 46 ff.
8.	S. 86	55.	S. 23–29	102.	S. 46 ff.
9.	konkludent	56.	S. 43 f.	103.	S. 51
10.	S. 18–22	57.	S. 43 f.	104.	S. 43–49
11.	konkludent	58.	S. 55	105.	S. 43–49
12.	konkludent	59.	nicht relevant	106.	S. 23–29, 70–81, 127–129
13.	konkludent und S. 36, 38 f.	60.	Für den WIT-1 → Bettinville et al. ⁵	107.	S. 23–29, 70–81, 127–129
14.	S. 36	61.	S. 56 f.	108.	S. 23–29
15.	S. 11	62.	S. 56 f.	109.	S. 23–29, 70–81, 127–129
16.	S. 11	63.	S. 32–39 ²	110.	S. 23–26, 73–75
17.	S. 11, 61–80	64.	S. 40–42	111.	S. 75, 78
18.	S. 50 f.	65.	S. 43–54	112.	S. 75–81, 127–129
19.	S. 50 f.	66.	S. 37 f.	113.	S. 23–29, 75–81, 127–129
20.	nicht relevant	67.	S. 40–46	114.	S. 75–81, 127–129

Beispiel aus: Kersting, M., Althoff, K. & Jäger, A. O. (2008). WIT-2. Der Wilde-Intelligenztest. Verfahrenshinweise. Göttingen: Hogrefe. u

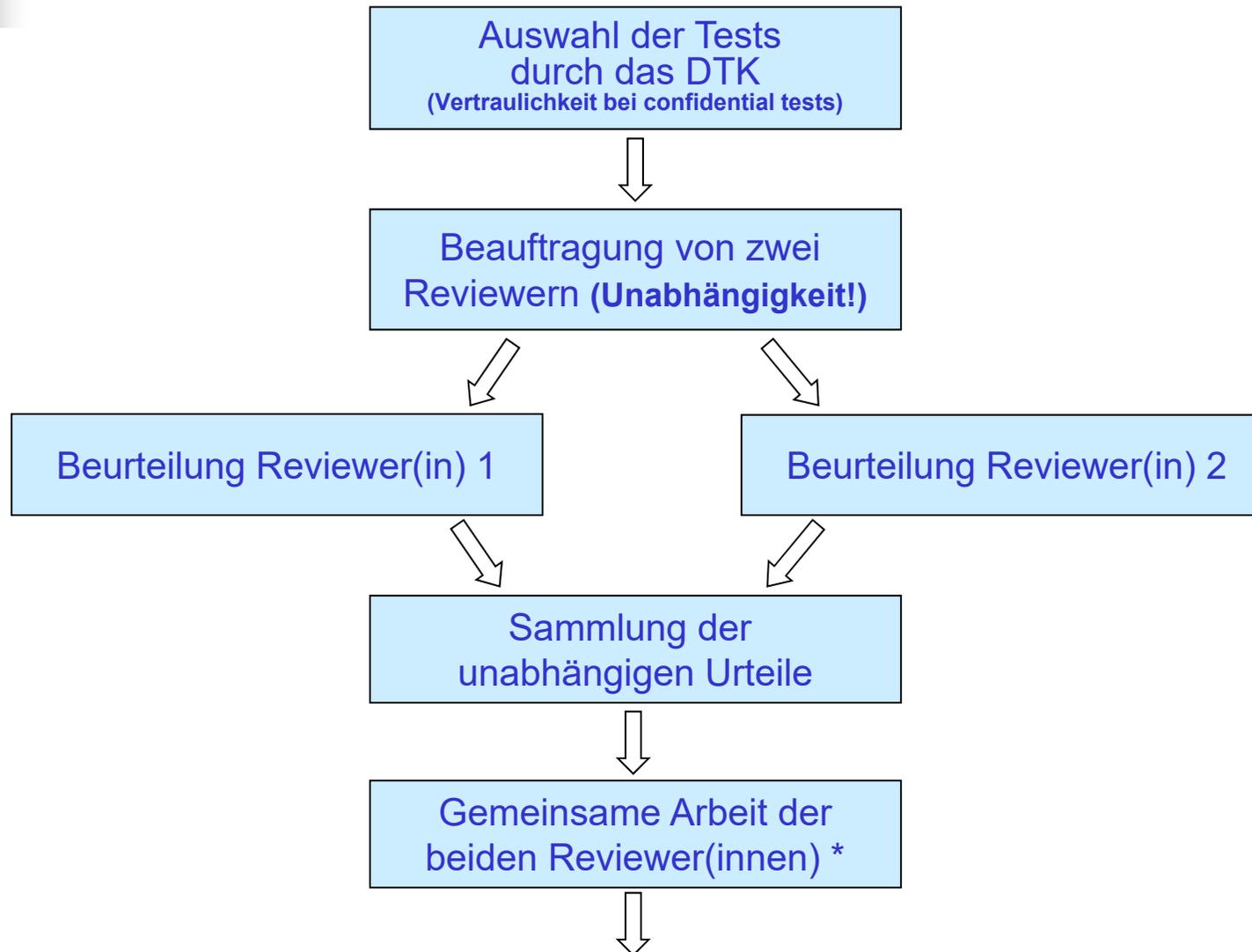
◆ Diagnostik- u. Testkuratorium ◆ Testbeurteilungssystem des Testkuratoriums (TBS-DTK) ◆ Stand: Sep. 2019 ◆

