

Kersting, M., Häcker, H. & Hornke, L. F. (2011). Qualitätsstandards in der Diagnostik. In L. F. Hornke, M. Amelang & M. Kersting (Hrsg.), Enzyklopädie der Psychologie. Themenbereich B, Methodologie und Methoden. Serie II, Psychologische Diagnostik. Band 1, Grundfragen und Anwendungsfelder psychologischer Diagnostik (S. 1-86). Göttingen: Hogrefe.

Enzyklopädie der Psychologie
Themenbereich B, Methodologie und Methoden
Serie II, Psychologische Diagnostik

Grundfragen und Anwendungsfelder psychologischer Diagnostik

herausgegeben von

Prof. em. Dr. Lutz F. Hornke, Aachen
Prof. em. Dr. Manfred Amelang, Heidelberg
Prof. Dr. Martin Kersting, Münster

2011



Hogrefe • Verlag für Psychologie
Göttingen • Bern • Toronto • Seattle

1. Kapitel

Qualitätsstandards in der psychologischen Diagnostik

*Martin Kersting, Hartmut O. Häcker
und Lutz F. Hornke*

1 Qualitätsstandards in der psychologischen Diagnostik

Wer Qualität in der psychologischen Diagnostik fordert, kann sich solange ungebrochener Zustimmung erfreuen wie im Ungewissen bleibt, was unter Qualität verstanden wird. Der Begriff „Qualität“ ist seines inflationären Gebrauchs zum Trotz nicht trivial. In dem vorliegenden Beitrag werden zunächst die Ziele von Qualitätsstandards formuliert und unterschiedliche Qualitätsauffassungen definiert. Ausgewählte Qualitätsstandards werden im zweiten Abschnitt vorgestellt und in eine Systematik gebracht, indem zwischen verfahrensunspezifischen Qualitätsstandards, kompetenzorientierten Qualitätsansätzen, testbezogenen Qualitätsstandards und Beurteilungssystemen für Tests sowie Qualitätsstandards zu weiteren Verfahren (z. B. Assessment-Center und Interviews) unterschieden wird. Im Mittelpunkt des Beitrags stehen die nordamerikanischen „Standards for educational and psychological testing“ (AERA, APA & NCME, 1999) und die DIN 33430 (DIN, 2002). Abschnitt 3 gibt die DIN 33430 wieder und ordnet, soweit möglich, die nordamerikanischen Standards den DIN-Aussagen zu, eine Auswertung der Gemeinsamkeiten und Unterschiede der 1999 publizierten US-amerikanischen Standards und der DIN 33430 wird im vierten Abschnitt vorgenommen. Der fünfte Abschnitt zieht ein Fazit und wagt eine Prognose zur Zukunft der Qualitätssicherung und -optimierung im Bereich der Diagnostik.

Diagnostisches Handeln und Entscheiden erfordert Qualität, die Prozesse und Produkte (z. B. Tests) sollten bestimmten Anforderungen genügen. Qualitätsstandards formulieren Regeln und Prinzipien für die psychologische Diagnostik. Sie dienen u. a. als

- Ausgangspunkt eines Qualitätsmanagements,
- Leitfaden für die Entwicklung diagnostischer Verfahren,
- Leitfaden für die Planung, Durchführung und Evaluation des diagnostischen Vorgehens,
- Leitfaden für die Qualifizierung diagnostisch tätiger Personen,
- Maßstab zur Bewertung von Angeboten sowie Leistungen im Bereich der Diagnostik,
- Schutz der untersuchten Personen vor unsachgemäßer oder missbräuchlicher Diagnostik.

Es gibt zahlreiche Qualitätsstandards für die Diagnostik. Im vorliegenden Beitrag werden nur solche berücksichtigt, die als Konventionen zwischen Experten unter dem Dach einer oder mehrerer, nicht primär gewinnorientierter Organisationen ausgehandelt und für die Allgemeinheit zusammengestellt wurden. Beiträge einzelner Autoren oder Autorengruppen werden nicht einbezogen, wenngleich einzelne Beiträge oder Bücher durchaus Qualitätsstandards setzen. Die Beschränkung auf gemeinschaftlich und durch berufsständische Organisationen im weiteren Sinne erarbeitete Qualitätsstandards ergibt sich aus der Auffassung, dass die Definition von Qualität nicht so sehr von objektiven Erkenntnisprozessen abhängt, sondern primär eine Konvention zwischen Experten über Gegenstände und sinnvolles Verhalten darstellt (Kersting, 2009).

