



GESIS Leibniz-Institut  
für Sozialwissenschaften

Martin Kersting

## DTK-Testinformationsstandard - Auszug aus der DIN SCREEN Checkliste, Version 4.0

Bitte zitieren Sie diesen Beitrag wie folgt:

Kersting, M. (2025). DTK-Testinformationsstandard. Auszug aus der DIN SCREEN Checkliste, Version 4.0. Mannheim: GESIS - Leibniz-Institut für Sozialwissenschaften.

Freier Download unter:

[https://zis.gesis.org/appendix/ZIS\\_DTK\\_DIN-Screen\\_DE\\_v4.pdf](https://zis.gesis.org/appendix/ZIS_DTK_DIN-Screen_DE_v4.pdf)

### **Nutzungsbedingungen**

CC BY-BY-ND 4.0



Zusammenstellung sozialwissenschaftlicher Items und Skalen  
[www.gesis.org/zis](http://www.gesis.org/zis)

## **DTK – Testinformationsstandard – Auszug aus der DIN SCREEN Checkliste, Version 4.0 (Kersting, 2025)**

### **Beschreibung**

Eine fundierte Beurteilung der Qualität von Fragebögen und Tests setzt sowohl die Kenntnis der theoretischen Grundlagen, auf denen das Verfahren beruht, als auch empirischer Studien voraus, in denen das Verfahren hinsichtlich verschiedener Gütekriterien geprüft wurde. Daher besteht der erste Schritt darin zu prüfen, ob die für die theoretische Fundierung und die empirische Qualitätsprüfung erforderlichen Informationen verfügbar sind. Diese Informationen sollen in den Verfahrenshinweisen (auch Testhandbuch oder -manual genannt) stehen. Welche Informationen für eine fachgerechte Testbeurteilung zwingend erforderlich sind, ist in der DIN 33430 (2016) eindeutig festgelegt. Obwohl sich die DIN 33430 auf die berufsbezogene Eignungsdiagnostik bezieht, sind diese Anforderungen auf Fragebögen und Tests aus allen Bereichen anzuwenden. Die DIN SCREEN – Checkliste operationalisiert diese Anforderungen. Der hier vorliegende Auszug aus der Checkliste dient als »Standard des Diagnostik- und Testkuratoriums (DTK) der Föderation Deutscher Psychologinnenvereinigungen hinsichtlich des Qualitätsanspruchs an Information und Dokumentation von Instrumenten zur Erfassung menschlichen Erlebens und Verhaltens« (kurz: DTK-Testinformationsstandard). Der Standard legt fest, welche Informationen zu einem Test oder Fragebogen bereitgestellt werden müssen. Insgesamt umfasst er 32 Muss-Anforderungen sowie 53 Soll-Anforderungen. Autorinnen und Autoren bzw. Anbieterinnen und Anbieter von Tests und Fragebögen, die die 32 Muss-Anforderungen erfüllen, können beim DTK ein „TBS-DTK-Transparenzzertifikat“ für ihr Verfahren beantragen (siehe DTK, 2023a).

Für die Qualitätsbewertung von Tests und Fragebögen anhand des Testbeurteilungssystems des DTK (TBS-DTK, 2023a) wird mithilfe der nachfolgenden Checkliste geprüft, ob die Angaben zur Anwendung und zum Verfahren alle Informationen enthalten, die gemäß DIN 33430 (2016) verpflichtend sind. Fehlen diese Informationen, gilt der Test oder Fragebogen ohne weitere Prüfung als „nicht anforderungsgerecht“ (DTK, 2023a).

Die erste Fassung des TBS-DTK wurde 2006 publiziert. 2009 wurde die zweite, 2018 die dritte und 2023 die vierte Fassung publiziert. Ein wichtiges Ziel bei der Erstellung der vierten Version des TBS-DTK-Systems war es sicherzustellen, dass das System auch für Fragebögen und Tests geeignet ist, die mithilfe von Algorithmen – etwa durch maschinelles Lernen oder künstliche Intelligenz – Personenkenwerte ermitteln (siehe DTK, 2023b). Dabei wurde deutlich, dass die bisherigen Bewertungskriterien auch für diese neuen Formen passend sind. Besonders die im TBS-DTK geforderte Transparenz hinsichtlich der verwendeten Datenbasis (z. B. zu Untersuchungs- und Normstichproben) bleibt ein zentrales Qualitätsmerkmal, auch wenn Algorithmen aus Trainingsdaten abgeleitet werden.

Für algorithmusbasierte Tests sind jedoch zusätzliche Informationen notwendig. Dazu gehören beispielsweise umfassende Angaben zur Entwicklung und Anwendung der eingesetzten Algorithmen. Ergänzt wurden außerdem Anforderungen zur Nachvollziehbarkeit der Ergebnisse, die durch solche Tests entstehen. Ein weiterer Schwerpunkt liegt auf dem Umgang mit möglichen Diskriminierungen,

die durch die Nutzung algorithmischer Verfahren auftreten können. Das TBS-DTK-System (2018) und die vorherige Version der DIN-Checkliste (Kersting, 2018) wurden um ein Addendum ergänzt, das 16 zusätzliche Anforderungen an die Verfahrenshinweise für messtheoretisch abgesicherte Tests und Fragebögen enthält, die Personenkennwerte mithilfe von Algorithmen bestimmen. Diese zusätzlichen Anforderungen gehören nicht zur DIN 33430 (2016). Die 16 Forderungen „AD1“ bis „AD16“ des Addendums sind aber Bestandteil des „DTK-Informationsstandards“. Hier und im Folgenden ist mit » DTK-Testinformationsstandard« stets die Kombination aus dem entsprechenden Auszug aus der DIN SCREEN Checkliste und dem Addendum gemeint.

Die nachfolgende Checkliste ist eine Teilmenge der „DIN SCREEN 4 – Checkliste zur Kontrolle und Optimierung der Qualität berufsbezogener Eignungsdiagnostik nach DIN 33430“ (Kersting, 2025), die alle Qualitätsanforderungen, die in der DIN 33430 (2016) formuliert sind, vollständig wiedergibt. Die hier dargestellte Checkliste umfasst nur die in der DIN 33430 (ebd.) formulierten Anforderungen an Handhabungshinweise (Anforderungen A1 bis A15) sowie an Verfahrenshinweise (Anforderungen B1 bis B54) zu Fragebögen und Tests.