Selbsterklärungen im Rahmen d. TBS-DTK Systems

- Seit der TBS-DTK Version 2018 besteht die Möglichkeit zu Selbsterklärungen: Verfahrensanbieter(innen) / können ein Zertifikat beantragen, über dessen Vergabe das DTK entscheidet. Das Zertifikat berechtigt die Verfahrensanbieter(innen) dazu, mit der folgende Aussage für das Verfahren zu werben:
 - *„Die Verfahrenshinweise zum Test (Bezeichnung) erfüllen den Qualitätsanspruch des Diagnostik- und Testkuratoriums an Information u. Dokumentation von Instrumenten zur Erfassung menschlichen Erlebens und Verhaltens. Damit erfüllt der Test (Bezeichnung) nach Ansicht des DTK die Voraussetzungen, um einer Qualitätskontrolle unterzogen zu werden. Diese Qualitätskontrolle selbst hat das DTK für den Test (Bezeichnung) nicht vorgenommen.“*
- Die Voraussetzungen für die Vergabe des Zertifikats sind in der Publikation (Diagnostik- und Testkuratorium , 2018) spezifiziert.

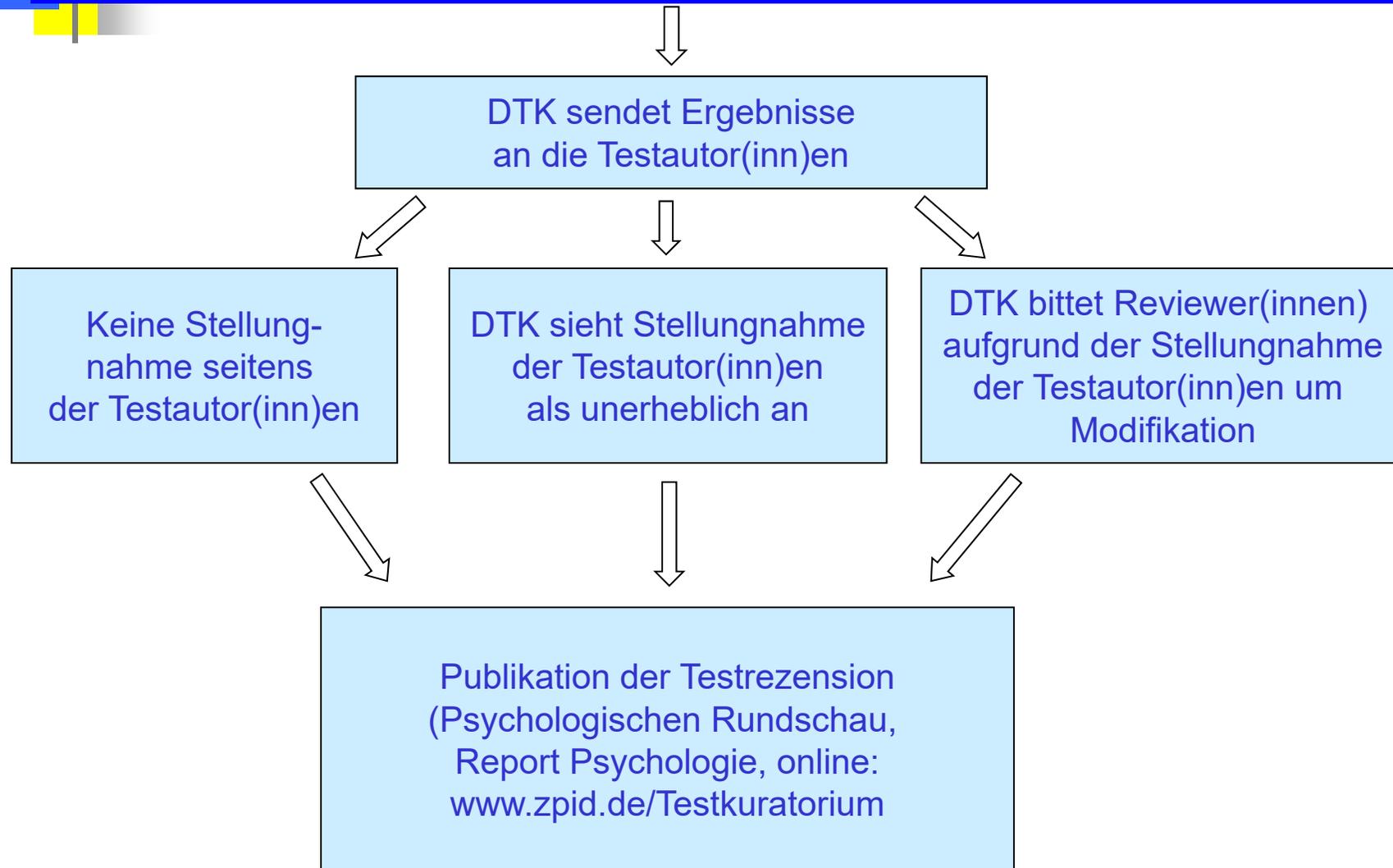
Das Testbeurteilungssystem des Diagnostik- und Testkuratoriums – formaler Ablauf I –