Qualitätsstandards kommt eine hohe Bedeutung zu, weil sie die diagnostischen Prozesse auf ein (Qualitäts-)Ziel ausrichten und (je nach Formulierung der Standards) den Erfolg bei der Zielerreichung messbar/nachprüfbar werden lassen. Diagnostik minderer Qualität ist häufig nicht auf mangelhaftes Bemühen, sondern auf mangelhaftes Feedback und mangelnde Nachvollziehbarkeit zurückzuführen. In der Praxis ist es beispielsweise häufig nicht möglich, die Reliabilität und Validität der eingesetzten diagnostischen Verfahren zu bestimmen, wenn es sich um eine Einzelfalldiagnostik handelt, für die einschlägige Forschungsergebnisse nicht zur Verfügung stehen und/oder nur einen begrenzten Stellenwert haben. Ob die aus der Diagnostik abgeleiteten Interventionsmaßnahmen ihr Ziel erreichen, zeigt sich vielleicht erst nach Jahren. Ein derart verzögertes Feedback kann nur schwer einen Anreiz für Qualität darstellen, denn die aktuelle Arbeit bleibt ohne Feedback und wird somit nicht in überschaubarer Zeit mit positiven oder negativen Wirkungen verbunden. Dies kann dazu führen, dass der Diagnostiker der Versuchung nachgibt, sich an schnell verfügbaren, aber eigentlich sekundären Kriterien zu orientieren (z. B. Akzeptanz oder soziale Validität der diagnostischen Verfahren, vgl. z. B. Kersting, 2008b; Schuler 1990, 1993; Schuler & Stehle, 1983). Aus Qualitätsstandards sollen sich konkrete Qualitätsforderungen für den Einzelfall ableiten lassen, so dass die Erfüllung oder Nichterfüllung im Sinne eines unmittelbaren Feedback-Systems be-

stimmt werden kann. Qualitätsstandards sind Leistungsstandards, die ein gerichtetes Verhalten motivieren.

Eine Quelle für Missverständnisse bezüglich Qualitätsstandards ist der Begriff „Qualität“, der umgangssprachlich anders genutzt werden kann als in der Fachsprache des Qualitätsmanagements. Umgangssprachlich wird Qualität häufig als „Güte“ definiert, womit in der Regel eine Wertung und häufig auch ein Vergleich verbunden wird. Demgegenüber wird Qualität im Kontext von Qualitätsmanagementsystemen (wie der Reihe ISO 9000 ff.) als Beschaffenheit definiert, womit zunächst keine Wertung verbunden ist. Zollondz (2006) gibt das Beispiel einer Wolledecke, die dick oder dünn, grob oder fein gewebt, weich oder hart beschaffen sein kann, ohne dass die Wolledecke deshalb eine bestimmte Güte aufweist. „Qualität als Beschaffenheit“ bezieht sich auf Qualitätsforderungen. Ein Qualitätsstandard ist somit ein Regelwerk, dessen Beachtung die Erfüllung bestimmter Forderungen unterstützt. Qualitätsmanagementsysteme (wie z. B. ISO 9000) sind Instrumente der sogenannten „Beschaffenheitsgestaltung“. Mithilfe derartiger Systeme soll die Differenz zwischen (An-)Forderung und Umsetzung so gering wie möglich gehalten werden.