### **Einsatz der Checkliste**

Bei der Nutzung der Checkliste wird für jede Qualitätsanforderung überprüft, ob sie erfüllt wurde. Das Ergebnis wird festgehalten, indem man entweder „ja“ (Anforderung erfüllt) oder „nein“ (Anforderung nicht erfüllt) markiert. Zusätzlich besteht die Möglichkeit, in der letzten Spalte der Checkliste die Quelle und die entsprechende Seitenzahl zu vermerken, aus der die Beurteilung abgeleitet wurde. Beurteilungsgrundlage für die Checkliste zum DTK-Testinformationsstandard sind die Handhabungs- und Verfahrenshinweise (das Testhandbuch). Und zwar die Version, die zum Zeitpunkt des Beginns der Rezension aktuell ist. Weitere – möglicherweise aktuellere – Informationen müssen nur dann einbezogen werden, wenn auch die Anwenderinnen und Anwender ausdrücklich auf diese Informationen hingewiesen werden und wenn diese Informationen für alle Anwenderinnen und -anwender zugänglich sind. Das DTK (2023a) weist ausdrücklich darauf hin, dass es ein wesentliches Qualitätsmerkmal eines Tests oder Fragebogens ist, dass sämtliche relevanten Informationen zum Test und zu seiner Anwendung zentral und zugänglich bereitgestellt werden und dass die Anwenderinnen und Anwender auf diese zentrale Informationsquelle hingewiesen werden. Daher können nur solche Informationen als Grundlage dienen, die von den Test anbietenden bzw. den Testautorinnen und -autoren allen Anwenderinnen und Anwendern zur Verfügung gestellt wurden. Bezüglich aller relevanten Informationen zum Test / Fragebogen stehen die Test anbietenden bzw. die Testautorinnen und -autoren in einer Bringschuld. Es ist nicht die Aufgabe von Anwenderinnen und Anwendern, zu prüfen, ob in anderen Quellen (z. B. Fachzeitschriftenartikeln, Abschlussarbeiten) relevante Informationen zu finden sind. Ist eine Einschätzung aus bestimmten Gründen nicht möglich, kann statt „ja“ oder „nein“ das Feld „nicht zu bewerten“ ausgewählt werden. Diese Option sollte jedoch nur in begründeten Ausnahmefällen genutzt werden. Falls Kommentare erforderlich sind, können diese in der Checkliste eingetragen werden oder aus Platzgründen an anderer Stelle – im zuletzt genannten Fall ist in der Checkliste auf die separaten Anmerkungen zu verweisen.

Bei Aussagen, die mit „und“ formuliert sind, muss die Art der Verknüpfung berücksichtigt werden. Ist „und“ durch eine Unterstreichung hervorgehoben, bedeutet dies, dass alle genannten Aspekte erfüllt sein müssen, um insgesamt ein „ja“ vergeben zu können. Ein Beispiel ist Aussage B23, bei der die Erläuterung zur Angemessenheit der Methode(n) zur Reliabilitätsbestimmung verschiedene Aspekte einbeziehen muss – nämlich die Art der Eignungsmerkmale, die angestrebte Entscheidungsart sowie die jeweiligen Bedingungen. Nur wenn sämtliche genannten Punkte berücksichtigt wurden, gilt die Anforderung als vollständig erfüllt. Verknüpfungen mit „und/oder“ werden ausdrücklich so benannt.

### **Verzweigungsfragen**

In 25 Fällen sind den eigentlichen Aussagen sogenannte Verzweigungsfragen vorangestellt. Diese Zeilen sind visuell hervorgehoben, um ihre besondere Funktion zu verdeutlichen. Durch diese Verzweigungsfragen wird gesteuert, dass nur die für den jeweiligen diagnostischen Einsatz relevanten Anforderungen der DIN 33430 (2016) geprüft werden. Zunächst wird die Verzweigungsfrage mit „ja“ oder „nein“ beantwortet. Anschließend ist festgelegt, wie nach dieser Antwort weiter vorzugehen ist. Wenn bestimmte Aussagen übersprungen werden sollen, wird präzise angegeben, bei welcher Aussage die Prüfung fortgesetzt wird. In Zeilen, die übersprungen werden müssen, dürfen keine Eintragungen vorgenommen werden. Zur besseren Orientierung sind alle Verzweigungsfragen nummeriert.

### **Hinweise zur Bewertung**

Die Checkliste enthält zusätzlich zu den Aussagen und Verzweigungsfragen drei „Hinweise“ (H1 bis H3), die auf inhaltliche Abhängigkeiten zwischen Anforderungen aufmerksam machen. Diese Hinweise beschreiben logische Verknüpfungen, die sich aus hierarchischen Beziehungen ergeben.

Ein Beispiel: Die DIN 33430 (2016) verlangt, dass die Verfahrenshinweise für Tests und Fragebogen Ergebnisse einer oder mehrerer empirischer Untersuchungen wiedergeben (Anforderung B3). Auf dieser Grundlage bauen elf weitere Anforderungen (B4 bis B13) auf, die eine nachvollziehbare Beschreibung dieser Untersuchungen verlangen – beispielsweise muss das Erhebungsjahr der Untersuchungen genannt werden (Anforderung B5). Wenn jedoch die grundlegende Anforderung B3 nicht erfüllt ist, können die darauf basierenden Anforderungen ebenfalls nicht erfüllt werden. Hinweis H1 legt daher fest, dass in einem solchen Fall die Anforderungen B4 bis B13 automatisch mit „nein“ zu bewerten sind.

Die Hinweise zeigen, dass das Fehlen bestimmter Informationen zwangsläufig weitere Mängel nach sich zieht. Diese abhängigen Aussagen dürfen daher nicht übersprungen werden; sie sind explizit als „nicht erfüllt“ zu markieren. Im Gegensatz dazu führen Verzweigungsfragen (siehe oben) dazu, dass bestimmte Aussagen für den jeweiligen Fall gar nicht relevant sind und deshalb ausgelassen werden – ohne Eintragung.

Schließlich enthält die Checkliste noch Fußnoten, die das Verständnis erleichtern, indem sie Fachbegriffe erklären oder Praxisbeispiele liefern. Diese stammen entweder direkt aus der DIN 33430 (2016) oder sind ergänzende Erläuterungen des Autors.

### **Qualitätsprüfung, Qualitätssicherung und Qualitätsentwicklung**

Mit der Checkliste lässt sich überprüfen, ob alle Informationen, die nach DIN 33430 (2016) zu einem Fragebogen oder Test vorliegen müssen, auch vorliegen. Die Anforderungen A1 bis A15 und B1 bis B54 spiegeln die Informationsvorgaben der Anhänge A und B der Norm wider. Obwohl sich die DIN 33430 (2016) auf die berufsbezogene Eignungsdiagnostik bezieht, sind diese Anforderungen nach Ansicht des DTK auf Tests und Fragebogen aus allen Bereichen anwendbar – ein Punkt, der für die Nutzung der Checkliste von besonderer Bedeutung ist. Die Anforderungen AD1 bis AD16 wurden vom DTK ergänzt (DTK, 2023a); sie gelten zusätzlich für Verfahren, bei denen z. B. Personenkenwerte algorithmisch geschätzt werden.

