



Online-Self-Assessments zur Studienfachwahl

Hinweis: Die Nutzer(innen) dürfen die Inhalte nur zum Zwecke der wissenschaftlichen Forschung ansehen, drucken, kopieren, herunterladen, sowie für Text- und Datamingen verwenden. Die Inhalte dürfen weder ganz noch teilweise wörtlich (wieder)veröffentlicht oder für kommerzielle Zwecke verwendet werden. Die Nutzer müssen sicherstellen, dass die Urheberpersönlichkeitsrechte des Autors sowie gegebenenfalls bestehende Rechte Dritter an den Inhalten oder Teilen der Inhalte nicht verletzt werden.

Dies ist das akzeptierte Manuskript (Author's Accepted Manuscript) des folgenden Beitrages:

Kersting, M. & Marquardt, C. (2022). Tests, Fragebogen & Co – Anforderungen an die Qualität der im OSA eingesetzten diagnostischen Verfahren auf welche Qualitätskriterien es für OSA ankommt. In: Weiß, Susanne & Stoll, Gundula (Hrsg.), Online Self-Assessments zur Studienfachwahl (S.107-129), Berlin: Springer, vervielfältigt mit Genehmigung von: Springer Verlag.



Gundula Stoll · Susanne Weis
(Hrsg.)

Online-Self-Assessments zur Studienfachwahl

Entwicklung – Konzepte –
Qualitätsstandards



Inhaltsverzeichnis

1	Ein Rahmenmodell zu Entwicklung, Wirkmechanismen und Bewertung von Online Self-Assessments	1
	Gundula Stoll, Susanne Weis, André Beauducel, Aline Becker, Constanze Brazel, Ralph Delzepich, Sylvia Dietrich, Melanie Ewert, Gabriela Gniewosz, Martin Kersting, Manuel Lange, Julia E.M. Leiner, Christian Marquardt, Tuulia M. Ortner, Sergei Pachtchenko, Malte Persike, Pascale Stephanie Petri, Andreas Pfeiffer, Neele Schütte, Manfred Schmitt, Marlene Schütz, Frank M. Spinath, Sarah Teige-Mocigemba, Tatjana Wachtel, Vera Weingardt und Katharina Zay	
2	Ziele von Online Self-Assessments: die Bedeutung einer frühzeitigen Auftragsklärung	25
	Gabriela Gniewosz, Julia E.M. Leiner und Tuulia M. Ortner	
3	Identifikation kriterienbezogener Anforderungen im Studium: Die Anforderungsanalyse als Ausgangspunkt und Qualitätsmerkmal für Online Self-Assessments	49
	Marlene Schütz und Vera Weingardt	
4	Mit OSA erfasste Personenmerkmale und deren Interaktion mit OSA-Rückmeldungen und Informationselementen	75
	Pascale Stephanie Petri, Neele Schütte und André Beauducel	
5	Tests, Fragebogen & Co – worauf es bei diagnostischen Verfahren für OSA ankommt	107
	Martin Kersting und Christian Marquardt	
6	Informationselemente in Online Self-Assessments	131
	Aline Becker, Constanze Brazel und Frank M. Spinath	
7	Die Rückmeldung als zentrales Element von Online Self-Assessments: Funktionen, Gütekriterien und Gestaltungsempfehlungen	153
	Malte Persike, Katharina Zay, Melanie Ewert, Tatjana Wachtel und Sarah Teige-Mocigemba	



8 Computergestützte Umsetzung von Online Self-Assessments 183
Sergei Pachtchenko, Andreas Pfeiffer, Ralph Delzepich
und Manuel Lange

9 Evaluation von Online Self-Assessments 203
Susanne Weis, Gundula Stoll, Sylvia Dietrich und Tatjana Wachtel

10 Qualitätsstandards 231
Pascale Stephanie Petri, Christian Marquardt, Susanne Weis
und Gundula Stoll

UNCORRECTED PROOF



1 Tests, Fragebogen & Co – worauf es 2 bei diagnostischen Verfahren für 3 OSA ankommt

4 Martin Kersting und Christian Marquardt

Hinweis: Die Nutzer(innen) dürfen die Inhalte nur zum Zwecke der wissenschaftlichen Forschung ansehen, drucken, kopieren, herunterladen, sowie für Text- und Dataming verwenden. Die Inhalte dürfen weder ganz noch teilweise wörtlich (wieder)veröffentlicht oder für kommerzielle Zwecke verwendet werden. Die Nutzer(innen) müssen sicherstellen, dass die Urheberpersönlichkeitsrechte des Autors sowie gegebenenfalls bestehende Rechte Dritter an den Inhalten oder Teilen der Inhalte nicht verletzt werden.

Dies ist das akzeptierte Manuskript (Author's Accepted Manuscript) des folgenden Beitrages: Kersting, M. & Marquardt, C. (2021). Tests, Fragebogen & Co – Anforderungen an die Qualität der im OSA eingesetzten diagnostischen Verfahren auf welche Qualitätskriterien es für OSA ankommt. In: Weiß, Susanne & Stoll, Gundula (Hrsg.), Online Self-Assessments zur Studienfachwahl (S.107-129), Berlin: Springer, vervielfältigt mit Genehmigung von: Springer Verlag.

5 Leitfragen:

- 6 1. Welche Arten von diagnostischen Verfahren gibt es?
- 7 2. Nach welchen Kriterien sollte man diagnostische Verfahren für OSA aus-
8 wählen und beurteilen?
- 9 3. Was gilt es bei der Neuentwicklung von diagnostischen Verfahren für
10 OSA zu beachten?

11 5.1 Überblick

12 „Zuverlässige Informationen sind unbedingt nötig für das Gelingen eines Unter-
13nehmens“. Dieses Zitat wird Christoph Kolumbus zugeschrieben, inhaltlich trifft
14 es auch auf Studieninteressierte zu, die sich auf eine Entdeckungsreise durch die
15 Angebote der (Fach-)Hochschulen begeben. Im Rahmen von OSA werden den
16 Studieninteressierten verschiedene Verfahren angeboten, mit denen zuverlässige
17 und gültige Informationen über die Passung ihrer Fähigkeiten, studienbezogenen
18 Persönlichkeitsmerkmale und Interessen zu den Anforderungen (z. B. eines spezi-
19fischen Studiengangs, siehe Kap. 3) gewonnen werden. Ziel des vorliegenden
20 Kapitels ist es, einen – um konkrete Beispiele angereicherten – Überblick über die

A1 M. Kersting (✉)
A2 Justus-Liebig-Universität Gießen, Psychologische Diagnostik, Gießen, Deutschland
A3 E-Mail: martin.kersting@psychol.uni-giessen.de

A4 PresentC. Marquardt
A5 Psychologisches Institut, Universität Heidelberg, Heidelberg, Deutschland
A6 E-Mail: christian.marquardt@psychologie.uni-heidelberg.de

A7 C. Marquardt
A8 Center für Lehr- und Lernservices (CLS), RWTH Aachen University, Aachen, Deutschland



21 in OSA eingesetzten diagnostischen Verfahren zu geben. Weitere Abschnitte gehen
22 auf die Entwicklung und die Qualität diagnostischer Verfahren ein.

23 5.2 Eignungsdiagnostische Verfahren nach DIN 33430

24 Die einzelnen diagnostischen Verfahren zur Studieneignungsfeststellung sind viel-
25 fältig, lassen sich aber letztendlich auf einige „Grundtypen“ von Verfahren zurück-
26 führen. Kersting (2011) hat die diagnostischen Verfahren mit dem sogenannten
27 CUBE System in fünf Kategorien eingeteilt. Diese Einteilung wurde für die DIN
28 33430 (2016, S. 12) übernommen. Man kann demnach die folgenden fünf Kate-
29 gorien von diagnostischen Verfahren unterscheiden:

- 30 „a) Dokumentenanalyse (z. B. die Analyse und Interpretation von Hochschul-, / Schul-
31 und Arbeitszeugnissen, dem Lebenslauf, von Beurteilungen, der Ergebnisse von Internet-
32 recherchen);
- 33 b) Direkte mündliche Befragungen (z. B. Interview mit Kandidaten; Gespräch mit
34 einem Referenzgeber);
- 35 c) Verfahren zur Verhaltensbeobachtung und -beurteilung (z. B. Rollenspiele,
36 Gruppendiskussionen, Präsentationsübungen, Arbeitsproben);
- 37 d) Messtheoretisch fundierte Fragebogen (z. B. Persönlichkeitsfragebogen, Interessen-
38 fragebogen);
- 39 e) Messtheoretisch fundierte Tests (z. B. Intelligenztests, Wissenstests, Situational
40 Judgement Tests).“

41 Die Einteilung der Verfahren in Kategorien ist deshalb bedeutsam, weil für jede
42 Kategorie bestimmte Qualitätsanforderungen gelten. Diese werden in der DIN
43 33430 (2016) wiedergegeben und im Buch zur DIN 33430 (Diagnostik- und Test-
44 kuratorium, 2018a) erläutert. Gemäß DIN 33430 kann die Verfahrenswahl erst
45 nach einer Anforderungsanalyse (Kap. 3) erfolgen, da zunächst geklärt werden
46 muss, welches Merkmal überhaupt erfasst werden soll, bevor ein dafür geeignetes
47 Verfahren ausgewählt werden kann.

48 Grundsätzlich können alle Verfahrenskategorien auch im Kontext von OSA
49 in Form einer „an sich selbst“ durchgeführten Diagnostik vorkommen. Da OSA-
50 Angebote im Normalfall aber als unbeaufsichtigte online-Tests durchgeführt
51 werden und skalierbar sein sollten, also ohne weitere Investitionen auch auf
52 größere Zahlen von Testpersonen anwendbar sein sollten, sind Dokumentenana-
53 lysen, direkte mündliche Befragungen sowie Verfahren zur Verhaltensbeobachtung
54 im Kontext von OSA seltener. Denn hier müssen in der Regel Menschen zur
55 Interpretation der Dokumente bzw. zur Durchführung von Interviews und
56 Beobachtungen eingesetzt werden, was derlei Verfahren sehr personalaufwendig
57 macht. Dies kann sich ändern, falls sich in Zukunft technische Lösungen für diese
58 Aufgaben etablieren (z. B. Programme zur automatischen Textanalyse, Chatbots
59 oder Bild- / Tonverarbeitungsgeräte). Wir gehen im Folgenden insbesondere auf
60 die in OSA weit verbreiteten messtheoretisch fundierten Tests und Fragebögen
61 ein, thematisieren der Vollständigkeit halber aber alle Verfahrenskategorien.