Die unterschiedlichen Qualitätsauffassungen münden in unterschiedlichen Qualitätssicherungsstrategien und Qualitätsbewertungen. Viele Wissenschaftler interpretieren Qualität als „Güte“, die Qualität hat ihren Zweck in sich selbst. Diese Anschauung geht häufig mit dem Prinzip der Input-Kontrolle einher. Qualität wird u. a. dadurch gesichert, dass die Entscheidungsbefugten festlegen, welche Personen und Ressourcen in das System eintreten dürfen. Für den Bereich der psychologischen Diagnostik gipfelt dies beispielsweise in der Forderung, dass nach Art der Zünfte nur im Sinne der höchstmöglichen Qualität ausgebildete und kooptierte Personen, z. B. Diplom-Psychologen, psychologische Diagnostik betreiben dürfen und dass psychometrische Tests nur an Diplom-Psychologen verkauft werden dürfen. (Die Frage, ob und in welchem Ausmaß zusätzlich zu den Psychologen Angehörige anderer Berufsgruppen an psychologischer Diagnostik im Allgemeinen und Testanwendungen im Besonderen beteiligt werden sollten, lässt sich bis in die Zeiten der Psychotechnik zurückverfolgen (Herrmann, 1966, S. 263, verweist in diesem Kontext auf eine Publikation von Giese aus dem Jahre 1927) und hat bis heute nicht an Aktualität verloren (vgl. z. B. Hall, Howerton & Bolin, 2005; Kersting, 2008a; Schuler, 1996, 2003; Turner, DeMers, Fox & Reed, 2001).) Auch für diagnostische Produkte findet sich ein derartiges Qualitätsverständnis. Ein qualitativ hochwertiger psychometrischer Test wäre dieser Auffassung entsprechend nur ein Test mit der höchstmöglichen Reliabilität und Validität.

Aus der Sicht vieler Praktiker hat demgegenüber eine Qualitätsanforderung selbst eine Qualität, sie ist angemessen, überzogen oder zu gering – und zwar

im Bezug auf den jeweiligen Anwendungszweck. Wissenschaftler müssen sich fragen lassen, inwieweit aus theoretischen Kenntnissen über das Machbare berechnete und notwendige Qualitätsforderungen für die Praxis abgeleitet werden können. Damit rückt im Sinne von Cronbach und Gleser (1965) die Anzahl der korrekten Entscheidungen, die aufgrund einer diagnostischen Prozedur gefällt werden, in den Mittelpunkt des Interesses (vgl. bereits Franke, 1969, sowie den Beitrag von Franke und Hornke in diesem Band). Die Frage lautet beispielsweise, wie reliabel ein diagnostisches Verfahren, z. B. ein psychometrischer Test, überhaupt sein muss, um unter den gegebenen Rahmenbedingungen einer konkreten Fragestellung sowie der gegebenen Basis- und Selektionsquoten eine Entscheidung treffen zu können, die eine festgesetzten Fehlerwahrscheinlichkeit nicht überschreitet. Ein Test hat die notwendige Qualität, wenn er diese Forderung an die Reliabilität erfüllt, die absolut betrachtet nicht notwendigerweise „hoch“ ausgeprägt sein muss. Außerdem können in der Praxis zusätzlich zu den klassischen Gütekriterien wie Reliabilität und Validität für die Entscheidung für oder gegen ein bestimmtes diagnostisches Vorgehen auch andere Aspekte, wie z. B. die Akzeptanz, Schnelligkeit und die Praktikabilität Berücksichtigung finden. Es gilt, die erbrachten psychodiagnostischen Leistungen im Verhältnis zu den Möglichkeiten zu bewerten, also eine relationale Bewertung von Leistungen an die Stelle der absoluten Leistungsbewertung zu setzen.

2 Überblick über ausgewählte Qualitätsstandards

Es gibt sehr viele Qualitätsstandards für die Diagnostik, dies vereitelt den Versuch einer vollständigen Aufzählung und Dokumentation. Der vorliegende Artikel konzentriert sich auf Qualitätsstandards, die von deutschen, US-amerikanischen, europäischen und/oder weltweiten Organisationen erarbeitet wurden. Mit Ausnahme der USA und Deutschland werden nationale Qualitätsstandards (z. B. Australien: Davidson, 1997; van de Vijver, 1997, und Kanada: Simner, 1996) nicht thematisiert, es sei denn, sie sind von beispielgebender Bedeutung, wie z. B. das niederländische Testbeurteilungssystem (vgl. Abschnitt 2.3.3.1) oder das britische und niederländische System der Zertifizierung von Testanwendern (vgl. Abschnitt 2.2.2.1). Im Vordergrund stehen Qualitätsstandards für diagnostisches Handeln und/oder für diagnostische Instrumente. Obwohl der Ethik im Rahmen der Diagnostik zweifelsohne eine hohe Bedeutung zukommt (vgl. z. B. Leach & Oakland, 2007), da es nicht nur um das Wissen und Können der Diagnostiker, sondern auch um das „Dürfen“ geht, werden Ethik-Richtlinien im engeren Sinne im vorliegenden Beitrag ausgegrenzt. Ethik-Richtlinien wie die von der Deutschen Gesellschaft für Psychologie e. V. (<http://www.dgps.de/dgps/aufgaben/003.php>) gemeinsam mit dem Berufsverband Deutscher Psychologinnen und Psychologen e. V. (BDP; <http://www.bdp-verband.org/bdp/verband/ethik.shtml>) 1998 (geringfügig geändert 2005) herausgegebenen „berufsethischen