Jedes „nein“ markiert ein Informationsdefizit. Für die Einhaltung des Standards sind strikt genommen nur die Muss-Anforderungen entscheidend. Wird eine Soll-Anforderung nicht erfüllt, bedeutet dies noch keine Abweichung von dem Standard, weist aber auf mögliche Verbesserungen hin.

### **Interpretation der Ergebnisse**

Die Checkliste kann quantitativ und qualitativ ausgewertet werden. Eine quantitative Betrachtung – etwa die Ermittlung der Erfüllungsquote in Prozent – ermöglicht eine erste Einschätzung. Eine allein zahlenbasierte Beurteilung ist jedoch nicht ausreichend. Die Relevanz einzelner Anforderungen variiert je nach Fragestellung und Einsatzkontext. Eine niedrige Erfüllungsquote kann trotz Einhaltung entscheidender Qualitätsmerkmale auftreten. Umgekehrt kann ein hoher Erfüllungsgrad bestehende erhebliche Mängel verdecken, sofern diese besonders kritische Aspekte betreffen.

Eine fundierte Gesamtbewertung erfordert deshalb eine inhaltliche Einordnung, die der Bedeutung jeder einzelnen Anforderung im jeweiligen diagnostischen Kontext gerecht wird. Die Anwendung der Checkliste setzt daher Fachwissen und eine reflektierte Bewertung voraus. Gleichzeitig gibt es Mindestkriterien, die zwingend erfüllt sein müssen – dies sind die „Muss-Kriterien“.

Die Erfüllung der Anforderungen A1 bis A15, B1 bis B54 sowie AD1 bis AD16 zeigt lediglich, dass die Verfahrenshinweise ausreichend Informationen enthalten. Ein hoher Informationsgrad allein garantiert jedoch nicht die Qualität des diagnostischen Verfahrens. Neben dem Informationsgehalt muss auch die eigentliche Verfahrensqualität beurteilt werden. Das DTK hat dafür ein eigenes Testbewertungssystem entwickelt (DTK, 2023a). Die Checkliste dient im ersten Schritt zur Prüfung, ob alle relevanten Informationen vorliegen; erst danach folgt eine separate Qualitätsbeurteilung anhand der Testbeurteilungsrichtlinien des DTK (2023a). Liegen alle notwendigen Informationen vorliegen; erst danach folgt eine separate Qualitätsbeurteilung anhand der Testbeurteilungsrichtlinien des DTK (2023a). Liegen alle notwendigen Informationen vor, erfüllt das Verfahren die Anforderungen an die Transparenz. Entsprechend des sogenannten „TBS-DTK-Transparenzzertifikats“ bedeutet dies: „Die Verfahrenshinweise zum Test erfüllen den Qualitätsanspruch des Diagnostik- und Testkuratoriums (DTK) an Information und Dokumentation von Instrumenten zur Erfassung menschlichen Erlebens und Verhaltens.“ (DTK, 2023a, S. 43). Damit erfüllt der jeweilige Test nach Ansicht des DTK die Voraussetzungen, um im nächsten Schritt einer Qualitätskontrolle unterzogen zu werden.

Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS): <https://www.gesis.org/zis>

## Fazit

Die vorliegende Checkliste bildet den „Standard des Diagnostik- und Testkuratoriums (DTK) der Föderation Deutscher Psychologinnenvereinigungen hinsichtlich des Qualitätsanspruchs an Information und Dokumentation von Instrumenten zur Erfassung menschlichen Erlebens und Verhaltens“. Sie bildet die Informationsanforderungen an Fragebogen und Tests vollständig ab und stellt sie in überprüfbarer Form bereit. Sie dient damit als Maßstab zur Bewertung des Informationsgehalts von Verfahrenshinweisen (Testhandbüchern, -manualen). Sie eignet sich sowohl zur nachträglichen Qualitätsbewertung als auch zur Planung und Gestaltung von Verfahrenshinweisen. Damit leistet die Checkliste nicht nur einen Beitrag zur Qualitätsprüfung, sondern wirkt auch prozesssteuernd und qualitätsfördernd.

## Literatur

- Diagnostik- und Testkuratorium (2018). TBS-DTK. Testbeurteilungssystem des Diagnostik- und Testkuratoriums. Revidierte Fassung vom 3. Januar 2018. *Report Psychologie*, 43 (3), 106–113 sowie *Psychologische Rundschau* (2018), 69, 109-148.
- Diagnostik- und Testkuratorium (DTK). (2023a). TBS-DTK – Testbeurteilungssystem des Diagnostik- und Testkuratoriums (TBS-DTK). Vierte, revidierte Fassung vom 31. Juli 2023. *Report Psychologie*, 48, Heft 11/12, S. 18-27 sowie *Psychologische Rundschau*, online veröffentlicht am 8. Dezember 2023. <https://doi.org/10.1026/0033-3042/a000653>, Print: 75, S. 39-49.
- Diagnostik- und Testkuratorium (DTK). (2023b). Qualitätssicherung und -optimierung von Tests. Die vierte Fassung des Testbeurteilungssystems des Diagnostik- und Testkuratoriums (TBS-DTK). *Report Psychologie*, 48, Heft 11/12, S. 14-17. <https://doi.org/10.1026/0033-3042/a000657>
- DIN. (2016). *DIN 33430: Anforderungen an berufsbezogene Eignungsdiagnostik*. Berlin: Beuth. <https://dx.doi.org/10.31030/2514220>
- 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. <https://doi.org/10.1007/978-3-662-53772-5>
- Kersting, M. (2025). DIN SCREEN 4 – Checkliste zur Kontrolle und Optimierung der Qualität berufsbezogener Eignungsdiagnostik nach DIN 33430. Gießen. <https://doi.org/10.22029/jlupub-20140>

**Zitation des Messinstruments:****Abschnitt A (normativ): Anforderungen an Handhabungshinweise für Verfahren**

		Soll (S) oder Muss (M)	Ja	Nein	Nicht zu bewerten	Anmerkungen	Quelle (Seite)
A1	In den Handhabungshinweisen ist die Zielsetzung des Verfahrens verständlich beschrieben	S					
A2	In den Handhabungshinweisen sind die Anwendungsbereiche verständlich benannt <sup>1</sup>	S					
V1	<i>Sind missbräuchliche Anwendungen des Verfahrens zur Eignungsbeurteilung nahe liegend? Falls „nein“ -&gt; Bitte weiter bei Verzweigungsfrage V2</i>						
A3	Die Handhabungshinweise enthalten spezifische warnende Hinweise vor missbräuchlicher Anwendung zur Eignungsbeurteilung	S					
V2	<i>Erfordert die Handhabung des Verfahrens besondere Qualifikationen? Falls „nein“ -&gt; Bitte weiter bei Anforderung A5</i>						
A4	Die für die Handhabung des Verfahrens erforderlichen besonderen Qualifikationen sind genannt	M					

<sup>1</sup> zu Anforderung A2: Es sollte z. B. angegeben sein, bei welcher Personengruppe (z. B. Bildungsstand) das Verfahren eingesetzt werden kann.