62 5.2.1 Dokumentenanalyse

63 Zur Dokumentenanalyse zählt die Interpretation von Schulzeugnissen und des
64 Lebenslaufs. Eine solche Analyse kommt im Rahmen von OSA so gut wie nicht
65 vor. Dies liegt u. a. daran, dass die Ausgangsdaten sehr individuell sind und sich
66 nicht ohne weiteres so standardisiert erfassen lassen, dass eine automatisiert
67 erstellte Rückmeldung möglich ist. Dabei sind Schulnoten kein schlechter Start-
68 punkt für eine reflektierte Studienfachwahl. Schulnoten können in vielerlei Hin-
69 sicht kritisiert werden, sind aber dennoch aussagekräftig. Eine Metaanalyse
70 von Trapmann et al. (2007) kommt auf der Basis der Ergebnisse aus 42 Studien
71 mit insgesamt 45.860 Personen auf einen durchschnittlichen (und messfehler-
72 korrigierten) Zusammenhang von $r = .53$ zur Vorhersage des Studienerfolgs durch
73 die Schulnote (Durchschnittsnote).

74 5.2.2 Direkte mündliche Befragungen

75 Unter den Begriff „direkte mündliche Befragungen“ fasst die DIN 33430 u. a. das
76 klassische Eignungsinterview, eines der meist genutzten Verfahren im Bereich der
77 Personalauswahl. Interviews sind bei den Personalverantwortlichen beliebt und
78 werden von den Testpersonen weithin akzeptiert. Dies liegt u. a. daran, dass die
79 Situationskontrolle als hoch wahrgenommen wird und sich zudem die Möglich-
80 keit bietet, sich möglichst positiv zu präsentieren (Westhoff & Strobel, 2005).
81 Interviews werden aus Aufwandsgründen im Rahmen von OSA nicht praktiziert,
82 zumindest ist uns aktuell kein OSA bekannt, welches ein Interview vorsieht.

83 5.2.3 Verfahren zur Verhaltensbeobachtung und 84 Verhaltensbeurteilung

85 Aufgrund des hohen personellen Aufwands kommen Verfahren zur Verhaltens-
86 beobachtung und -beurteilung wie Rollenspiele und Gruppendiskussionen im
87 Rahmen von OSA kaum vor. Eine Ausnahme bildet das OSA für das Lehramts-
88 studium an der Universität Salzburg, welches mit einem asynchronen Video-
89 Rollenspiel arbeitet (Praxisbeispiel: [Asynchrones Video-Rollenspiel im OSA des Hochschulverbundes Cluster Mitte](#)).

91 Da bei diesen Verfahren Beobachtung und Bewertung nicht voll-standardisiert
92 vorgenommen werden, sondern die beteiligten Menschen einen hohen Verhaltens-
93 und Entscheidungsspielraum haben, entstehen im Vergleich zu standardisierten
94 Testverfahren viele Verzerrungsrisiken – sowohl in der Gewinnung als auch
95 bei der Auswertung und Interpretation von Daten. Personen, die mit den
96 Beobachtungen und Bewertungen betraut sind, sollten daher im Vorfeld ein
97 Training durchlaufen (Höft & Kersting, 2018). Hohe Anforderungen bei der
98 Konstruktion der Verfahren bestehen insbesondere darin, die Eignungsmerkmale



99 so verhaltensnah zu operationalisieren, dass verschiedene Personen zu einer
100 gleichsinnigen Einschätzung kommen.

101 **Praxisbeispiel: Asynchrones Video-Rollenspiel im OSA des Hochschulverbundes** 102 **Cluster Mitte**

103 Beim OSA für das Lehramtsstudium an der Universität Salzburg bekommen
104 Studieninteressierte in einem Begrüßungsvideo die Aufgabe gestellt, sich in ein
105 Szenario als Lehrkraft in zehn Jahren zu versetzen. Für die inhaltliche Vorbereitung
106 der Aufgabe steht beliebig viel Vorbereitungszeit zur Verfügung. Die Aufgabe
107 selbst muss in einem Video von maximal zwei Minuten Länge beantwortet und
108 hochgeladen werden. Technisch wird dies entweder per Webcam direkt im Web-
109 browser oder durch Upload eines z. B. vom Smartphone aufgenommenen Videos
110 ermöglicht. Misslungene Aufnahmen, etwa durch äußere Störungen während des
111 Aufnehmens, können durch Neuaufnahme korrigiert werden.

112 Da dieser Teil des OSA auch ausgelassen werden kann, wird die freiwillige
113 Einlassung auf diese unbekannt Situation zunächst als Indikator für Auf-
114 geschlossenheit im Hinblick auf neue, soziale Situationen gewertet. Darüber
115 hinaus wird beurteilt, inwieweit es den Studieninteressierten gelungen ist, die
116 vorgegebene Frage zu beantworten, insbesondere wie sie dabei argumentieren
117 und in wieweit sie es verstehen, klar und verständlich zu formulieren (inhalt-
118 licher und sprachlicher Aspekt). Die Bewertungskriterien sind transparent und
119 werden den Studieninteressierten vorab im Rahmen der Aufgabeninstruktion
120 mitgeteilt. Die Auswertung der hochgeladenen Videos findet durch studentische
121 Hilfskräfte der Universität Salzburg statt. Die Hilfskräfte müssen die Aus-
122 wertung nicht zur gleichen Zeit vornehmen, zu der die Testpersonen die Videos
123 erstellen – daher die Bezeichnung „asynchrone“ Rollenspiele. Eine Rück-
124 meldung zum Video erhalten die Studieninteressierten in der Regel binnen
125 weniger Tage. In der Verfahrensentwicklung wurde ein Beobachtermanual zur
126 Videoanalyse erstellt, welches auch die Grundlage für ausführliche Rating-
127 Schulungen bildet. Scherndl et al. (2019) berichten, dass die Akzeptanz des
128 asynchronen Rollenspiels seitens der Studieninteressierten hoch sei. ◀

129 **5.2.4 Messtheoretisch fundierte Fragebogen**

130 Schriftliche Befragungen sind weit verbreitet. Sind Frage- und Antwortmöglich-
131 keiten dabei vollständig oder weitgehend festgelegt, spricht man von einem
132 standardisierten Fragebogenverfahren (Eid et al., 2014).

133 In OSA werden standardisierte Fragebogenverfahren häufig dann eingesetzt,
134 wenn bestimmte Personeneigenschaften wie z. B. studienbezogene Persön-
135 lichkeitseigenschaften oder Interessen beurteilt und auf ihre Passung zu einem
136 bestimmten Studiengang verglichen werden sollen. Die Fragen werden meist in
137 Aussageform präsentiert. In den meisten Fällen sollen die Testpersonen auf einer
138 mehrstufigen Antwortskala diesen Aussagen zustimmen bzw. diese Aussagen



139 ablehnen (sog. Likert-Skala). Für die einzelnen Antwortstufen werden in der Regel
140 Punkte vergeben (z. B. *trifft völlig zu*: 5 Punkte bis *trifft überhaupt nicht zu*: 1
141 Punkt), die dann pro Skala zu einem Gesamtwert aufsummiert werden.

142 ► **Der Eigenschaftsansatz in der psychologischen Diagnostik**

143 Mit einem OSA möchte man Eignungs- und Passungsmerkmale von
144 Studieninteressierten erfassen (Kap. 4, z. B. Intelligenz, Interessen). Nach
145 dem Eigenschaftsansatz der psychologischen Diagnostik fasst man
146 diejenigen Merkmale als Eigenschaften einer Person auf, die man nicht direkt
147 beobachten kann. Man spricht von so genannten „Konstrukten“. Um diese
148 latenten Eigenschaften messbar zu machen (auch: zu operationalisieren),
149 bedarf es so genannter manifester, d. h. beobachtbarer Indikatoren in Form
150 von Antworten auf einzelne Testaufgaben oder Fragen in einem Fragebogen.

151 Zwingende Voraussetzung für das Schließen von den Antworten auf
152 einzelne Testaufgaben oder Fragen in einem Fragebogen auf Gesamt-
153 aussagen über das zugrunde liegende Eignungsmerkmal einer Person ist
154 ein theoretisches Messmodell. Das Messmodell bestimmt, in welcher
155 Beziehung die manifesten Indikatoren zu dem interessierenden latenten
156 Konstrukt stehen. Es sagt somit aus, welcher Indikator zu welchem
157 latenten Merkmal „gehört“. Dieses Modell muss theoretisch verankert
158 und empirisch überprüft sein, der Fragebogen oder Test selbst muss test-
159 theoretischen Gütekriterien genügen (Abschn. 5.3.2). Erst dann lässt sich
160 von einem messtheoretisch fundierten Fragebogen oder Test als geeigneter
161 Operationalisierung für ein latentes Personenmerkmal sprechen.

162 So kann beispielsweise in einem Test, der logisches Denken erfassen
163 soll, die korrekte Fortsetzung einer Zahlenreihe als Indikator für die
164 Fähigkeit logisch zu denken herangezogen werden.

165 Alternativ kann in einem Persönlichkeitsfragebogen die Beurteilung
166 bestimmter Aussagen durch eine Testperson als Indikator für bestimmte
167 Persönlichkeitseigenschaften herangezogen werden. In standardisierten
168 Fragebogen gibt die Testperson z. B. auf einer Skala von 1 (stimme über-
169 haupt nicht zu) bis 5 (stimme völlig zu) an, wie sehr sie einer bestimmten
170 Aussage (z. B. „Ich gehe gerne auf Partys.“) zustimmt. Das Ausmaß
171 der Zustimmung gilt dann als Indikator für das der Aussage zugrunde
172 liegende latente Konstrukt (hier: Extraversion).

173 Ein Beispiel für standardisierte Fragebogen sind Persönlichkeitsfragebogen (zum
174 Begriff und zur Bedeutung der Persönlichkeit siehe Kap. 4). In den OSA der Uni-
175 versität Koblenz-Landau etwa wird mit Hilfe von 80 Fragen ein Profil der studien-
176 bezogenen Persönlichkeit erstellt, welches dann mit dem durchschnittlichen Profil
177 von 200 erfolgreich Studierenden verglichen wird (normorientierte Interpretation,
178 siehe Abschn. 5.3.2; Kap. 7).