Richtlinien“ (zugleich Berufsordnung für Psychologen des BDP), wie die nordamerikanischen „Ethical Principles of Psychologists and Code of Conduct“ der American Psychological Association (APA, 2002, vgl. <http://www.apa.org/ethics/code/index.aspx>) oder wie der Meta-Code of Ethics der European Federation of Professional Psychologists Associations (<http://www.efpa.eu/ethics>, vgl. z. B. Lindsay, Koene, Øvreide & Lang, 2008) betreffen häufig alle Anwendungsgebiete der Psychologie und nicht nur die Diagnostik. Ethik-Standards erfüllen auch berufsständische Funktionen (vgl. Schuler, 1991, S. 343; Schuler, 1996, S. 188 ff.) Ausgeblendet werden auch organisationsspezifische Qualitätsstandards für Diagnostik wie z. B. die der Bundesagentur für Arbeit (Hilke, 2004) und der Bundeswehr (Puzicha, 1997).

Zunächst werden Qualitätsstandards thematisiert, die sich auf den gesamten diagnostischen Prozess beziehen und sich nicht auf spezifische diagnostische Verfahren (z. B. Tests) beschränken. Abschnitt 2.2 widmet sich dem kompetenzorientierten Qualitätsansatz, wobei zwischen einer verfahrensunspezifischen Variante und einer verfahrensspezifischen Variante (insbesondere in Bezug auf Tests) unterschieden wird. In Abschnitt 2.3 werden Qualitätsstandards besprochen, die sich explizit auf Tests und die Testanwendung beziehen, wobei auch Testbeurteilungssysteme und die testspezifische Qualifizierung von Anwendern thematisiert werden. Qualitätsstandards zu weiteren Verfahren (z. B. Assessment-Center und Interviews) finden in Abschnitt 2.4 Berücksichtigung. Diese Einteilung erleichtert zwar die Orientierung, ist aber nicht immer trennscharf. Viele Qualitätsstandards müssen als Mischformen bezeichnet werden: So konzentrieren sich die Standards des „Arbeitskreises Assessment Center“ auf ein bestimmtes Verfahren (z. B. Assessment-Center), dieses Verfahren besteht aber aus einem Methoden-Mix. Außerdem sind die Standards des Arbeitskreises Assessment Center explizit prozessorientiert und berücksichtigen u. a. auch die Kompetenzen der Verantwortlichen. Umgekehrt nehmen zahlreiche Qualitätsstandards, die sich auf allein Tests beziehen, für sich in Anspruch, prinzipiell auf alle diagnostischen Verfahren anwendbar zu sein und legen den Begriff „Test“ explizit sehr weit aus.

2.1 Verfahrensunspezifische Qualitätsstandards

2.1.1 Standards für pädagogisches und psychologisches Testen

Von zentraler Bedeutung sind die in Gemeinschaftsarbeit verschiedener US-amerikanischer Organisationen [American Educational Research Association (AERA), American Psychological Association (APA) & National Council on Measurement in Education (NCME), 1999] herausgegebenen „Standards for educational and psychological testing“ (nachfolgend mit „Standards“ bezeichnet). Wenngleich

überwiegend Leistungstests und Fragebogen behandelt werden, ist der Begriff „Test“ sehr weit gefasst und bezieht prinzipiell auch Interviews und verhaltensorientierte Verfahren wie z. B. das Rollenspiel ein. Aus diesem Grunde und vor allem aufgrund der zentralen Bedeutung der Standards werden sie im vorliegenden Text in diesem Abschnitt besprochen, der eigentlich keine Beschränkung auf spezifische Verfahren (wie Tests) vorsieht. Die Standards waren und sind Vorbild verschiedener nationaler und nationenübergreifender sowie fachspezifischer Richtlinien. Die Standards werden hier überblicksartig sowie anhand von Beispielen beschrieben, in Abschnitt 3 der DIN 33430 gegenübergestellt und im Kontext dieses Vergleichs im vierten Abschnitt noch einmal näher erläutert.