		Soll (S) oder Muss (M)	Ja	Nein	Nicht zu bewerten	Anmerkungen	Quelle (Seite)
	Die Handhabungshinweise liefern Informationen, aus denen der Anwender den hinsichtlich der folgenden Aspekte entstehenden Aufwand abschätzen kann:						
A5	<ul style="list-style-type: none"> <li>Materialien</li> </ul>	S					
A6	<ul style="list-style-type: none"> <li>Personal</li> </ul>	S					
A7	<ul style="list-style-type: none"> <li>Räumlichkeiten</li> </ul>	S					
	Die Handhabungshinweise liefern Informationen, aus denen der Anwender den hinsichtlich der folgenden Aspekte entstehenden zeitlichen Aufwand abschätzen kann:						
A8	<ul style="list-style-type: none"> <li>für den Kandidaten</li> </ul>	S					
A9	<ul style="list-style-type: none"> <li>für den Anwender bei der Routinevorbereitung</li> </ul>	S					
A10	<ul style="list-style-type: none"> <li>für den Anwender bei der Durchführung</li> </ul>	S					
A11	<ul style="list-style-type: none"> <li>für den Anwender bei der Auswertung</li> </ul>	S					

		Soll (S) oder Muss (M)	Ja	Nein	Nicht zu bewerten	Anmerkungen	Quelle (Seite)
V3	<i>Gibt es eine Interaktion des Anwenders mit den Kandidaten? Falls „nein“ -&gt; Bitte weiter bei Anforderung A13</i>						
A12	Die Handhabungshinweise beinhalten verständliche Instruktionen für den Kandidaten, die dazu beitragen, die Wahrscheinlichkeit von Nachfragen zu verhindern <sup>2</sup>	S					
	Die Handhabungshinweise sind so gestaltet, dass verschiedene Personen mit den erforderlichen Qualifikationen in der Lage sind, die Verfahren allein aufgrund dieser Handhabungshinweise auf die gleiche Art und Weise:						
A13	• durchzuführen	M					
A14	• auszuwerten	M					
A15	• und deren Ergebnisse zu interpretieren	M					

<sup>2</sup> zu Anforderung A12: Beispiele für häufige, aber (durch entsprechende Instruktionen zu Beginn des Verfahrens) vermeidbare Nachfragen: Darf man sich Notizen machen? Wird die zur Verfahrensbearbeitung zur Verfügung stehende Zeit bekannt gegeben? Darf man Teilaufgaben überspringen? Gibt es Minuspunkte bzw. Abzüge für falsche Antworten?

**Abschnitt B (normativ): Anforderungen an Verfahrenshinweise für messtheoretisch fundierte Fragebögen und Tests****B.1 Allgemeine Anforderungen**

		Soll (S) oder Muss (M)	Ja	Nein	Nicht zu bewerten	Anmerkungen	Quelle (Seite)
B1	Die theoretischen Grundlagen des Verfahrens sind beschrieben	M					
B2	In den Verfahrenshinweisen ist angemessen <sup>3</sup> dargestellt, wie das standardisierte Verfahren konstruiert wurde <sup>4</sup>	M					
B3	In den Verfahrenshinweisen werden die Ergebnisse einer oder mehrerer empirischen / empirischer Untersuchung(en) berichtet	M					
H1	<i>Hinweis: Falls „nein“ gewählt wird, muss bei den Anforderungen B4 bis B13 ebenfalls „nein“ angekreuzt werden</i>						
B4	Alle in den Verfahrenshinweisen aufgeführten relevanten empirischen Untersuchungen sind nachvollziehbar beschrieben / dokumentiert	M					

<sup>3</sup>zu Anforderung B2: im Sinne von ausführlich, verständlich und nachvollziehbar.

<sup>4</sup>zu Anforderung B2: z. B. wird erläutert, wie und warum die Fragen eines Fragebogens oder die Aufgaben eines Tests ausgewählt oder konstruiert wurden.

		Soll (S) oder Muss (M)	Ja	Nein	Nicht zu bewerten	Anmerkungen	Quelle (Seite)
	Der Bericht über empirische Untersuchungen enthält:						
B5	• eine Angabe über das Jahr der Datenerhebung	M					
B6	• deskriptive Statistiken über die Merkmale der Untersuchungsteilnehmer <sup>5</sup>	M					
B7	• Angaben, mit welchem Ziel der Test von Teilnehmern bearbeitet wurde <sup>6</sup>	M					
B8	• Angaben, ob die Datenerhebung unter Aufsicht oder unter nicht kontrollierten Bedingungen <sup>7</sup> stattgefunden hat	M					
B9	• Angaben, ob und falls ja <sup>8</sup> , wie die Teilnahme „belohnt“ <sup>9</sup> wird	M					
B10	• Informationen über den Stichprobenplan	S					
B11	• Informationen zu den Teilnehmerquoten	S					

<sup>5</sup>zu Anforderung B6: z. B. Angaben zu Alter, Geschlecht, Bildung, Status (z. B. Schüler, Studenten, Azubis, Berufstätige usw.).

<sup>6</sup>zu Anforderung B7: z. B. ohne für die Teilnehmer relevantes Ziel, zum Zwecke der persönlichen Orientierung oder im Zusammenhang mit Personalentscheidungen.

<sup>7</sup>zu Anforderung B8: z. B. über das Internet von „zu Hause“ aus.

<sup>8</sup>zu Anforderung B9: „Ob“ und „wie“ bedeutet, dass die Teilnahme z. B. ergebnisabhängig belohnt wird.

<sup>9</sup>zu Anforderung B9: „Belohnt“ bedeutet z. B. vergütet.

		Soll (S) oder Muss (M)	Ja	Nein	Nicht zu bewerten	Anmerkungen	Quelle (Seite)
B12	Die Dokumentation / der Bericht der empirischen Arbeit folgt den üblichen Kriterien für wissenschaftliche Publikationen <sup>10</sup>	S					
B13	Die Anzahl der in den empirischen Studien untersuchten Personen ist für die jeweilige Fragestellung <sup>11</sup> angemessen	M					
V4	<i>Ist mit einer Verfälschung des Verfahrens zu rechnen? Falls „nein“ -&gt; Bitte weiter bei Verzweigungsfrage V5</i>						
B14	Es ist aufgeführt, ob und falls ja, wie einer Verfälschung durch die Art der Verfahrensvorgabe <u>und</u> -durchführung – sowie ggf. auch bei der Auswertung – entgegengewirkt werden kann	S					
V5	<i>Erfolgt die Auswertung manuell? Falls „nein“ -&gt; Bitte weiter bei Verzweigungsfrage V6</i>						
B15	In den Verfahrenshinweisen sind Regeln aufgestellt, wie bei der Auswertung mit nicht bearbeiteten Fragen bzw. (Teil-) Aufgaben umgegangen wird	M					

<sup>10</sup>zu Anforderung B12: Erläuterung: siehe z. B. Deutsche Gesellschaft für Psychologie (2016), sowie Wilkinson und APA Task Force on Statistical Inference (1999). Dabei gilt jeweils die letzte Ausgabe dieser Publikation.