179 Ein generelles Problem von Fragebogenverfahren ist die vergleichsweise
180 hohe Anfälligkeit für Verfälschungen. Dies kann zum einen durch Unkenntnis
181 oder Fehlannahmen über die eigene Person geschehen („blinde Flecken“, Selbst-
182 täuschung), zum anderen aber auch durch bewusste Beschönigung der eigenen
183 Person (soziale Erwünschtheit, Selbstdarstellung). Das Problem der absichtlichen
184 Verfälschung stellt sich aber bei OSA weniger, da die Testpersonen kein Interesse
185 an einer Verfälschung der Ergebnisse hat. Auf diesen Umstand kann man in der
186 Instruktion zu einem OSA-Fragebogen hinweisen.

187 In OSA zur Studienorientierung werden standardisierte Fragebogen vor
188 allem auch für studien- oder berufsbezogene Interessentests eingesetzt. (Zum
189 Begriff und zur Bedeutung der Interessen siehe Kap. 4). Die Bedeutsamkeit
190 für die Studienorientierung liegt auf der Hand: Passen die persönlichen Vor-
191 lieben und Neigungen zu den Inhalts- und Tätigkeitsfeldern eines Studienganges
192 besonders gut, sollte sich dies auch auf den Erfolg eines Studiums auswirken.
193 Wie aber lassen sich verschiedene Interessenbereiche differenzieren? Fast alle
194 in OSA eingesetzten Fragebogen zu berufsbezogenen Interessen orientieren sich
195 am Interessenmodell nach Holland (1997), dem sogenannten RIASEC-Modell
196 (Kap. 4). In Fragebogen zum RIASEC-Modell müssen Studieninteressierte
197 beispielsweise einschätzen, wie gerne sie bestimmte Tätigkeiten selbst aus-
198 führen würden, die jeweils einem von sechs grundlegenden Interessenbereichen
199 zugeordnet werden können. Die resultierenden Interessensprofile können dann
200 mit zuvor festgelegten Soll-Profilen einzelner Studiengänge abgeglichen werden
201 (Anforderungsanalyse, Kap. 3).

202 Zusammenfassend kann festgehalten werden, dass messtheoretisch fundierte
203 Fragebogen vielfältig in OSA zum Einsatz kommen können und vermutlich
204 wertvolle Reflexionsprozesse bei den Studieninteressierten anstoßen. Die Vor-
205 hersagekraft der aus Persönlichkeitsfragebogen gewonnenen Werte über den wahr-
206 scheinlichen Erfolg in einem bestimmten Studiengang, also die Kriteriumsvalidität
207 (Abschn. 5.3.2), ist jedoch eher gering. Dies gilt insbesondere, wenn man den
208 Erfolg an Studiennoten oder Studiendauer misst. Bessere Validitäten können sich
209 für die Vorhersage nicht-kognitiver Kriterien wie der Studienzufriedenheit ergeben
210 (Kap. 4).

211 5.2.5 Messtheoretisch fundierte Tests

212 Messtheoretisch fundierte Tests sind Bestandteil vieler OSA. Sie ermöglichen
213 es den Studieninteressierten zu prüfen, ob ihre Fähigkeiten und ihr Wissen
214 (siehe Kap. 4) den Anforderungen eines Studiengangs gerecht werden. Zu den
215 messtheoretisch fundierten Tests zählen alle Verfahren mit einem vorab fest-
216 gelegten „richtig“/„falsch“ Schema. Dies sind in erster Linie Fähigkeits- und
217 Leistungstests, aber auch Wissenstests (z. B. fachspezifische Vorwissenstests oder
218 „Erwartungschecks“) und Konzentrationstests gehören in diese Kategorie.

219 Ein typisches Beispiel für einen im Rahmen eines OSA genutzten Fähig-
220 keitstest ist der Gießener kognitive Kompetenz Test, 66-Item Version, kurz



221 GkKT-66 (Ulfert et al., 2017; [Beispiel zur Erfassung kognitiver Fähigkeiten: Der](#)
222 [Gießener kognitive Kompetenz Test](#)). Der GkKT wurde an der Universität Gießen
223 entwickelt. Die Gießener Arbeitsgruppe ist Teil des Netzwerks Online Self-
224 Assessment. Sie stellt den GkKT auf Anfrage für die Nutzung in nicht-gewinn-
225 orientierten OSA mit einer freien Lizenz zur Verfügung.

226 **Beispiel zur Erfassung kognitiver Fähigkeiten: Der Gießener kognitive Kompetenz**
227 **Test (Ulfert et al., 2017)**

228 Der Gießener kognitive Kompetenz Test (GkKT-66) zielt auf die Erfassung des
229 logischen Denkens. Diese Fähigkeit wird mit 12 Aufgabengruppen und ins-
230 gesamt 66 Aufgaben erfasst: 22 verbale Aufgaben, 23 numerische Aufgaben
231 und 21 figurale Aufgaben. Die gesamte Testdurchführung dauert durchschnitt-
232 lich 28,5 min. Für die normbezogene Interpretation (siehe unten Abschnitt
233 „Normen“) liegt eine Eichstichprobe im Umfang von N= 803 Personen (62,9 %
234 Frauen) vor. Über die Qualität des Tests berichten wir weiter unten ([Praxisbei-
235 spiel: Validierung des GkKT für den Einsatz im OSA der Universität Gießen](#)),
236 eine ausführliche Darstellung des Tests geben Petri et al. (2019).

237 Abb. 5.1 und 5.2 geben Beispielaufgaben des Tests wieder. ◀

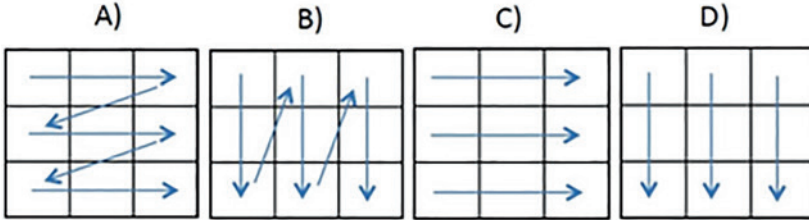
238 „Erwartungschecks“ thematisieren weit verbreitete „Fehlannahmen“ über ein
239 Studium zumeist in Form von Wissensfragen (für einen Überblick siehe Hasen-
240 berg & Stoll, 2015). Dazu werden gängige (Fehl-) Annahmen über einen Studien-
241 gang in Aussageform präsentiert. So könnte ein Item eines Erwartungschecks
242 beispielsweise lauten „Wenn man Chemie studiert, kann man in der vorlesungs-
243 freien Zeit arbeiten, um Geld zu verdienen.“ Die Studieninteressierten müssen
244 den Realitätsgehalt dieser Aussage mit „trifft zu“ / „trifft nicht zu“ einschätzen.
245 Die häufig unmittelbar erfolgende Auswertung zeigt ihnen dann, inwieweit
246 die Erwartung mit der Realität übereinstimmt. Im Beispiel würde eine Person,
247 die der Aussage zustimmt, nicht nur erfahren, dass ihre Antwort „falsch“ ist,
248 sondern sie würde entsprechende Informationen über das Studium erhalten. (In
249 unserem Beispiel z. B. die Information, dass die vorlesungsfreie Zeit häufig
250 mit Arbeiten im Labor verbracht wird usw.). Auf diese Weise können Studien-
251 interessierte individuell genau an den Stellen weitere Informationen einholen,
252 an denen sie tatsächlich auch Bedarf sehen, statt sich durch einen umfassenden
253 Informationstext „zu quälen“ und dabei die Lust zu verlieren. Insgesamt sollen
254 durch Erwartungschecks Fehl-Erwartungen minimiert und das Wissen über das
255 Studium maximiert werden. Es geht bei den Erwartungschecks also nicht vor-
256 rangig um eine „Diagnostik“, sondern darum, in interaktiver und somit anregender
257 Form über relevante Studieninhalte, Abläufe, Erwartungen oder spätere Berufsaus-
258 sichten aufzuklären. Insofern können Erwartungschecks auch als spezielle Form
259 von Informationselementen verstanden werden (vgl. Kap. 6).

260 Bei der Interpretation der Daten von Wissens- und insbesondere Leistungs-
261 tests ist zu beachten, dass das Ergebnis sowohl von der Fähigkeit und dem Wissen,
262 als auch von der Anstrengung und somit von der Motivation der Testpersonen



Bei den folgenden Aufgaben geht es darum, das fehlende Feld im Zahlengitter durch die richtige Zahl zu ergänzen. Ihre Aufgabe besteht darin, die Regel zu erkennen, nach der das Zahlengitter aufgebaut ist und die fehlende Zahl mit Hilfe dieser Regel einzusetzen. Bitte schreiben Sie Ihre Antworten in das unten stehende Freitextfeld.

Die Anordnung der Zahlen kann nach den folgenden 4 Prinzipien erfolgen:



Beispielaufgabe:

2	8	14
4	10	16
6	12	?

Lösung: Dieses Zahlengitter funktioniert nach dem Prinzip B; die richtige Lösung lautet „18“, da immer „+2“ gerechnet wird.

2	8	14
4	10	16
6	12	18

Abb. 5.1 Beispielaufgabe „Zahlengitter“ aus dem GkKT-66

263 abhängt. Eine geringe Leistung kann somit sowohl auf eine gering ausgeprägte
264 Fähigkeit als auch auf mangelnde Anstrengungsbereitschaft zurückgeführt
265 werden. Dies ist insbesondere bei der Interpretation der Daten aus Leistungstests

Im folgenden Aufgabentyp geht es darum, den Weg vom Start zum Ziel zu finden. Die Wegstücke können durch eine Links- oder Rechtsdrehung zu einem Weg verbunden werden. Überlegen Sie, wie die jeweiligen Wegstücke *gedreht* werden müssen, damit Sie vom Startpunkt einen *durchgängigen Weg* zum Zielpunkt finden. Jedes Wegstück muss in den Weg mit eingebaut werden. Sie können die Wegstücke um *90 Grad nach links* (gegen den Uhrzeigersinn) *oder rechts* (mit dem Uhrzeigersinn) *drehen oder in ihrer Position lassen*.