client
logo



Das Testbeurteilungssystem des Diagnostik- und Testkuratoriums – formaler Ablauf II –

client
logo



Die Unterlagen sind ausschließlich zum persönlichen Gebrauch.

Diagnostik- und Testkuratorium der Föderation Deutscher Psychologinnenvereinigungen (2018). TBS-DTK. Testbeurteilungssystem des Diagnostik- und Testkuratoriums der Föderation Deutscher Psychologinnenvereinigungen. Revidierte Fassung vom 3. Januar 2018. *Psychologische Rundschau* (2018), 69, 109-148.

Datei: ...Kersting\TBS_TK_Präsentation_Stand_Sep_2019.ppt

Das Testbeurteilungssystem des Diagnostik- und Testkuratoriums



Die Unterlagen sind ausschließlich zum persönlichen Gebrauch.
* Zur Skala: Eine Ausnahme ist die Kategorie 3, bei der eine dichotome Bewertung (Ja / Nein) vorgenommen wird.

Grundlage der Prüfung auf Stufe 1 ist die DIN SCREEN „Checkliste 1“, Version 3 (Kersting, 2018), die offiziell als „Standard zur Information und Dokumentation von Instrumenten zur Erfassung menschlichen Erlebens und Verhaltens des Testkuratoriums der Föderation Dt. Psychologengvereinigungen“ gilt.

Prüfung der Informationsgrundlage nach DIN 33430



Testbeurteilung und Rezension



5 Aspekte (z. B. Normierung) frei



4 Aspekte (z. B. Reliabilität, Validität) frei und formalisiert, 1 Aspekt nur formalisiert



Der Test erfüllt die Anforderungen...

voll – weitgehend – teilweise – nicht*

nicht prüffähig



Kersting, M. (2018). Zur Information über und Dokumentation von Instrumenten zur Erfassung menschlichen Erlebens und Verhaltens – Die DIN SCREEN Checkliste 1, Version 3. In: Diagnostik- und Testkuratorium (Hrsg.) Personalauswahl kompetent gestalten: Grundlagen und Praxis der Eignungsdiagnostik nach DIN 33430 (S. 223-224)

Besprechungs- und Beurteilungskategorien des TBS-DTK Systems

client
logo

Kategorien	Bewertung
Beschreibung des Tests und seiner diagnostischen Zielsetzung	frei
Bewertung des Informationsgehalts der Verfahrenshinweise	frei und formalisiert*
Prüfung, ob in den Verfahrenshinweisen verzeichnet ist, wo die nach dem DTK-Testinformationsstandard notwendigen Informationen zu finden sind	formalisiert*
Theoretische Grundlagen als Ausgangspunkt der Testkonstruktion	frei
Objektivität	frei und formalisiert*
Normierung (Eichung)	frei
Zuverlässigkeit (Reliabilität, Messgenauigkeit)	frei und formalisiert*
Gültigkeit (Validität)	frei und formalisiert*, auch unter Berücksichtigung der Fairness (soweit in Anspruch genommen)
Weitere Gütekriterien (Störanfälligkeit, Unverfälschbarkeit und Skalierung)	frei
Abschlussbewertung	frei

Die Gesamtlänge der Bewertung beträgt 12000 Zeichen erhöht.