<sup>11</sup>zu Anforderung B13: z. B. Berechnung von Normwerten, erwartbaren Effektstärken.

		Soll (S) oder Muss (M)	Ja	Nein	Nicht zu bewerten	Anmerkungen	Quelle (Seite)
V6	<i>Handelt es sich um ein Verfahren, welches einen Vergleich mit Normwerten anbietet? Falls „nein“ -&gt; Bitte weiter bei Anforderung B22</i>						
B16	Die Bezugsgruppe, an der die Normdaten gewonnen wurden, entspricht hinsichtlich zentraler Merkmale <sup>12</sup> der Personengruppe, für die das Verfahren laut Verfahrenshinweisen eingesetzt wird / werden soll oder es wird nachgewiesen, dass die vorhandenen Normdaten für die Zielgruppe verwendet werden können	M					
B17	Die Angemessenheit der Normwerte wurde in den letzten acht Jahren überprüft <sup>13</sup>	S					
V7	<i>Wurde die Angemessenheit der Normwerte in den letzten acht Jahren überprüft? Falls „ja“-&gt; Bitte weiter bei Verzweigungsfrage V8</i>						
B18	Es wird begründet, warum <u>und</u> unter welchen Umständen das Verfahren für einen Anwendungsfall ausgewählt werden kann, obwohl die Angemessenheit der Normwerte nicht in den letzten acht Jahren überprüft wurde <sup>14</sup>	M					

<sup>12</sup>zu Anforderung B16: z. B. Alter, Bildungsstand, Berufserfahrung. Eine solche Entsprechung liegt z. B. nicht vor, wenn etwa Englischkenntnisse von Managern untersucht werden sollen, die Normwerte zum Verfahren aber an Schülern gewonnen wurden.

<sup>13</sup>zu Anforderung B17: Es geht hierbei nur um eine Überprüfung der Angemessenheit der Normwerte. Ob eine Neunormierung durchgeführt werden muss, ergibt sich in Abhängigkeit von den Ergebnissen der Überprüfung. In der DIN 33430 (2016) wird nicht gefordert, dass spätestens alle acht Jahre neu normiert werden muss.

<sup>14</sup>zu Anforderung B18: Die Anforderung lautet im Text der DIN 33430 (2016, S. 29): „Wurde die Angemessenheit der Normwerte in den letzten acht Jahren nicht überprüft, muss begründet werden, warum das Verfahren dennoch ausgewählt wird.“ Diese Begründung obliegt den Eignungsdiagnostikern. Die in der vorliegenden Checkliste gewählte Formulierung richtet sich an die Verantwortlichen für die Verfahrenshinweise.

		Soll (S) oder Muss (M)	Ja	Nein	Nicht zu bewerten	Anmerkungen	Quelle (Seite)
V8	<i>Zielt das Verfahren auf die Erfassung eines Eignungsmerkmals ab, dessen Ausprägung in der Referenzgruppe möglicherweise relativ kurzfristigen Veränderungen unterliegt?<sup>15</sup> Falls „nein“-&gt; Bitte weiter bei Verzweigungsfrage V9</i>						
B19	Die Angemessenheit der Normwerte wurde bereits vor Ablauf der acht-Jahres-Frist empirisch gezeigt	M					
V9	<i>Besteht für den Verfahrensanwender die Möglichkeit, die Werte einer Person anhand unterschiedlicher Normgruppen (auch gruppenspezifische Normen genannt) zu bewerten?<sup>16</sup> Falls „nein“ -&gt; Bitte weiter bei Anforderung B22</i>						
B20	Zur Sicherung der Interpretationsobjektivität sind eindeutige Hinweise gegeben, wie die Entscheidung, welche Normgruppe in welchem Fall heranzuziehen ist, zu treffen ist	S					
B21	Die Effekte der Anwendung dieser gruppenspezifischen Normen <sup>17</sup> sind nachvollziehbar erläutert	S					

<sup>15</sup>zu Verzweigungsfrage V8: z. B. EDV-Kenntnisse.

<sup>16</sup>zu Verzweigungsfrage V9: Unterschiedliche Normgruppen wären z. B. bildungsspezifische und bildungsunspezifische Normen.

<sup>17</sup>zu Anforderung B21: z. B. Bildungsnormen.

**B.2 Zuverlässigkeit**

		Soll (S) oder Muss (M)	Ja	Nein	Nicht zu bewerten	Anmerkungen	Quelle (Seite)
B22	In den Verfahrenshinweisen werden Angaben zur Zuverlässigkeit des Verfahrens gemacht, die aus empirischen Studien abgeleitet wurden	M					
H2	<i>Hinweis: Falls keine Angaben zur Zuverlässigkeit des Verfahrens gemacht werden, ist die Anforderung B22 mit „nein“ zu beantworten</i>						
V10	<i>Werden Angaben zur Zuverlässigkeit des Verfahrens gemacht? Falls „nein“ -&gt; Bitte weiter bei Anforderung B27</i>						
B23	Die Angemessenheit der für die Zuverlässigkeitsbestimmung genutzten Methode(n) wird erläutert <sup>18</sup> . Die Erläuterung berücksichtigt die Art der untersuchten Eignungsmerkmale und der angestrebten Entscheidung ebenso wie die jeweiligen Anwendungs- <u>und</u> Untersuchungsbedingungen	S					
V11	<i>Sollen mit dem Verfahren Merkmale erfasst werden, für die eine zumindest relative Zeit- <u>und</u> Situationsstabilität angenommen wird? Falls „nein“ -&gt; Bitte weiter bei Anforderung B27</i>						
B24	Die Zuverlässigkeit wurde über die Retest-Methode bestimmt oder die Retest-Reliabilität wurde durch einen geeigneten Untersuchungsplan geschätzt	S					

<sup>18</sup>zu Anforderung B23: Die Bestimmung der internen Konsistenz ist z. B. keine angemessene Art der Zuverlässigkeitsbestimmung für Verfahren mit heterogenen Inhalten; die Bestimmung der Retest-Reliabilität ist keine angemessene Art der Zuverlässigkeitsbestimmung für Verfahren zur Messung rasch veränderlicher Merkmale (z. B. Stimmungen).