Bitte kreuzen Sie bei den folgenden Aufgaben an, in welche Richtung das jeweilige Wegstück gedreht werden soll.

Wie müssen die Wegstücke A, B, C, und D gedreht werden, um vom Start zum Ziel zu kommen?

Beispielaufgabe

Drehungen

Lösung

<input type="checkbox"/> A0	<input type="checkbox"/> A1	<input checked="" type="checkbox"/> A2
<input checked="" type="checkbox"/> B0	<input type="checkbox"/> B1	<input type="checkbox"/> B2
<input type="checkbox"/> C0	<input checked="" type="checkbox"/> C1	<input type="checkbox"/> C2
<input type="checkbox"/> D0	<input type="checkbox"/> D1	<input checked="" type="checkbox"/> D2

Die richtige Antwort lautet: **A2** (Drehung nach rechts), **B0** (keine Drehung), **C1** (Drehung nach links), **D2** (Drehung nach rechts)

Abb. 5.2 Beispielaufgabe „Labyrinth“ aus dem GkKT-66

266 im Kontext von OSA zu beachten, da bei einer Testung ohne Aufsicht die Gefahr
267 besteht, dass die Testpersonen die Testung nicht „ernst nehmen“. Entsprechend
268 sollte man bereits in der Instruktion und in der anschließenden Rückmeldung
269 zu solchen Tests darauf verweisen, wie wichtig Motivation und eine ungestörte
270 Arbeitsumgebung für die Interpretierbarkeit der Ergebnisse sind.

271 Von besonderer Bedeutung für OSA ist die Akzeptanz (Abschn. 5.3.2).
272 Leistungstests sind besonders aussagekräftig (valide). Dennoch scheuen viele,
273 die für die Gestaltung von OSA verantwortlich sind, vor dem Einsatz von
274 Leistungstests zurück, weil sie Akzeptanzprobleme und damit verbundene Ein-
275 brüche in der Nutzung befürchten. In der Tat ist die Bearbeitung von Leistungs-
276 tests anstrengend. Außerdem merken die Studieninteressierten selbst – anders als
277 bei Persönlichkeitsfragebogen und Interessentests – wenn sie den Anforderungen
278 nicht gewachsen sind. Es bestehen daher Befürchtungen, dass sich der Einsatz
279 von Leistungstests im Rahmen von OSA ungünstig auf die Akzeptanz der OSA
280 auswirkt. Die Sorgen sind aber unbegründet: Metaanalytisch konnten Anderson
281 et al. (2010) beispielsweise zeigen, dass Intelligenztests in Bezug auf die Gesamt-



282 akzeptanz einen mittleren Platz (nominell vor den Persönlichkeitsfragebogen)
283 erzielten. Intelligenztests belegten sogar den ersten Platz, was die Wahrung der
284 Privatsphäre angeht. Die Akzeptanz für Leistungstests im Kontext von OSA dürfte
285 noch höher sein, da Studieninteressierte in der Regel freiwillig und anonym an der
286 Testung teilnehmen.

287 **5.3 Auswahl und Konstruktion diagnostischer Verfahren**

288 Wenn man ein OSA etablieren möchte, muss man zunächst festlegen, welche
289 Ziele man mit dem OSA verfolgt (Kap. 2), welche Anforderungen der infrage
290 stehende Studiengang stellt (Kap. 3) und wie das OSA zur Zielerreichung beiträgt.
291 Die Anforderungen des Studienfachs werden in psychologische Konstrukte über-
292 setzt. Für diejenigen Bereiche, für die man eine Selbst-Diagnostik vorsehen will,
293 benötigt man diagnostische Verfahren, die geeignet sind, Indikatoren für diese
294 Konstrukte zu liefern. Die informativen Anteile eines OSA (Kap. 6) stellen keine
295 diagnostischen Verfahren dar.

296 **5.3.1 make or buy?**

297 Wie kommt man zu den diagnostischen Verfahren, die man im Rahmen eines OSA
298 einsetzen will? Diesbezüglich muss man sich zwischen einer eigenständigen Ent-
299 wicklung und einem Fremdbezug entscheiden. Formal hängt die Entscheidung
300 u. a. von den verfügbaren Ressourcen (Personal, finanzielle Mittel) ab. Inhaltlich
301 hängt diese Entscheidung u. a. davon ab, ob es überhaupt bereits diagnostische
302 Verfahren für das, was man erfassen möchte, gibt. Dies ist umso wahrscheinlicher,
303 je allgemeingültiger und abstrakter die interessierenden Studieneignungsmerkmale
304 sind. Für etablierte Konstrukte wie Intelligenz oder Interessen (Kap. 4) gibt es
305 zahlreiche etablierte diagnostische Verfahren (für die Intelligenz siehe oben [Bei-
306 spiel zur Erfassung kognitiver Fähigkeiten: Der Gießener kognitive Kompetenz
307 Test](#)). Will man es den Studieninteressierten aber ermöglichen, ihre Vorkenntnisse
308 für ein spezifisches Fach zu überprüfen, wird man das entsprechende Verfahren
309 voraussichtlich selbst entwickeln müssen. Ein Sonderfall ist die Modifikation
310 eines bestehenden Verfahrens. Dies minimiert den Entwicklungsaufwand, der Auf-
311 wand für die Qualitätskontrolle ist aber annähernd genauso hoch wie bei einer
312 Neuentwicklung.

313 Die Entscheidung, welches diagnostische Verfahren für die Fragestellung
314 geeignet ist, hängt vor allem von der Qualität des Verfahrens ab. Die Verfahren
315 müssen, u. a., gültig (valide) und zuverlässig (reliabel) sowie für die Zielgruppe
316 (hier: Studieninteressierte) geeignet sein und einen eindeutigen Anforderungs-
317 bezug aufweisen.

318 Bei der Nutzung bereits vorhandener Verfahren sind formale und inhaltliche
319 Aspekte zu beachten. Formal ist zu klären, ob die Nutzung rechtlich zulässig ist.
320 Einschlägig sind hier die Regelungen des Urheber- und Wettbewerbsgesetzes.



321 Verstöße gegen diese Gesetze können zivil- und strafrechtliche Konsequenzen
322 nach sich ziehen. Das Diagnostik und Testkuratorium (DTK) hat 2019 eine
323 Broschüre mit Informationen zum Urheberrecht publiziert, die sich spezifisch
324 auf psychologisch diagnostische Verfahren bezieht ([https://www.bdp-verband.de/
325 binaries/content/assets/beruf/testrezensionen/dtk_tests_in_lehre_und_forschung.
326 pdf](https://www.bdp-verband.de/binaries/content/assets/beruf/testrezensionen/dtk_tests_in_lehre_und_forschung.pdf)). Es gibt drei Möglichkeiten, diagnostische Verfahren für den Einsatz im
327 Rahmen eines OSA zu gewinnen:

- 328 1. man kauft ein Verfahren (bzw. die Lizenzen)
- 329 2. man nutzt ein Verfahren, für das freie Lizenzen vorliegen
- 330 3. man entwickelt ein Verfahren

331 Die Modifikation eines vorhandenen Verfahrens setzt ebenfalls das Einverständnis
332 derer, die die Rechte innehaben, voraus, sodass diese Variante unter (1) oder (2)
333 fällt.

334 Für den Einsatz in OSA besonders interessant sind Fragebogen und Tests, zu
335 denen seitens der Urheberschaft freie Lizenzen vergeben wurden. Diese Frage-
336 bogen und Tests dürfen ohne zusätzliche Genehmigung genutzt werden, sofern
337 erstens die Personen die die Urheberrechte innehaben namentlich genannt werden
338 und zweitens die Art der Nutzung (z. B. die Frage, ob Abwandlungen erlaubt sind)
339 den Vorgaben der Lizenz entspricht (DTK, 2019). Tests und Fragebogen oder auch
340 Items, für die freie Lizenzen vorliegen, finden sich u. a. auf den folgenden open-
341 access-Plattformen:

- 342 • International Personality Item Pool (IPIP, <http://ipip.ori.org>)
- 343 • Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS, [https://zis.
344 gesis.org](https://zis.gesis.org))
- 345 • Elektronisches Testarchiv ([https://www.testarchiv.eu/index.php?wahl=testarchi
346 veintro](https://www.testarchiv.eu/index.php?wahl=testarchivintro))
- 347 • PSYCHOMETRIKON, psychologisch-medizinisches Testportal ([http://www.
348 psychometrikon.de](http://www.psychometrikon.de))
- 349 • International Cognitive Ability Resource (ICAR, <https://icar-project.com>)

350 Ggf. finden sich auch unter einer offenen Lizenz veröffentlichte und brauch-
351 bare Verfahren auf der Website „Open Educational Resources (OER)“ ([https://
352 open-educational-resources.de](https://open-educational-resources.de)) sowie in der „Datenbank zur Qualität von Schule
353 (DaQS)“ (<https://daqs.fachportal-paedagogik.de/show-test>).

354 Die Neuentwicklung von diagnostischen Verfahren ist ein aufwendiger und
355 hochkomplexer Prozess, der spezifische und umfassende Erfahrung erfordert.
356 Im vorliegenden Text können wir den Prozess nur grob skizzieren. Für ausführ-
357 liche Darstellungen verweisen wir auf einschlägige Fachbücher (z. B. Beauducel
358 & Leue, 2014; Bühner, 2021; Eid & Schmidt, 2014; Moosbrugger & Kelava,
359 2020 sowie Rost, 1996). Ausgangspunkt der Testentwicklung sind die Ergebnisse
360 der Auftragsklärung (Kap. 2) und Anforderungsanalyse (Kap. 3), anschließend
361 folgt – vor dem Hintergrund einschlägiger theoretischer Erkenntnisse – eine
362 Präzisierung dessen, was der Test messen soll. Im nächsten Schritt werden Test-



363 und Fragenformate festgelegt und ein erster Testentwurf wird entwickelt und
364 erprobt ([Der Eigenschaftsansatz in der psychologischen Diagnostik](#)). Dazu werden
365 in empirischen Studien Daten erhoben. Die empirischen Daten werden dann einer
366 Testanalyse unterzogen (z. B. Verteilungsanalyse, Iteanalyse, Skalenanalyse).
367 Dabei geht es beispielsweise darum, zu prüfen, ob die Aufgaben eines Leistungs-
368 tests eine angemessene Schwierigkeit haben. Anschließend wird der Testentwurf
369 auf Basis der empirischen Erkenntnisse revidiert und erneut empirisch erprobt.
370 Dies wird so lange wiederholt, bis eine akzeptable Fassung des Tests vorliegt, die
371 den Qualitätsanforderungen an Tests (Abschn. [5.3.2](#)) genügt. Sofern eine norm-
372 orientierte Interpretation erfolgen soll, muss der Test schließlich anhand einer für
373 die Fragestellung repräsentativen großen Gruppe an Personen geeicht werden.
374 Dieser kurze Überblick über die erforderlichen Entwicklungsschritte zeigt, dass
375 die Testentwicklung – ebenso wie die Modifikation eines vorhandenen Tests – mit
376 erheblichen Herausforderungen verbunden ist und daher einschlägig qualifiziertes
377 Personal sowie finanzielle Ressourcen erfordert.