Nachdem zunächst 1950 „ethical standards for the distribution of psychological tests and diagnostic aids“ (APA, 1950) und 1952 ein erstes „preliminary proposal“ veröffentlicht worden waren, erschien 1954 die erste Ausgabe der Standards. Weitere, inhaltlich jeweils zum Teil deutlich geänderte Fassungen erschienen 1966, 1974, 1985 und 1999, an einer weiteren Fassung wird aktuell gearbeitet. Anhand der verschiedenen Ausgaben der Standards kann aufgezeigt werden, wie sich die Vorstellungen über Qualitätsmerkmale der Diagnostik über die Zeit hinweg wandelten. Sireci (1998) arbeitet dies am Beispiel der unterschiedlichen Aussagen der jeweiligen Ausgaben der Standards zur Validität heraus. Von

Tabelle 1:

Inhaltsverzeichnis der Standards for educational and psychological testing (AERA et al., 1999)

Teil I: Testkonstruktion, Evaluation und Dokumentation	<ol style="list-style-type: none"> 1. Validität 2. Reliabilität und Messfehler 3. Testentwicklung und Testrevison 4. Skalierung, Normen und Testwertvergleichbarkeit 5. Testdurchführung, Testbewertung und Ergebnisdarstellung 6. Unterstützende Dokumentation für Tests
Teil II: Testfairness	<ol style="list-style-type: none"> 7. Fairness bei der Testanwendung und bei der Testung 8. Rechte und Pflichten der Testteilnehmer 9. Das Testen von Personen mit unterschiedlichem sprachlichen Hintergrund 10. Das Testen von Personen mit Behinderungen
Teil III: Testanwendungen	<ol style="list-style-type: none"> 11. Pflichten der Testanwender 12. Psychologisches Testen und psychologische Diagnostik 13. Pädagogisches Testen und pädagogische Diagnostik 14. Testen im Kontext der Eignungsdiagnostik und der beruflichen Zulassung und Zertifizierung 15. Testen im Kontext von (Lehr-)Programmevaluationen sowie zur Unterstützung (bildungs)politischer Entscheidungen

der 1985 publizierten Version wurde 1998 eine deutschsprachige Ausgabe veröffentlicht (Häcker, Leutner & Amelang, 1998).

Seit 1999 liegt die fünfte, überarbeitete Fassung vor (AERA et al., 1999; Eignor, 2001). Die 264 Standards dieser aktuellen Auflage gliedern sich in drei Teile und 15 Abschnitte (vgl. Tab. 1). Jeder Teil sowie jeder Abschnitt beginnt mit einem einführenden Erläuterungstext, bevor die eigentlichen Standards präsentiert werden. Zu jedem Standard findet sich dann noch ein kurzer Kommentar. Die folgenden Ausführungen gehen von den eigentlichen Standards aus, nicht aber von den Erläuterungen in den Einführungstexten oder Kommentaren.

Der erste Teil der Standards thematisiert die Entwicklung, Evaluation und Dokumentation von Tests. Im Einzelnen werden 123 Standards zu den Themengebieten (1) Validität, (2) Reliabilität, (3) Testentwicklung und Testrevison, (4) Skalierung und Normierung, (5) Testanwendung, Testauswertung und Ergebnisdarstellung sowie (6) Dokumentation formuliert.

Validierung bezieht sich immer auf die Interpretation eines Tests und nicht auf den Test selbst, dies stellen die Standards in der Erläuterung zum ersten Abschnitt klar. Entsprechend fordern die Standards eine logische/theoretische Begründung und nach Möglichkeit empirische Fundierung für jede empfohlene Interpretation und Verwendung von Testwerten sowie zu jeder Aussage über Testeigenschaften (z. B. zur Resistenz gegenüber Übungseffekten, Standard 1.9) und für alle getroffenen methodischen Annahmen (z. B. Korrektur gegen Streuungseinschränkungen). Sofern Methoden der statistischen Adjustierung, wie zum Beispiel Minderungskorrektur zur Anwendung kommen, sind den Standards entsprechend sowohl die ursprünglich erhaltenen als auch die korrigierten Koeffizienten sowie alle in Zusammenhang mit der Adjustierung verwendeten Statistiken anzuführen (Standard 1.18).