		Soll (S) oder Muss (M)	Ja	Nein	Nicht zu bewerten	Anmerkungen	Quelle (Seite)
B25	Der aktuellste Nachweis der Geltung der Zuverlässigkeitskennwerte ist jünger als acht Jahre	S					
V12	<i>Ist eine Überprüfung der Geltung der Zuverlässigkeitskennwerte erforderlich, da in den letzten acht Jahren kein Überprüfung stattgefunden hat? Falls „nein“ -&gt; bitte weiter bei Anforderung B27</i>						
B26	Es wird begründet, warum <u>und</u> unter welchen Umständen das Verfahren für einen Anwendungsfall gewählt werden kann, obwohl die Geltung der Zuverlässigkeitskennwerte in den letzten acht Jahren nicht überprüft wurde <sup>19</sup>	M					

<sup>19</sup>zu Anforderung B26: Die Anforderung lautet im Text der DIN 33430 (2016, S. 29): „Wurden die Zuverlässigkeitskennwerte in den letzten 8 Jahren nicht überprüft, muss begründet werden, warum das Verfahren dennoch ausgewählt wird.“ Diese Begründung obliegt den Eignungsdiagnostikern. Die in der vorliegenden Checkliste gewählte Formulierung richtet sich an die Verantwortlichen für die Verfahrenshinweise.

**B.3 Gültigkeit****B.3.1 Allgemeine Anforderungen**

		Soll (S) oder Muss (M)	Ja	Nein	Nicht zu bewerten	Anmerkungen	Quelle (Seite)
B27	In den Verfahrenshinweisen werden Angaben zur Gültigkeit des Verfahrens gemacht, die aus empirischen Studien abgeleitet wurden	M					
H3	<i>Hinweis: Falls keine empirisch fundierten Angaben zur Gültigkeit des Verfahrens gemacht werden, sind die Anforderung B27 - B34 und B40 - B41 mit „nein“ zu beantworten</i>						
B28	Aus den Verfahrenshinweisen wird deutlich, welche empirischen Nachweise der Inhalts- und / oder Kriteriums- und / oder Konstruktgültigkeit eine Anwendung des Verfahrens bzw. der Verfahrensklasse für den laut Verfahrenshinweisen intendierten Anwendungszweck des Verfahrens rechtfertigen	M					
	In den Verfahrenshinweisen wird angegeben, welche Gültigkeitswerte:						
B29	<ul style="list-style-type: none"> <li>in Bezug zu welchem Kriterium<sup>20</sup> erzielt wurden</li> </ul>	M					
B30	<ul style="list-style-type: none"> <li>für welche Referenzgruppen erzielt wurden</li> </ul>	M					

<sup>20</sup>zu Anforderung B29: z. B. Bewährungskriterium.

		Soll (S) oder Muss (M)	Ja	Nein	Nicht zu bewerten	Anmerkungen	Quelle (Seite)
	In den Verfahrenshinweisen wird angegeben, welche Gültigkeitswerte:						
B31	• in welcher Untersuchung erzielt wurden	M					
B32	• für welches Verfahrensergebnis erzielt wurden <sup>21</sup>	M					
B33	• zu welchem Zeitpunkt erzielt wurden	M					
B34	Der aktuellste Nachweis über die Gültigkeit des Verfahrens ist jünger als acht Jahre	S					
V14	<i>Wurden zur Bestimmung der Gültigkeit Methoden der statistischen Adjustierung / Optimierung eingesetzt<sup>22</sup>? Falls „nein“ -&gt; Bitte weiter bei Verzweigungsfrage V15</i>						
	Bei der Dokumentation der Analysen zur Gültigkeit:						
B35	• sind sowohl die ursprünglich erhaltenen als auch die korrigierten Kennwerte aufgeführt	M					
B36	• werden alle in Zusammenhang mit der Adjustierung verwendeten Statistiken genannt	M					

<sup>21</sup>zu Anforderung B32: Bezieht sich der Gültigkeitswert z. B. auf das Gesamtergebnis oder auf ein Teilergebnis (etwa auf eine einzelne Skala oder einzelne Items)? Bezieht sich der Gültigkeitswert auf einen Rohwert oder auf einen standardisierten Wert?

<sup>22</sup>zu Verzweigungsfrage V14: z. B. Minderungskorrektur, Varianzeinschränkungskorrektur, multiple Regression.

		Soll (S) oder Muss (M)	Ja	Nein	Nicht zu bewerten	Anmerkungen	Quelle (Seite)
	Bei der Dokumentation der Analysen zur Gültigkeit:						
B37	• werden neben den statistisch optimierten Schätzungen auch die einfachen Schätzungen angegeben <sup>23</sup>	M					
B38	• werden die optimierten Schätzungen auf eine andere Personengruppe aus dem Geltungsbereich des Verfahrens angewendet <u>und</u> in ihrer Gültigkeit bestätigt <sup>24</sup>	S					
B39	• sind die statistischen Optimierungen in handlungsleitende Beurteilungsregeln umgesetzt <sup>25</sup>	S					
V15	<i>Wird der Gültigkeitsanspruch damit begründet, dass Gültigkeitshinweise aus anderen Untersuchungen in Anspruch genommen werden<sup>26</sup>? Falls „nein“-&gt; Bitte weiter bei Anforderung B42</i>						
	Es wird nachvollziehbar aufgeführt:						
B40	• welche Befunde generalisiert werden können <sup>27</sup>	S					
B41	• weshalb <sup>28</sup> sich die Gültigkeitshinweise übertragen lassen, die sich aus anderen Studien ergeben	S					

<sup>23</sup> zu Anforderung B37: statistisch optimierte Schätzungen sind z. B. multiple Regressionen; einfache Schätzungen sind z. B. einfache Korrelationen.

<sup>24</sup> zu Anforderung B38: z. B. Kreuzvalidierung.

<sup>25</sup> zu Anforderung B39: Wenn z. B. gezeigt wird, dass die multiple Vorhersagbarkeit eines Kriteriums unter Einbezug mehrerer Prädiktoren (z. B. mehrere Skalen eines Tests) deutlich höher ist als die einfachen Korrelationen zwischen einzelnen Prädiktoren und diesem Kriterium, so sollte dem Anwender erläutert werden, wie er die verschiedenen Prädiktoren so kombinieren / gewichten kann, dass der Vorteil praktisch nutzbar wird.

<sup>26</sup> zu Verzweigungsfrage V75: z. B. Validitätsgeneralisierung.

<sup>27</sup> zu Anforderung B40: Dies erfolgt durch Darstellung der entsprechenden Studien, Literaturübersichten und Metaanalysen.

<sup>28</sup> zu Anforderung B41: und in welchem Ausmaß.

**B.3.2 Konstruktgültigkeit**

		Soll (S) oder Muss (M)	Ja	Nein	Nicht zu bewerten	Anmerkungen	Quelle (Seite)
B42	Aufgrund von inhaltlichen Überlegungen wird dargelegt, wie sich das fragliche Konstrukt zu ähnlichen Konstrukten verhält <sup>29</sup>	S					
B43	Aufgrund von empirischen Ergebnissen wird dargelegt, wie sich das fragliche Konstrukt zu ähnlichen Konstrukten verhält <sup>29</sup>	S					
B44	Aufgrund von inhaltlichen Überlegungen wird dargelegt, wie sich das fragliche Konstrukt zu unähnlichen Konstrukten verhält <sup>30</sup>	S					
B45	Aufgrund von empirischen Ergebnissen wird dargelegt, wie sich das fragliche Konstrukt zu unähnlichen Konstrukten verhält <sup>30</sup>	S					
<sup>29</sup> zu Anforderung B42 - B43: Dies ist die konvergente Gültigkeit. <sup>30</sup> zu Anforderung B44 - B45: Dies ist die diskriminante Gültigkeit.							