378 **5.3.2 Beurteilung der Güte diagnostischer Verfahren**

379 In dem einschlägigen Qualitätsstandard, der DIN 33430 ([2016](#), siehe auch das
380 vom DTK ([2018a](#)) herausgegebene Buch zur DIN), werden für alle Kategorien von
381 diagnostischen Verfahren Qualitätsanforderungen formuliert. Wir gehen im Folgenden
382 überblicksmäßig auf die Anforderungen an die im Rahmen von OSA am häufigsten
383 genutzten Verfahren ein: die psychometrisch fundierten Fragebogen und Tests.
384 Darüber hinaus nennen wir auch ausgewählte Anforderungen an die computer- und
385 internetgestützte Darbietung von Verfahren, die für OSA ebenfalls einschlägig sind.

386 Übereinstimmend werden in der DIN 33430 ([2016](#)) und im Testbeurteilungss-
387 ystem des Diagnostik- und Testkuratoriums (DTK, [2018b](#)) sechs Qualitäts-
388 kriterien für psychometrisch fundierte Fragebogen und Tests genannt:

- 389 • Theoretische Fundierung als Ausgangspunkt der Testkonstruktion
- 390 • Objektivität
- 391 • Normierung
- 392 • Zuverlässigkeit (Reliabilität, Messgenauigkeit)
- 393 • Gültigkeit (Validität)
- 394 • weitere Gütekriterien (z. B. Störanfälligkeit, Unverfälschbarkeit, Fairness)

395 Die Bedeutung dieser einzelnen Begriffe wie Reliabilität erläutern wir weiter
396 unten, **eine** spezifisch auf OSA bezogene Darstellung der zuletzt genannten fünf
397 Gütekriterien findet sich auch bei Schmitt und Schmidt-Atzert ([2019](#)). Ob die
398 Qualitätskriterien erfüllt werden, lässt sich letztendlich nur anhand von Kenn-
399 werten beurteilen, die in empirischen Untersuchungen ermittelt wurden. Die
400 Betrachtung des Fragebogens oder Tests allein lässt zwar bereits wichtige Ein-
401 schätzungen zu, etwa zur theoretischen Fundierung, zur Verständlichkeit oder
402 zur „Passung“ des Verfahrens zur Fragestellung und zur Zielgruppe – für die



403 Qualitätsbeurteilung sind aber zusätzlich empirische Studien unabdingbar.
404 Diese Studien können entweder von Dritten durchgeführt worden sein (z. B. bei
405 „gekauften“ Tests von der test anbietenden Institution) oder es handelt sich um
406 eigene Studien. In jedem Fall ist die Qualität der empirischen Studien zu prüfen.
407 Ein weiterer Gesichtspunkt ist die Frage, ob die Zielgruppe und die Unter-
408 suchungsbedingungen repräsentativ für die OSA-Zielgruppe und die Unter-
409 suchungsbedingungen sind. Wenn beispielsweise ein Test sich in einer Studie mit
410 lebensälteren Personen, die unter Aufsicht getestet wurden, bewährt hat, lässt sich
411 daraus nicht ohne weiteres schließen, dass sich der Test auch für den Einsatz im
412 Kontext von OSA (unbeaufsichtigte Testung, überwiegend junge Testpersonen)
413 eignet.

414 Aus der Tatsache, dass für die Qualitätsbeurteilung die empirischen Unter-
415 suchungen von zentraler Bedeutung sind, ergibt sich, dass diese Untersuchungen
416 und ihre Ergebnisse umfassend und nachvollziehbar dokumentiert sein müssen.
417 Deshalb verlangt die DIN 33430 (2016), dass zu jedem Fragebogen und Test
418 sogenannte „Verfahrenshinweise“ vorliegen. Diese Texte, häufig auch „Test-
419 manual“ oder „Testhandbuch“ genannt, müssen diejenigen Informationen
420 umfassen, auf deren Basis die Qualität des Verfahrens erst beurteilt werden kann.
421 Diese Qualitätsbewertung ist wiederum die Grundlage für die Entscheidung, ob
422 das Verfahren im Rahmen eines OSA eingesetzt werden kann.

423 In den Anhängen A und B der DIN (2016) wird definiert, welche Informationen
424 zu einem Test oder Fragebogen vorliegen müssen. Kersting (2018) hat diese
425 Informationsanforderungen in Frageform umformuliert und daraus eine Check-
426 liste entwickelt. Diese Checkliste „DIN SCREEN, Checkliste 1, Version 3 /
427 Anforderungen an Verfahrenshinweise“ umfasst 54 Fragen. Sie gilt offiziell als
428 „Standard zur Information und Dokumentation von Instrumenten zur Erfassung
429 menschlichen Erlebens und Verhaltens des Diagnostik- und Testkuratoriums der
430 Föderation Deutscher Psychologinnenvereinigungen“ ([http://kersting-internet.de/
431 pdf/Kersting_2017_in_DTK_DIN-SCREEN_Checkliste_1.pdf](http://kersting-internet.de/pdf/Kersting_2017_in_DTK_DIN-SCREEN_Checkliste_1.pdf)). Fragebogen und
432 Tests, die im Rahmen von OSA eingesetzt werden, müssen diesen Informations-
433 standard erfüllen.

434 Tab. 5.1 zeigt beispielhaft einen Auszug aus der Checkliste. Auf der Basis
435 dieser Informationen erfolgt dann die Qualitätsbeurteilung. In den Test-
436 beurteilungsrichtlinien des DTK (2018b) ist festgehalten, wie die oben genannten
437 fünf Kriterien zu beurteilen sind. Dabei geht es keinesfalls nur um die numerische
438 Ausprägung der jeweiligen Koeffizienten. Bezüglich der Reliabilität ist beispiels-
439 weise zu prüfen, ob die gewählte Schätzmethode überhaupt angemessen ist. Ein
440 auf den ersten Blick überzeugend hoher Wert für die interne Konsistenz (als
441 Schätzer für die Reliabilität) verliert beispielsweise an Wert, wenn dieser Wert auf
442 nahezu identisch gestaltete Items zurückzuführen ist (DTK, 2018b).

443 Da OSA computerbasiert sind und internetgestützt eingesetzt werden, müssen
444 zudem die Systemstabilität und die Funktionsfähigkeit auf den relevanten
445 Betriebssystemplattformen gewährleistet werden (siehe DIN 33430 sowie Kap. 8).
446 Das Ergebnis des OSA darf laut DIN 33430 (2016) nicht von Aspekten abhängen,
447 die nicht zur Fragestellung gehören, beispielsweise von dem Ausmaß der



Tab. 5.1 Beispiel für die Fragen und Antwortkategorien der „DIN SCREEN, V3/Checkliste 1/ Anforderungen an Verfahrenshinweise“ sowie für eine Bewertungsverknüpfung (Kersting, 2018, S. 229)

		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?</i> <i>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?</i> <i>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"/>			

448 Computererfahrung und / oder -kenntnissen. Will man den Qualitätsanforderungen
449 gerecht werden, die die DIN 33430 (2016) formuliert, muss für die Nutzung von
450 OSA ein Support für technische Fragen angeboten werden.

451 Im Folgenden möchten wir nun auf die eingangs genannten Qualitätskriterien
452 für psychometrisch fundierte Fragebogen und Tests nach DIN 33430 (2016) ein-
453 gehen.

454 **Theoretische Fundierung**

455 Bei der theoretischen Fundierung geht es darum, dass ein Test auf einer
456 etablierten überzeugenden Theorie basieren soll (Kap. 4 gibt einen Überblick
457 über wichtige Theorien zu den Konstrukten Intelligenz, Persönlichkeit und beruf-
458 liche Interessen). Nur dann kann auch nach dem [Eigenschaftsansatz](#) [der psychologischen Diagnostik](#)
459 ein theoretisches Messmodell spezifiziert und der Test im
460 Hinblick auf seine Modellpassung überprüft werden.

461 **Objektivität**

462 Das in einem Test erzielte Ergebnis muss unabhängig von den Bedingungen
463 sein, unter denen der Test (a) durchgeführt, (b) ausgewertet und (c) interpretiert



464 wird. Während die Auswertungs- und Interpretationsobjektivität bei den meisten
465 OSA aufgrund der mit der computergestützten Darbietung einhergehenden Auto-
466 matisierung gegeben sind, stellt die Durchführung aufgrund der unbeaufsichtigten
467 Testung ein Problem von OSA dar. Möglicherweise haben einige Testpersonen
468 den Test unter ungünstigen Bedingungen (z. B. keine störungsfreie Umgebung)
469 oder mit ungeeigneter technischer Ausstattung (z. B. wurde ein Smartphone mit
470 einem kleinen Bildschirm genutzt) bearbeitet. In Kap. 8 gehen Pachtchenko und
471 Kollegen auf in diesem Kontext relevante Fragen wie etwa die Usability ein. Ein-
472 schränkungen in der Durchführungsobjektivität lassen sich aber bei Testungen
473 ohne Aufsicht grundsätzlich nicht ausschließen. Daher ist es wichtig, vor der
474 Ergebnisrückmeldung eines OSA entsprechende Plausibilitätsprüfungen durchzu-
475 führen (beispielsweise kann man prüfen, ob die Bearbeitungszeit unplausibel kurz
476 ausfällt) und beim Feedback einen entsprechenden Interpretationsvorbehalt zu
477 formulieren (Kap. 7).