(*) Zur formalen Bewertung ist eine vierstufige Skala vorgesehen: „Der Test erfüllt d. Anforderungen (...)“
 (a) voll,
 (b) weitgehend,
 (c) teilweise und
 (d) nicht.“

Eine Ausnahme ist die Kategorie 3, bei der eine dichotome Bewertung (Ja / Nein) vorgenommen wird.

Die Unterlagen sind ausschließlich zum persönlichen Gebrauch.

Diagnostik- und Testkuratorium der Föderation Deutscher Psychologinnenvereinigungen (2018). TBS-DTK. Testbeurteilungssystem des Diagnostik- und Testkuratoriums der Föderation Deutscher Psychologinnenvereinigungen. Revidierte Fassung vom 3. Januar 2018. Psychologische Rundschau (2018), 69, 109-148.

Datei: ...Kersting\TBS_TK_Präsentation_Stand_Sep_2019.ppt

TBS-DTK Rezensionen

→ TBS-DTK Rezensionen werden sowohl in der Psychologischen Rundschau als auch in Report Psychologie veröffentlicht. Darüber hinaus ist die Rezension über die ZPID online abrufbar: Freier Download der Rezensionen unter: <http://www.zpid.de/Testkuratorium>

→ Es ist der Wunsch des Diagnostik- und Testkuratoriums (DTK), dass Testrezensionen in allen deutschsprachigen Fachzeitschriften dem TBS-DTK System folgen. In diesem Fall erfolgt die Betreuung der Rezension in Kooperation mit dem DTK. Beispielsweise wurde die Rezension zum MSCEIT in der Zeitschrift für Arbeits- und Organisationspsychologie publiziert. Sollten Sie von einer Fachzeitschrift um eine Testrezension gebeten werden, so fragen Sie bitte nach der Möglichkeit, diese als TBS-DTK Rezension zu gestalten!

Bewertungsformat des neuen Systems zur Testbeurteilung



Beispiel: TBS-DTK Bewertung zu: Liepmann, D., Beauducel, A., Brocke, B. & Amthauer, R. (2007). Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R). 2., erweiterte und überarbeitete Auflage. Göttingen: Hogrefe.

TBS-TK R e z e n s i o n	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		●		

Schmidt-Atzert, L. & Rauch, W. (2008). TBS-TK Rezension: „Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R)“. *Report Psychologie*, 33, 303-304.

Schmidt-Atzert, L. & Rauch, W. (2008). TBS-TK Rezension: „Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R)“. *Psychologische Rundschau*, 59, 186.

◆ Diagnostik- u. Testkuratorium ◆ Testbeurteilungssystem des Testkuratoriums (TBS-DTK) ◆ Stand: Sep. 2019 ◆

Sachstand TBS-DTK Rezensionen, Stand August 2018

Download: <http://www.bdp-verband.org/psychologie/testrezensionen/index.html>

client
logo

Testname	Allg. Info	Objektivität	Zuverlässigkeit	Validität
DISK-Gitter	++	++	++	++
DERET	++	++	+	++
FPD	++	++	+	++
BIP-6 F	++	++	+	+
FPI-R	++	++	+	+
IBES	++	++	+	+
IEA	+	++	++	+
IST2000-R	++	++	+	+
OLMT	++	++	+	+
OPQ32	++	++	+	+
BDI-II	+	+	+	++
CBL/6-18R, TRF/6-18R, YSR/ 11-18R	++	+	+	+
FAKT-II	+	++	+	+
INSBAT	+	+	++	+
KABC	+	++	+	+
KABC-II	++	+	+	+
NEO-PI-R	++	+	+	+
WISC-IV	++	+	++	-
CFT 20-R	+	+	+	+
EXPLORIX	++	+	+	-
WIE	++	+	+	-
AVEM	+	+	+	-
AID 3	+	+	+	-
d2-R	+	+	+	-
ET_6-6-R	++	+	-	-

Testname	Allg. Info	Objektivität	Zuverlässigkeit	Validität
FEEL-KJ	+	+	+	-
Klasse 4	++	+	-	-
MSCEIT	+	+	+	-
SURT	+	+	-	+
DCS-II	+	+	-	-
SCL-90-R	+	+	-	-
TEA-CH	+	+	-	-
VVKI	+	-	+	-
ET_6-6 (1.Auflage)	+	-	-	-
SKEI	+	-	-	-
Frakis	+	-	-	--
Persolog	-	-	+	--
PPI-R	-	-	-	-
GPOP	-	-	-	--
Familie in Tieren	-	--	--	--
InCheck	»InCheck« erfüllt die durch DIN 33430 festgelegten Anforderungen bezüglich Information und Dokumentation nicht.			