**B.3.3 Kriteriumsgültigkeit**

		Soll (S) oder Muss (M)	Ja	Nein	Nicht zu bewerten	Anmerkungen	Quelle (Seite)
B46	Bei der Analyse der Kriteriumsgültigkeit des Verfahrens wird beschrieben, warum das in der Analyse jeweils verwendete Kriterium angemessen ist <u>und</u> valide erfasst wurde	M					
B47	Sowohl die Objektivität als auch die Zuverlässigkeit jedes verwendeten Kriterienmaßes wird nach Möglichkeit dargestellt	S					
B48	Die Angemessenheit der für die Analyse der Kriteriumsgültigkeit herangezogenen Untersuchungsgruppe wird erläutert <sup>31</sup>	M					

<sup>31</sup> zu Anforderung B48: z. B. sollten die demografischen Merkmale der Untersuchungsgruppe (z. B. Bildungsstand, Alter, Berufserfahrung usw.) vor dem Hintergrund der als Zielgruppe des Verfahrens genannten Gruppe diskutiert werden.

**B.3.4 Inhaltsgültigkeit (sofern für das jeweilige Verfahren relevant)**

		Soll (S) oder Muss (M)	Ja	Nein	Nicht zu bewerten	Anmerkungen	Quelle (Seite)
V16	<i>Wird für das Verfahren Inhaltsgültigkeit in Anspruch genommen? Falls „nein“ -&gt; Bitte weiter mit Verzweigungsfrage V18</i>						
B49	Der im Verfahren abgebildete Inhaltsbereich wird nachvollziehbar beschrieben	S					
B50	Die Kriterien zur Beschreibung des dem Verfahren zugrunde liegenden, hypothetischen Itemuniversums sind angegeben	S					
B51	Die Regeln, nach denen das Verfahren als systematisch zusammengestellte Itemstichprobe aus dem Itemuniversum abgeleitet wurde, werden dargestellt	S					
V17	<i>Wird die Frage, ob das Verfahren den definierten Inhaltsbereich repräsentiert, durch Experten beurteilt? Falls „nein“ -&gt; Bitte weiter mit Verzweigungsfrage V18</i>						
	Es wird:						
B52	<ul style="list-style-type: none"> <li>sowohl der fachbezogene Ausbildungsstand <u>und</u> die Erfahrung als auch die Qualifikation der beteiligten Experten beschrieben</li> </ul>	S					
B53	<ul style="list-style-type: none"> <li>erläutert, wie die Experten zu ihrer Einschätzung gekommen sind</li> </ul>	S					
B54	<ul style="list-style-type: none"> <li>angegeben, inwieweit die Expertenbeurteilungen übereinstimmen</li> </ul>	S					

**Addendum: Zusätzliche Anforderungen an Verfahrenshinweise für messtheoretisch fundierte Fragebögen und Tests, die mittels Algorithmen Personenkenwerte schätzen**

Die Anforderungen B1 bis B54 an Verfahrenshinweise stammen aus der DIN 33430 (2016). Seit der Veröffentlichung der DIN 33430 haben sich Fragebögen und Tests etabliert, die mittels Algorithmen (die z. B. mithilfe von Modellen des maschinellen Lernens (ML) oder der künstlichen Intelligenz (KI) entwickelt wurden) Personenkenwerte schätzen. Die in der DIN 33430 formulierten Anforderungen – in der vorliegenden Checkliste als Forderungen B1 bis B54 operationalisiert – gelten uneingeschränkt auch für diese neuen, algorithmusbasierten Fragebögen und Tests. Besonders wichtig ist dabei die Transparenz der Datenbasis, z. B. im Hinblick auf die verwendeten Untersuchungs- oder Normgruppen (siehe z. B. Anforderungen B3 bis B13).

Allerdings reichen diese Anforderungen allein bei algorithmusbasierten Verfahren nicht aus. Es bedarf zusätzlicher Informationen, etwa zur Entwicklung und Anwendung der eingesetzten Algorithmen sowie zur Nachvollziehbarkeit der daraus resultierenden Bewertungen. Ein besonderes Augenmerk ist bei diesen Tests schließlich auf potenzielle Diskriminierungen zu richten.

Um diesen besonderen Anforderungen gerecht zu werden, hat das Diagnostik- und Testkuratorium (DTK) im Jahr 2023 sein Testbeurteilungssystem (TBS-DTK) um Beurteilungskriterien für Fragebögen und Tests ergänzt, die auf Algorithmen basieren. Die nachfolgend dargestellten Forderungen AD1 bis AD16 geben diese zusätzlichen Qualitätskriterien wider und ergänzen die bestehenden Anforderungen aus der DIN 33430.

## Addendum

		Soll (S) oder Muss (M)	Ja	Nein	Nicht zu bewerten	Anmerkungen	Quelle (Seite)
V18	Handelt es sich um ein Verfahren, das mittels Algorithmen <sup>32</sup> Personenkennwerte schätzt? Falls „nein“ -> Bitte weiter mit Verzweigungsfrage V19						
AD1	Es liegen nachvollziehbare Informationen über die Entstehung <u>und</u> die Nutzung der eingesetzten Algorithmen vor <sup>33</sup>	S					
V19	Handelt es sich um ein »Blackbox-Verfahren« <sup>34</sup> ? Falls „nein“ -> Bitte weiter bei Anforderung A3						
AD2	Es liegen Darstellungen von Analysen zu interpretierbarem ML bzw. Explainable Artificial Intelligence vor <sup>35</sup>	S					
AD3	Die zur Schätzung von Personenkennwerten eingesetzten Modelle <u>und</u> statistischen Schätzmethoden sind so dargestellt, dass die Anwender des Verfahrens den logischen Aufbau der Auswertung des Verfahrens nachvollziehen können	S					

<sup>32</sup>zu Verzweigungsfrage V18: Ein Algorithmus ist eine eindeutige Handlungsvorschrift, die z. B. mittels ML oder KI erstellt wurde und Daten verarbeitet, analysiert, Vorhersagen trifft oder Entscheidungsvorschläge unterbreitet.

<sup>33</sup>zu Anforderung AD1: z. B. genutzte Prädiktoren, Zusammensetzung von Trainings- oder Testdatensatz, Begründung des Analysemodells (z. B. Schätzverfahren linear vs. non-linear, mehrparametrisches Item-Response-Theorie (IRT) vs. Raschmodell), Aktualisierungsmaßnahmen.