478 **Reliabilität**

479 Bei der Reliabilität (Zuverlässigkeit) geht es nach der DIN 33430 (2016, S. 9) um
480 den „Grad der Genauigkeit eines Verfahrens“, mit dem es ein Merkmal erfasst“,
481 wobei mit Merkmal das „Eignungsmerkmal“ gemeint ist, z. B. die kognitive
482 Kompetenz. „Die Zuverlässigkeit eines Verfahrens kann“, so wird im Text der DIN
483 33430 weiter erläutert, „unter der Voraussetzung gleichbleibender Merkmalsaus-
484 prägung als die Replizierbarkeit von Messergebnissen verstanden werden“ (DIN,
485 2016, ebd.).

486 Um zu bestimmen, wie reliabel ein Test misst, wird der Zusammenhang
487 zwischen zwei oder mehreren Werten, die mit dem Test gewonnen wurden,
488 berechnet, z. B. zwischen

- 489 • zwei (oder mehr) Messungen, die mit demselben Test zu unterschiedlichen
- 490 Zeitpunkten durchgeführt wurden (Retest-Reliabilität)
- 491 • zwischen zwei (oder mehr) Formen des gleichen Tests (Paralleltest-Reliabilität)
- 492 • zwischen einzelnen Aufgaben des Tests (split-half oder interne Konsistenz)

493 Welche Form der Reliabilitätsbestimmungen sinnvoll ist, hängt vom Test und den
494 diagnostischen Zielen ab. Bei einem Test mit heterogenen Inhalten oder bei einem
495 Speed Test, bei dem an sich einfache Items unter hohem Zeitdruck bearbeitet
496 werden, ist es beispielsweise nicht sinnvoll, die interne Konsistenz zu bestimmen.
497 Die Reliabilität wird überschätzt, wenn die interne Konsistenz bei einem Test
498 bestimmt wird, der aus mehreren nahezu gleich gestalteten Items besteht. Bei der
499 Test–Retest-Methode muss das Intervall zwischen Test und Retest angemessen
500 gesetzt werden.

501 **Gültigkeit (Validität)**

502 Die Validität thematisiert die Aussagekraft der aus dem Fragebogen oder Test
503 abgeleiteten Schlussfolgerungen oder Entscheidungen. Es geht letztendlich um
504 Argumente, die belegen, dass die vorgesehene Interpretation der Ergebnisse



505 zulässig ist. Dies setzt, wie Petri, Schütte und Beauducel in Kap. 4 bereits aus-
506 führen, zunächst ein theoretisches Verständnis voraus. Man muss wissen, welches
507 Merkmal man mit dem Test erfasst und mit welchen anderen Merkmalen dieses
508 Merkmal systematisch zusammenhängen sollte. Aus diesem theoretischen Ver-
509 ständnis heraus kann man Hypothesen ableiten und in empirischen Studien prüfen.
510 Dabei gibt es nicht nur eine Art der Validierung, sondern es gibt verschiedene
511 Validierungsstrategien. Zu jeder Validierungsstrategie wiederum sollten mehrere
512 empirische Studien durchgeführt werden, sodass die Validierung insgesamt ein
513 komplexer Prozess ist, der kontinuierlich fortgeschrieben wird. Man unterscheidet
514 drei Validierungsstrategien, die alle zur Validierung beitragen:

515 **Konstruktvalidierung** Bei der Strategie der Konstruktvalidierung wird geprüft,
516 ob der Test das erfasst, was er erfassen soll, und ob die Testergebnisse im Sinne
517 einer überzeugenden Theorie erklärt werden können. Wenn ein Test beispiels-
518 weise kognitive Kompetenzen erfassen soll, so müssen die Ergebnisse dieses Tests
519 gleichsinnig zu den Ergebnissen eines anderen Tests zum gleichen Merkmal sein.
520 Zur Konstruktvalidierung des oben beschriebenen Gießener kognitive Kompetenz
521 Test (GkKT-66; Ulfert et al., 2017) wurde der Test beispielsweise gemeinsam mit
522 dem Wonderlic Intelligenztest eingesetzt ([Praxisbeispiel: Validierung des GkKT](#)
523 [für den Einsatz im OSA der Universität Gießen](#)).

524 **Kriteriumsvalidierung** Bei der Strategie der Kriteriumsvalidierung wird geprüft,
525 ob die Testergebnisse systematisch mit relevantem Verhalten „in der wirklichen
526 Welt“ zusammenhängen und solches Verhalten ggf. vorhersagen können (prä-
527 diktive Kriteriumsvalidität). Mit OSA möchte man beispielsweise dazu beitragen,
528 dass Personen ein Studium wählen, die dieses Studium in der Regelstudien-
529 zeit mit guten Noten absolvieren, zumindest aber das Studium nicht abbrechen.
530 Vielleicht ist es den OSA-Verantwortlichen auch besonders wichtig, dass die
531 Studierenden mit ihrem Studium zufrieden sind. Entsprechend wären die Studien-
532 dauer, die Studiennoten und die Zufriedenheit mit dem Studium geeignete
533 „Kriterien“. Die Kriteriumsvalidierung besteht dann darin, dass man empirische
534 Studien durchführt, in denen der Zusammenhang zwischen den Ergebnissen im
535 Test (Prädiktoren) mit diesen Kriterien analysiert wird. Tab. 4.1 in Kap. 4 gibt eine
536 Übersicht über bislang durchgeführte Studien zur Kriteriumsvalidierung der aus
537 OSA abgeleiteten Interpretationen.

538 **Kontentvalidierung** Bei der Strategie der Kontentvalidierung wird geprüft, ob
539 die Fragen und Aufgaben des Tests das interessierende Merkmal repräsentieren.
540 Diese Strategie lässt sich dann anwenden, wenn das Merkmal, das man mit dem
541 Test erfassen will, wenig komplex ist, z. B. bei einem Vokabeltest. Hier ist es
542 möglich, die Menge aller relevanten Vokabeln (das sogenannte Itemuniversum)
543 zu definieren und daraus regelgeleitet die Items für den Test abzuleiten.
544 Bei komplexen Merkmalen wie der Intelligenz lässt sich die Strategie der
545 Kontentvalidierung streng genommen nicht anwenden, weil das Itemuniversum zu
546 groß ist, um es in einem Test abzubilden.



547 **Normen**

548 Die Auswertung und Interpretation zahlreicher OSA beruht auf einem so
549 genannten normativen Ansatz. Man vergleicht den gemessenen Wert mit den
550 Werten anderer Personen. Die Qualität des Tests hängt in diesem Fall auch davon
551 ab, ob die herangezogene Vergleichsgruppe „passend“ für die Gruppe ist, über die
552 man eine Aussage treffen will. Dies setzt zunächst eine inhaltliche Entscheidung
553 voraus. Will man die Leistung einer Person, die sich für ein Studium der Physik
554 an der RWTH Aachen interessiert, mit Studierenden im Allgemeinen oder nur mit
555 Physik-Studierenden im Besonderen vergleichen? Mit Studierenden irgendeiner
556 Hochschule oder nur mit Studierenden der RWTH Aachen? Mit Studierenden im
557 ersten Semester oder mit fortgeschrittenen Studierenden? Mit durchschnittlichen
558 Studierenden oder nur mit besonders erfolgreichen Studierenden? (Wobei die
559 Frage, was als Erfolg gilt, ebenfalls definiert werden muss; siehe Kap. 3) In jedem
560 Fall muss im Rahmen eines OSA die Normgruppe beschrieben werden. Gut ist ein
561 Test unter diesem Gesichtspunkt, wenn die Normdaten in jüngerer Zeit an einer
562 ausreichend großen Gruppe von Personen erhoben wurden, die hinsichtlich ent-
563 scheidungsrelevanter Merkmale der Gruppe entsprechen, die man als Zielgruppe
564 des Verfahrens ansieht.

565 **Testfairness**

566 Neben den bereits behandelten Kriterien gehen die DIN 33430 (2016) und das
567 DTK (2018b) auf weitere Qualitätskriterien wie die Störanfälligkeit, Unver-
568 fälschbarkeit und die Skalierung ein. Ein weiteres Qualitätskriterium für Tests ist
569 auch die Testfairness. Kubinger und Proyer (2010, S. 178) definieren den Begriff
570 in einem Erläuterungswerk zur DIN 33430 wie folgt: „Ein Test erfüllt das Güte-
571 kriterium Fairness, wenn die resultierenden Testwerte zu keiner systematischen
572 Diskriminierung bestimmter Testpersonen z. B. aufgrund ihrer ethnischen, sozio-
573 kulturellen oder geschlechtsspezifischen Gruppenzugehörigkeit führen.“

574 Gemeint ist natürlich nicht die Diskriminierung zwischen Personen, die für
575 das Studium geeignet sind und Personen, die nicht geeignet sind – diese „Dis-
576 kriminierung“ ist intendiert. Darauf geht das sogenannte Regressions-Modell
577 der fairen Vorhersage nach Cleary (1968) ein. Aussagen zur Fairness lassen sich
578 diesem Modell zufolge erst treffen, wenn man nicht nur die Ergebnisse im Test,
579 sondern simultan auch zukünftiges Verhalten der getesteten Personen, also die
580 Vorhersagekraft des Tests (z. B. Vorhersage des – wie auch immer definierten –
581 Studienerfolgs) für verschiedene Gruppen betrachtet.