... erfüllt die entsprechenden TBS-TK Anforderungen

- -- = nicht
- - = teilweise
- + = weitgehend
- ++ = voll

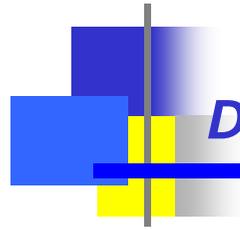
Hinweis:
Die Verfahren sind nach ihrer Bewertung geordnet dargestellt. Bei gleicher Bewertung erfolgt die Anordnung alphabetisch.

Stand Aug. 2018: 41 Rezensionen, weitere in Vorbereitung / in Druck

TBS-DTK System: Qualitätsmerkmale

- Das Diagnostik- und Testkuratorium (DTK) wählt die zu rezensierenden Tests aus. Im Falle von „confidential tests“ sichert das DTK den Testanbietern die Vertraulichkeit bestimmter Informationen zu. Diese Zusicherung ist notwendig, weil sich einige Testanbieter aktuell einer Rezension ihrer Tests mit der Begründung entziehen, sie könnten ihr Kapital (die Testentwicklung) und / oder vertrauliche Kundenbeziehungen (etwa in Form von Angaben zu den Normgruppen usw.) nicht bekannt geben.
- Für jeden zu beurteilenden Test beauftragt das Diagnostik- und Testkuratorium (DTK) zwei Reviewer. Das DTK bürgt für die einschlägige Expertise und Unabhängigkeit sowie Unvoreingenommenheit der Reviewer. Die beiden Reviewer arbeiten zunächst in Unkenntnis und unabhängig voneinander ihre Beurteilungen für die vorgegebenen Beurteilungsaspekte aus, fixieren diese Urteile auf der vorgegebenen Beurteilungsskala und übermitteln sie an das DTK. Erst wenn die Urteile beider Rezensenten vorliegen, werden die beiden Reviewer aufgefordert, im zweiten Schritt gemeinsam zu arbeiten und Konsensurteile zu finden sowie eine endgültige Rezension zu verfassen. Jeder Rezensent entscheidet für sich, ob er bei der Publikation der Rezension namentlich genannt werden will oder ob er anonym bleiben will
- Die von beiden Reviewern gemeinsam erstellte Rezension wird vom TK in anonymisierter Form an die Testautoren geschickt, um ihnen Gelegenheit einzuräumen, innerhalb einer gesetzten Frist Stellung zu beziehen. Im Falle einer solchen Stellungnahme entscheidet das DTK, ob es die beiden Rezensenten bittet, aufgrund der Stellungnahme eine Modifikation der Testrezension vorzunehmen.

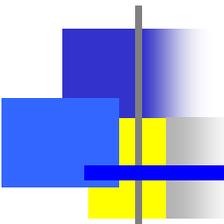
Diagnostik- und Testkuratorium. (2018). TBS-DTK. Testbeurteilungssystem des Diagnostik- und Testkuratoriums der Föderation Deutscher Psychologinnenvereinigungen. Revidierte Fassung vom 03. Jan. 2018. *Psychologische Rundschau*, 69, 109-116.



Die Beurteilungsrichtlinien – Beispiele

client
logo

Auf den nachfolgenden Folien finden Sie Beispiele für Beurteilungsaspekte.