<sup>34</sup>zu Verzweigungsfrage V19: Der Begriff "Blackbox-Verfahren" bezeichnet ein System oder Modell, dessen Funktionsweise nicht transparent ist.

<sup>35</sup> zu Anforderung AD2: z. B. Analysen zu den Fragen: Worauf basieren Vorhersagen? Welche Parameter gehen ein? Änderungen in welchen Wertebereichen der eingehenden Variablen verursachen welchen mittleren Effekt in der Vorhersage des Kriteriums? Solche Informationen können z. B. in Form von Model Cards dargeboten werden, die darüber hinaus auch über die Qualität der Entscheidungsvorschläge informieren (siehe Validität). Besonders positiv zu bewerten ist es, wenn die Testanbieter den Anwendern den Modellcode und / oder das fertig trainierte Modell für eigene Analysen zur Beurteilung der Angemessenheit des Modells (z. B. Verfahren des interpretierbaren ML; Überprüfung von Modellbias und Modellfairness) für die geplante Anwendung zur Verfügung stellen.

		Soll (S) oder Muss (M)	Ja	Nein	Nicht zu bewerten	Anmerkungen	Quelle (Seite)
V20	Werden zur Schätzung von Personenkennwerten Normdaten eingesetzt? Falls „nein“ -> Bitte weiter bei Verzweigungsfrage V21						
AD4	Die zur Schätzung von Personenkennwerten eingesetzten Normdaten sind so dargestellt, dass die Anwender des Verfahrens den logischen Aufbau der Auswertung des Verfahrens nachvollziehen können	S					
V21	Handelt es sich um ein Verfahren, bei denen die Algorithmen mit Hilfe von Trainingsdatensätzen gewonnen wurden? Falls „nein“ -> Bitte weiter bei Verzweigungsfrage V22						
AD5	Der Trainingsdatensatz ist, bezogen auf relevante Personenmerkmal, repräsentativ für jede angestrebte (Sub-)Population	S					
V22	Werden mit dem Verfahren Merkmale erfasst, für die eine zumindest relative Zeit- und Situationsstabilität angenommen wird, sodass gemäß Anforderung B24 ein Wert für die Retest-Reliabilität angegeben wird <sup>36</sup> ? Falls „nein“ -> Bitte weiter bei Verzweigungsfrage V23						
AD6	Das Modell ermöglicht eine Abschätzung der Messpräzision für individuelle Personenkennwerte <sup>37</sup>	S					

<sup>36</sup>zu Verzweigungsfrage V22: Dabei ist darauf zu achten, dass Angaben zur Accuracy eher die Modellgüte anzeigen und nicht die Reliabilität des zu interpretierenden Kennwerts.

<sup>37</sup>zu Anforderung AD6: z. B. durch die Angabe von Konfidenzintervallen.

		Soll (S) oder Muss (M)	Ja	Nein	Nicht zu bewerten	Anmerkungen	Quelle (Seite)
V23	Werden statistische Modelle eingesetzt <sup>38</sup> ? Falls „nein“ -> Bitte weiter bei Verzweigungsfrage V24						
AD7	Die Güte der statistischen Modelle ist mit Hilfe von für die entsprechende Modellklasse angemessenen Kennwerten <sup>39</sup> angegeben	S					
V24	Wurde das Modell kreuzvalidiert? Falls „nein“ -> Bitte weiter bei Anforderung AD9						
AD8	Es ist ausgewiesen, welche Daten für die Kreuzvalidierung des Modells genutzt wurden	S					
AD9	Es ist angegeben, auf welche Kriterien die Algorithmen trainiert wurden	S					
AD10	Es ist angegeben, welche Kriterien für die Schätzung der Kriteriumsvalidität genutzt wurden	S					

<sup>38</sup>Verzweigungsfrage V23: z. B. Strukturgleichungsmodelle, IRT-Modelle, ML- oder KI-Modelle.

<sup>39</sup>zu Anforderung AD7: Angemessene Kennwerte sind z. B., aber nicht ausschließlich, Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA), und Standardized Root Mean Square Residual (SRMR) bei Strukturgleichungsmodellen, Ergebnisse von Anderson- oder Martin-Löf-Test bei IRT Modellen und F1-Score, Sensitivität, Spezifität oder Root Mean Square Error (RMSE) bei KI-Modellen.

		Soll (S) oder Muss (M)	Ja	Nein	Nicht zu bewerten	Anmerkungen	Quelle (Seite)
V25	Werden entscheidungstheoretische Kennzahlen berichtet? Falls „nein“ -> Bitte weiter bei Anforderung AD12						
AD11	Der Bericht der entscheidungstheoretischen Kennzahlen erfolgt multiperspektivisch <sup>40</sup>	S					
AD12	Die für den jeweiligen Anwendungsfall relevanten Kennwerte <sup>41</sup> wurden empirisch ermittelt	S					
AD13	Die für den jeweiligen Anwendungsfall relevanten Kennwerte <sup>41</sup> werden in den Verfahrenshinweisen nachvollziehbar berichtet	S					
AD14	Es ist angegeben, ob sich die Daten zur Validierung der Verfahrens auf den Trainingsdatensatz und / oder auf einen Anwendungsdatensatz beziehen	S					

<sup>40</sup>zu Anforderung AD11: Multiperspektivische Betrachtung meint, dass z. B. nicht nur die Accuracy (Anteil von korrekten Vorhersagen an allen getroffenen Vorhersagen), sondern auch die Precision bzw. Sensitivität (Anteil der korrekten positiven Vorhersagen von allen als positiv klassierten Fällen) sowie die Spezifität (Anteil der korrekten negativen Vorhersagen von allen als negativ klassierten Fällen) berichtet werden.

<sup>41</sup>zu Anforderung AD12 und AD13: Im Kontext der Kriteriumsvalidierung relevante Kennwerte sind z. B. die Prädiktor-Kriteriums Korrelation, Trefferquoten, Sensitivität und Spezifität.

		Soll (S) oder Muss (M)	Ja	Nein	Nicht zu bewerten	Anmerkungen	Quelle (Seite)
AD15	In den Verfahrenshinweisen wird die Frage diskutiert, ob die laut den Verrechnungsvorschriften resultierenden Testwerte die empirischen Verhaltensrelationen adäquat <u>und</u> mit ausreichender Messpräzision abbilden	S					
AD16	In den Verfahrenshinweisen werden potenzielle Diskriminierungen diskutiert und die Ergebnisse von Bias-Checks dokumentiert <sup>42</sup>	S					

<sup>42</sup>Zu Anforderung AD16: Dies gilt für Variablen, für die es aufgrund theoretischer Überlegungen oder empirischer Befunde naheliegend ist, dass Sie Varianz in den erzeugen (z. B. Geschlecht, Dialekt bei Sprachanalysen, Piercing bei Gesichtsanalysen)