582 Diese wissenschaftlich fundierte Betrachtungsweise steht im Widerspruch
583 zu verbreiteten „Alltagsvorstellungen“. Viele Menschen empfinden einen Test
584 als „unfair“, wenn Angehörige einer Gruppe (z. B. Frauen) systematisch bessere
585 oder schlechtere Testergebnisse erzielen als die Angehörigen einer anderen
586 Gruppe (z. B. Männer). Häufig werden dann gruppenspezifische Testnormen
587 (z. B. separate Normen für Frauen und Männer) oder „Quoten“ gefordert.
588 Solche Forderungen sind aber rein politischer Natur und lassen sich nicht
589 fachlich begründen. Die von verschiedenen Fachorganisationen gemeinsam
590 herausgegebenen „Standards“ (American Educational Research Association,



591 Psychological Association und National Council on Measurement in Education,
592 2014) betonen explizit, dass Gruppenunterschiede in Verfahrensergebnissen kein
593 hinreichender Indikator für Unfairness sind. Unter Validitätsgesichtspunkten geht
594 es darum, ob der Test das vorherzusagende Kriterium (z. B. Studienerfolg oder
595 Studienzufriedenheit) für verschiedene Gruppen gleich gut schätzt. Konkret: Wenn
596 Männer in einem bestimmten Studiengang im Durchschnitt signifikant schlechter
597 abschneiden als Frauen, ist es nicht unfair, wenn die Männer auch in dem Studien-
598 eignungstest für diesen Studiengang im Durchschnitt schlechter abschneiden. Zu
599 dieser Frage gibt es nur wenige Studien, die sich auf universitäre Auswahlver-
600 fahren (nicht aber auf OSA) beziehen. Schult et al. (2013) stellten beispielsweise
601 fest, dass die Studienleistung von Frauen mit numerischen Tests etwas unter-
602 schätzt wurde, während verbale Tests die Leistung von Männern etwas unter-
603 schätzten. Bergholz und Stegt (2018) prüfen die Vorhersage der Leistung im
604 Studium der Rechtswissenschaften. Für den Auswahltest insgesamt ergaben sich
605 keine signifikanten Unterschiede bezüglich der Studienerfolgsvorhersage, die Vor-
606 hersageleistung der einzelnen Subtests variierte aber geringfügig in Abhängigkeit
607 der untersuchten Gruppen (Frauen und Männer).

608 Akzeptanz

609 Bei den bisher beschriebenen Qualitätskriterien Objektivität, Reliabilität, Validität,
610 Normen und Testfairness handelt es sich um sachliche Aspekte, die Verantwort-
611 liche bei der Entscheidung darüber nutzen, wie sie Verfahren gestalten oder
612 welches Verfahren sie auswählen. Mit der „sozialen Akzeptanz“ wird hingegen die
613 Perspektive der getesteten Personen berücksichtigt. Zur Akzeptanz diagnostischer
614 Verfahren und Prozesse liegen zahlreiche Studien vor, die mittlerweile auch meta-
615 analytisch zusammengefasst wurden (z. B. Anderson et al., 2010; Hausknecht
616 et al., 2004). Diese Studien zeigen die hohe Bedeutung der Akzeptanz psycholo-
617 gischer Diagnostik für die wahrgenommene Attraktivität einer Organisation. Die
618 Akzeptanz wirkt sich u. a. auf die Bereitschaft aus, eine angebotene Stelle anzu-
619 nehmen und anderen Stelleninteressierten die Organisation zu empfehlen. Im OSA
620 Kontext entscheiden die getesteten Personen selbst, ob sie der ausgesprochenen
621 Empfehlung folgen oder nicht (Kap. 4). Eine geringe Akzeptanz wirkt sich dies-
622 bezüglich womöglich ungünstig aus.

623 Bei OSA kommt der Akzeptanz der psychologischen Diagnostik eine weitere
624 Bedeutung zu: Da die Teilnahme am Verfahren im Regelfall freiwillig ist,
625 müssen die Studieninteressierten motiviert sein, um ein OSA aufzurufen und bis
626 zum Ende zu bearbeiten. Dabei spielt die Akzeptanz eine große Rolle. Hier sind
627 zunächst Aspekte der Usability zu betrachten, die im vorliegenden Buch in Kap. 8
628 behandelt werden (siehe auch Ott et al., 2017). Eine wichtige Voraussetzung für
629 die Akzeptanz ist ein fehlerfrei und schnell funktionierendes, intuitiv verständ-
630 liches, benutzerfreundliches und optisch attraktives Tool.

631 Eine hohe Bedeutung für die Akzeptanz haben aber auch die Testaufgaben.
632 Hier kommt es u. a. auf die Verständlichkeit der Instruktionen, auf die Schwierig-
633 keit und – bei Persönlichkeitsfragebogen – auf die Wahrung der Privatsphäre
634 an. Kersting hat einen Fragebogen entwickelt, mit dem sich verschiedenen



635 Dimensionen der Akzeptanz psychologisch-diagnostischer Verfahren messen
636 lassen (siehe z. B. Beermann et al., 2013; Kersting, 2008). Besonders bedeut-
637 sam für die Akzeptanz ist den entsprechenden empirischen Studien zu Folge
638 die Augenscheinvalidität der Tests (Kersting, 2008). Die Augenscheinvalidität
639 bedeutet im Kontext von OSA die persönliche Überzeugung der Testpersonen,
640 dass die Inhalte des OSA eine gute Simulation der Studienanforderungen dar-
641 stellen und zur Vorhersage des Studienerfolgs geeignet sind. Diese persönliche
642 Überzeugung hängt nicht unbedingt mit der tatsächlichen (empirischen) Aussage-
643 kraft der OSA zusammen. OSA werden besonders dann positiv wahrgenommen,
644 wenn in den Aufgaben und Fragen die „Welt“ des Studiums im Allgemeinen und /
645 oder des konkreten Studienganges im Besonderen aufgegriffen werden, wenn der
646 Test also quasi eine Vorschau auf das Studium gibt.

647 Sofern abstrakte Aufgaben eingesetzt werden, empfiehlt es sich, mit einer
648 kurzen Erläuterung den Bezug der Aufgaben zum infrage stehenden Studium her-
649 zustellen. Von hoher Bedeutung für die Akzeptanz ist auch die individuelle Rück-
650 meldung zu den Testergebnissen (Kap. 7).

651 Qualitätsbeurteilung

652 Die Beurteilung der Qualität eines Tests kann nicht mechanisch erfolgen,
653 auch wenn die aus den Studien resultierenden „Kennwerte“ zu einer solchen
654 mechanischen Beurteilung verführen. Die Ausprägung der Reliabilität und
655 Validität wird in der Regel als Korrelationskoeffizient angegeben. Dieser nimmt
656 Werte zwischen 0,00 und 1,00 an. Daher liegt es erst einmal nahe, anzunehmen,
657 dass im Vergleich zweier Verfahren dasjenige positiver zu bewerten ist, welches
658 die höheren Kennwerte erzielt hat. Diese Annahme ist aber nicht immer sinn-
659 voll. Zunächst handelt es sich bei den Kennwerten immer nur um statistische
660 Schätzungen. Darüber hinaus hängen die Kennwerte nicht nur vom Test, sondern
661 auch von den Umständen der jeweiligen Untersuchung ab. Ganz wesentlich ist
662 z. B., dass die Testwerte möglichst stark variieren. Setzt man z. B. einen Test A
663 zur Merkfähigkeit bei 100 zufällig ausgewählten Personen ein und einen anderen
664 Test B zur Merkfähigkeit bei 100 in Gedächtniskunst trainierten Personen, so wird
665 die letztgenannte Gruppe kaum Fehler im Test aufweisen. Dies hat zur Folge, dass
666 die Werte im Test B nicht variieren, alle in Gedächtniskunst Trainierten erhalten
667 den gleichen (hohen) Testwert. Somit fallen alle Korrelationen (und somit Kenn-
668 werte) mit anderen Variablen, die man für diesen Test B berechnet, zwangsläufig
669 gering aus. Dies liegt aber nicht am Test B, der Test A hätte in dieser Gruppe auch
670 versagt. Die Studie war schlecht konzipiert, man hätte keine Versuchspersonen
671 mit Expertise in Gedächtniskunst mit einem Test untersuchen dürfen, der für
672 Menschen mit einer Merkfähigkeit im durchschnittlichen Bereich konstruiert
673 wurde. (Für weitere Erläuterungen siehe Kersting, 2006.)

674 Die Beurteilung der Qualität eines Tests erfordert Expertise im Bereich
675 empirischer Forschung, um die zugrunde liegenden empirischen Studien bewerten
676 zu können. Hierzu gehört beispielsweise die Frage, ob die untersuchte Stichprobe
677 der Zielgruppe des Verfahrens entspricht.



678 Da zu einem Test im Idealfall mehrere Studien (beispielsweise zur Reliabili-
679 tät und Validität) vorliegen, wobei unterschiedliche Schätzmethoden für die
680 Reliabilitätsbestimmung und unterschiedliche Validierungsstrategien genutzt
681 wurden, müssen für die Qualitätsbeurteilung eine Vielzahl der Informationen zu
682 einem Gesamturteil integriert werden. Wie die Testbeurteilung im Einzelnen vor-
683 genommen werden kann, wird in den Richtlinien zum Testbeurteilungssystem des
684 Diagnostik- und Testkuratoriums (2018b) ausgeführt.

685 **Praxisbeispiel: Validierung des GkKT für den Einsatz im OSA der Universität**
686 **Gießen**

687 Die Qualität des GkKT für den Einsatz in OSA zur Studienfachwahl wurde
688 anhand mehrerer Studien, mit Hinblick auf die oben dargestellten Güte-
689 kriterien, untersucht. Die Ergebnisse dieser Studien sprechen für die Qualität
690 des Verfahrens:

691 Der Test ist reliabel (interne Konsistenz des Gesamtwerts: McDonalds
692 Omega $\omega_t = .900$, $N = 802$). Die Konstruktvalidität wurde anhand der Überein-
693 stimmung mit anderen, etablierten Intelligenztests nachgewiesen. So korreliert
694 der GkKT beispielsweise zu $r = .77$ mit dem Wonderlic Intelligenztest
695 ($N = 181$). Zur Kriteriumsvalidierung wurden bislang u. a. Korrelationen zu
696 Schulnoten (logisch positiver Zusammenhang in Höhe von $r = -.31$ ($N = 101$))
697 zwischen dem GkKT und der Abiturnote) bestimmt.

698 Grundlage der hier geschilderten Validitätsuntersuchungen war die Version
699 1.0 des Tests, die nur über 45 Items verfügte.

700 Unter Akzeptanzgesichtspunkten erzielte der Test auf einer sechsstufigen
701 Skala (1 = „sehr negativ“ bis 6 = „sehr positiv“) mit einem Wert von 3,5 eine
702 Bewertung im positiven Bereich ($N = 512$). Das ist ein typischer Befund für
703 Leistungstests (Kersting, 2008). Eine ausführliche Darstellung des Tests geben
704 Petri et al., (2019)¹. ◀

705 **Fazit**

706 Für den Einsatz in OSA kommen verschiedene diagnostische Verfahren infrage,
707 am häufigsten werden psychometrisch fundierte Fragebogen und Tests ein-
708 gesetzt. Man kann entweder eigene diagnostische Verfahren entwickeln oder auf
709 vorhandene Verfahren zurückgreifen. Im letztgenannten Fall ist das Urheber-
710 recht zu beachten. In beiden Fällen müssen die Verfahren umfassenden Qualitäts-
711 anforderungen gerecht werden.