Bewertung des Informationsgehalts der Verfahrens-hinweise - Beispiele

client
logo

DIN SCREEN A4 (V2) BIS A11, B3 BIS B13

- Zugänglichkeit von Informationen / Informationspolitik
- Informationsgehalt der Darstellung von empirischen Untersuchungen
- Durchführung und Durchführungsvoraussetzungen (z. B. Qualifikation der Testleiter(innen), relevante ethische und rechtliche Aspekte des vorgesehenen Testeinsatzes).
- Auswertung und Interpretation (Vorgehen bei der Auswertung [ggf. Schablonen, Auswertungsprogramme], Vergabe von Punktwerten für eine Antwort, Berechnung von Skalen und / oder Gesamtwerten, gegebenenfalls Umrechnung in Normwerte, Interpretationshilfen wie Cut-off-Werte, Normen, Vertrauensgrenzen, kritische Differenzen)

Theoretische Grundlagen als Ausgangspunkt der Testkonstruktion - Beispiele

client
logo

- Schließt der Test an eine bestehende Theorie an oder entwickeln die Testautor(inn)en eine eigene Theorie?
- Wird diese Theorie ausreichend beschrieben? Wird das Konstrukt hinlänglich beschrieben?
- Wird deutlich, was und was nicht zu dem zu messenden Bereich gerechnet wird?
- Wird deutlich, ob ein beliebiges Item zum Test gehören könnte oder nicht?
- Werden das oder die zu messenden Konstrukt(e) auf solche Weise (z. B. mit Hilfe von Facetten-Analyse) analysiert, sodass deutlich wird, welche Aspekte innerhalb des Konstrukts oder der Konstrukte unterschieden werden können?

Objektivität - Beispiele

Hinsichtlich der *Durchführungsobjektivität* soll insbesondere auf folgende Punkte geachtet werden:

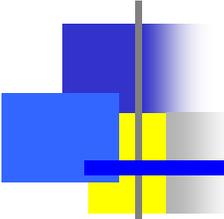
- Die Instruktionen für die Testleiter(innen) müssen
 - möglichst wörtlich vorschreiben, was der Testleiter sagen soll und was nicht (so ist z. B. die Empfehlung "der Testleiter erklärt das Ziel des Tests" als mangelhaft zu werten)
 - genau angeben, welche Handlungen der Testleiter konkret zu verrichten hat (z. B. das Testmaterial in einer bestimmten Art ordnen)
 - genau ausführen, wie auf Fragen eingegangen werden muss (es können z. B. Standardtexte gegeben werden für Antworten auf häufig vorkommende Fragen)
- Die Instruktionen für die getesteten Personen sollten Beispiel- und Übungssitem enthalten sowie Informationen über die Art, wie die Reaktionen (Antworten) zu geben sind.

Reliabilität - Beispiele

Zuverlässigkeit (Reliabilität / Messgenauigkeit)

DIN SCREEN Aussagen B22 bis B26

- Wird die Reliabilität anhand von Daten der (Sub-)Population der Zielgruppe geschätzt, für die der Test laut diagnostischer Zielsetzung eingesetzt werden soll?
- Prüfung, ob sehr hohe interne Konsistenz auf nahezu identisch gestaltete Items zurückzuführen sind
- Bei Tests mit Speed-Komponente: keine Homogenitätsmaße

A decorative graphic consisting of overlapping blue and yellow squares with a vertical grey bar and a horizontal blue bar.

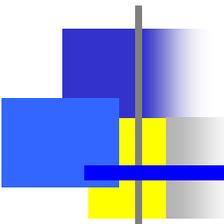
Gültigkeit (Validität) - Beispiele

DIN SCREEN AUSSAGEN B27 BIS B54

Grundsätzlich geht es nicht um die Validität eines Tests, sondern um die Validität der Interpretation der Ergebnisse, die mit dem Test gewonnen werden.

Bei der Bewertung der Validität gilt:

Die Rezensent(inn)en berücksichtigen bei ihrem Urteil über die Angaben zur Validität des Tests, dass die Validitätswerte in Abhängigkeit von den untersuchten Gruppen (eine besondere Bedeutung kommt der Homogenität der Gruppe hinsichtlich des gemessenen Konstrukts zu) und in Abhängigkeit vom Untersuchungsdesign variieren.



Normierung - Beispiele

Normierung (Eichung) DIN SCREEN Aussagen B16 (V6) bis B19

- Sofern die diagnostische Zielstellung bei der Interpretation der Testwerte Normen (Eichtabellen) nötig macht, prüfen die Rezensent(inn)en, ob tatsächlich für jedes genannte diagnostische Ziel Normen (Eichtabellen) zur Verfügung stehen.
- Die Eichstichprobe muss repräsentativ sein für jede angestrebte (Sub-)Population. Die Rezensent(inn)en prüfen, ob die Repräsentativität für die Zielgruppen nachvollziehbar dargestellt ist. Dabei geht es um eine angemessene Beschreibung sowohl der Population als auch der Art der Stichprobenziehung oder Datensammlung.