¹Der GkKT wird von der Gießener Arbeitsgruppe auf Anfrage für die Nutzung in nicht-gewinnorientierten OSA mit einer freien Lizenz zur Verfügung gestellt.



712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735

Checkliste zum Einsatz von diagnostischen Elementen in OSA

1. Sind die psychologischen Konstrukte (die Studieneignungsmerkmale), die untersucht werden sollen, präzise definiert?
2. Gibt es bereits diagnostische Verfahren, mit denen das zu untersuchende Konstrukt (das Studieneignungsmerkmal) erfasst werden kann? Wenn ja, sind diese diagnostischen Verfahren frei verfügbar oder müssen Lizenzen erworben werden?
3. Liegen für das diagnostische Verfahren, das man einsetzen will, Verfahrenshinweise vor, die den Anforderungen nach „DIN SCREEN Checkliste 1, Version 3“ genügen?
4. Wie ist die Qualität der diagnostischen Verfahren hinsichtlich der Aspekte theoretische Fundierung, Objektivität, Reliabilität, Validität, Normen, Fairness und Akzeptanz zu beurteilen?
5. Muss das diagnostische Verfahren mit Blick auf die vorliegenden Gegebenheiten (Zielgruppe, Kriterium, Testsituation, Teilnahmemotivation, ...) angepasst werden? Wenn ja, inwieweit beeinflusst dies die Qualitätskriterien des Instruments? Sind Modifikationen im Rahmen der Urheberschaft gestattet? Ist das Verfahren für computerbasierte Durchführungen geeignet?
6. Besteht im Falle einer Neuentwicklung eines diagnostischen Verfahrens die erforderliche fachliche Expertise? Stehen die notwendigen zeitlichen, personellen und finanziellen Ressourcen zur Verfügung?
7. Wie gut wurde das diagnostische Verfahren in bislang durchgeführten Studien von den Testpersonen akzeptiert?

736

Literatur

737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754

- American Educational Research Association, Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Anderson, N., Salgado, J. F., & Hülsheger, U. R. (2010). Applicant reactions in selection: Comprehensive meta-analysis into reaction generalization versus situational specificity. *International Journal of Selection and Assessment*, 18, 291–304.
- Beauducel, A., & Leue, A. (2014). *Psychologische Diagnostik*. Hogrefe.
- Beermann, D., Kersting, M., Stegt, S., & Zimmerhofer, A. (2013). Vorurteile und Urteile zur Akzeptanz von Persönlichkeitsfragebogen. *Personal Quarterly*, 65, 41–45.
- Bergholz, L., & Stegt, S. J. (2018). Validität und Fairness eines Studierfähigkeitstests für Rechtswissenschaften. *Zeitschrift für Hochschulentwicklung*, 13, 57–79.
- Bühner, M. (2021). *Einführung in die Test- und Fragebogenkonstruktion* (4, korrigierte und erweiterte Aufl.). Pearson.
- Cleary, T. A. (1968). Test bias: Prediction of grades of negro and white students in integrated colleges. *Journal of Educational Measurement*, 5, 115–124.
- Diagnostik- und Testkuratorium der Föderation Deutscher Psychologenvereinigungen. (2018a). *Personalauswahl kompetent gestalten: Grundlagen und Praxis der Eignungsdiagnostik nach DIN 33430*. Springer.



- 755 Diagnostik- und Testkuratorium der Föderation Deutscher Psychologenvereinigungen. (2018b).
756 TBS-DTK. Testbeurteilungssystem des Diagnostik- und Testkuratoriums der Föderation
757 Deutscher Psychologenvereinigungen. Revidierte Fassung vom 3. Januar 2018. *Psycho-*
758 *logische Rundschau*, 69, 109–148.
- 759 Diagnostik- und Testkuratorium der Föderation Deutscher Psychologenvereinigungen. (2019).
760 Tests in Lehre und Forschung. Informationen zum Testschutz und zum Urheberrecht. Verfüg-
761 bar unter: [https://www.dgps.de/fileadmin/user_upload/foederation/dtk_tests_in_lehre_und_](https://www.dgps.de/fileadmin/user_upload/foederation/dtk_tests_in_lehre_und_forschung.pdf)
762 [forschung.pdf](https://www.dgps.de/fileadmin/user_upload/foederation/dtk_tests_in_lehre_und_forschung.pdf) [Stand 03.02.2020]
- 763 DIN. (2016). *DIN 33430: Anforderungen an berufsbezogene Eignungsdiagnostik*. Beuth.
- 764 Eid, M., Gollwitzer, M., & Schmidt, K. (2014). *Testtheorie und Testkonstruktion*. Hogrefe.
- 765 Hasenberg, S., & Stoll, G. (2015). Erwartungschecks in Self-Assessments: Zur Erfassung und
766 Korrektur von Studiererwartungen. *Das Hochschulwesen*, 63, 104–108.
- 767 Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection
768 procedures: An updated model and meta-analysis. *Personnel Psychology*, 57, 639–683.
- 769 Holland, J. L. (1997). Making vocational choices: A theory of vocational personalities and work
770 environments. Psychological Assessment Resources.
- 771 Höft, S., & Kersting, M. (2018). Anforderungsprofil, Verhaltensbeobachtung und Verhaltensbe-
772 urteilung. In Diagnostik- und Testkuratorium (Hrsg.), *Personalauswahl kompetent gestalten:*
773 *Grundlagen und Praxis der Eignungsdiagnostik nach DIN 33430* (S. 27–63). Springer.
- 774 Kersting, M. (2006). Zur Beurteilung der Qualität von Tests: Resümee und Neubeginn. *Psycho-*
775 *logische Rundschau*, 57, 243–253.
- 776 Kersting, M. (2008). Zur Akzeptanz von Intelligenz- und Leistungstests. *Report Psychologie*, 33,
777 420–433.
- 778 Kersting, M. (2011). Managementdiagnostik: Verfahren und Qualitätsaspekte. In C.
779 Niedereichholz, J. Niedereichholz, & J. Staude (Hrsg.), *Handbuch der Unternehmens-*
780 *beratung. Organisationen führen und verwalten* (Kz. 3960, S. 1–18). Schmidt.
- 781 Kersting, M. (2018). Zur Information über und Dokumentation von Instrumenten zur Erfassung
782 menschlichen Erlebens und Verhaltens – Die DIN SCREEN Checkliste 1, Version 3. In
783 Diagnostik- und Testkuratorium (Hrsg.), *Personalauswahl kompetent gestalten: Grundlagen*
784 *und Praxis der Eignungsdiagnostik nach DIN 33430* (S. 223–244). Springer.
- 785 Kubinger, K. D., & Proyer, R. (2010). Gütekriterien. In K. Westhoff, C. Hagemeyer, M.
786 Kersting, F. Lang, H. Moosbrugger, G. Reimann, & G. Stemmler (Hrsg.), *Grundwissen für*
787 *die berufsbezogene Eignungsbeurteilung nach DIN 33430* (3 neubearbeitete, S. 173–180).
788 Pabst Science Publishers.
- 789 Moosbrugger, H., & Kelava, A. (2020). *Testtheorie und Fragebogenkonstruktion* (3., korrigierte
790 und erweiterte Aufl.). Springer.
- 791 Ott, M., Ulfert, A., & Kersting, M. (2017). „Online-Assessments“ und „Self-Assessments“ in der
792 Eignungsdiagnostik. In D. E. Krause (Hrsg.), *Personalauswahl* (S. 215–242). Springer Fach-
793 medien.
- 794 Petri, P. S., Weingardt, V., & Kersting, M. (2019). Let’s get to the hard facts – Erfassung
795 kognitiver Fähigkeiten im Rahmen von Online-Self-Assessments. In C. Marquardt & K.
796 Zay (Hrsg.), *Online-Self-Assessments zur Studienorientierung* (Themenheft), *Empirische*
797 *Pädagogik*, 33(3), 307–330.
- 798 Rost, J. (1996). *Lehrbuch Testtheorie, Testkonstruktion*. Huber.
- 799 Scherndl, T., Ortner, T., & Leiner, J. (2019). Indikatoren der Akzeptanz eines multimethodischen
800 Online-Self-Assessments unter der Bedingung der Teilnahmeverpflichtung für Studien-
801 interessierte. In C. Marquardt & K. Zay (Hrsg.), *Online-Self-Assessments zur Studien-*
802 *orientierung* (Themenheft), *Empirische Pädagogik*, 33(3), 362–382.
- 803 Schmitt, M., & Schmidt-Atzert, L. (2019). Gütekriterien für Online-Self-Assessments. In L.
804 Schmidt-Atzert, M. Schütz, & G. Stemmler (Hrsg.), *Online-Self-Assessments an Hochschulen*
805 (S. 99–114). Pabst Science Publishers.
- 806 Schult, J., Hell, B., Päßler, K., & Schuler, H. (2013). Sex-specific differential prediction of
807 academic achievement by German ability tests. *International Journal of Selection and Assess-*
808 *ment*, 21, 130–134.



- 809 Trapmann, S., Hell, B., Weigand, S., & Schuler, H. (2007). Die Validität von Schulnoten zur Vor-
810 hersage des Studienerfolgs – eine Metaanalyse. *Zeitschrift für Pädagogische Psychologie*, 21,
811 11–27.
- 812 Ulfert, A.-S., Ott, M., Bothe, P. S., & Kersting, M. (2017). Gießener kognitiver Kompetenztest,
813 66 Item-Version (GkKT-66). Verfahrenshinweise. Version 2.0. Unveröffentlichtes Manuskript.
814 Justus-Liebig-Universität Gießen.
- 815 Westhoff, K., & Strobel, A. (2005). Eignungsinterview. In K. Westhoff, L. J. Hellfritsch, L.
816 F. Hornke, K. D. Kubinger, F. Lang, H. Moosbrugger, A. Püschel, & G. Reimann (Hrsg.),
817 *Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33430* (S. 93–110).
818 Pabst.

UNCORRECTED PROOF