Im Abschnitt 2 über Reliabilität wird großer Wert auf den Bericht der jeweils angewendeten Methoden sowie den Bericht der Standardschätzfehler oder Testinformationsfunktionen gelegt. Es werden explizit Verbindungen zwischen der Reliabilität und der Validität hergestellt, wenn z. B. in Standard 2.3 gefordert wird, Reliabilitätswerte für Differenzwerte (zwischen Personen, zwischen Gruppen) anzuführen, sofern bei der Testinterpretation auf eben diese Differenzen Wert gelegt wird. Vergleichbar fordert Standard 2.7, dass das Vorgehen der Reliabilitätsschätzung einem etwaigen multifaktoriellen Charakter des Instruments entsprechen muss. Darüber hinaus werden zielgruppenspezifische Reliabilitätsschätzungen eingefordert (z. B. Standard 2.11), wozu auch normgruppenspezifische (Standard 2.12), regionalspezifische (Standard 2.13) und stichprobenspezifische (Standard 2.20) Reliabilitätsschätzungen zählen.

Sowohl hinsichtlich der Validität (Standard 1.5) als auch hinsichtlich der Reliabilität (Standard 2.4) und der Normierung (Standard 4.6) fordern die Standards eine detaillierte Beschreibung der Untersuchungsgruppen, einschließlich der jeweiligen Stichprobenpläne sowie einen umfassenden Bericht von Teststatistiken.

Im dritten Absatz werden umfassende Reflexionen und Dokumentationen zu den Zielen und Vorgehensweisen der Testentwicklung und Testrevision gefordert. Die Standards behandeln hierbei auch Detailfragen wie beispielsweise Fragen der Testinstruktion und -durchführung (z. B. Standards 3.19 bis 3.21) und Fragen der Auswahl der Itemformate sowie der Items (z. B. bei Testverkürzungen) oder Itemsets (z. B. Standards 3.7, 3.9, 3.10, 3.12, 3.16).

Der vierte Abschnitt greift die Aspekte Skalierung und Normierung ausführlich auf, wobei z. B. auch die Formulierung von „cut scores“ sowie die Äquivalenz von Testformen (inklusive der Equating-Technik) thematisiert werden. Bezüglich der Normen fordern die Standards beispielsweise über den Zeitraum des Testgebrauchs hinweg ausreichend häufige Überarbeitungen, um eine anhaltend genaue und angemessene Interpretation zu ermöglichen (Standard 4.18).

Testanwendung, Testauswertung und Ergebnisdarstellung sind Gegenstand des fünften Abschnitts, wobei es im Wesentlichen um die Objektivität und den Testschutz geht (z. B. Standard 5.7) sowie um das Spannungsfeld zwischen Standardisierung einerseits und der Berücksichtigung der Bedürfnisse spezifischer Individuen/Gruppen andererseits. Ein weiterer Aspekt ist der Datenschutz (z. B. Standards 5.13, 5.15 und 5.16).

Der sechste und letzte Abschnitt im ersten Teil der Standards behandelt das Informationsmanagement zu einem Test. Hierzu zählen Testhandbücher und Instruktionstexte für Testadministratoren und -auswerter ebenso wie Informationsmaterialien für die getesteten Personen. Neben den Informationen über Ziele und Nutzen eines Tests, fordern die Standards beispielsweise auch Informationen über nicht gerechtfertigte Testverwendungen oder absehbare missbräuchliche Anwendungen (z. B. Standards 6.4, 6.15).

Der zweite Teil der Standards widmet sich mit insgesamt 48 Standards der Frage der Testfairness. Allgemeine Aspekte der Testfairness und des Testbias werden in Abschnitt 7, die Rechte und Pflichten der Testteilnehmer werden in Abschnitt 8 behandelt. Die Abschnitte 9 und 10 setzen sich mit den Besonderheiten der Testung von Personen mit unterschiedlichem sprachlichen Hintergrund bzw. mit der Testung von behinderten Menschen auseinander.

In Abschnitt 7 werden u. a. gesonderte Validierungshinweise für relevante Untergruppen gefordert, falls aufgrund von Forschungen unterschiedliche Bedeu-

tungen der Testwerte oder unterschiedliche Itemfunktionen zwischen den Teilnehmergruppen erwartet werden können. Auch auf die Empfindungen der Testteilnehmer sollte Rücksicht genommen werden, so empfiehlt Standard 7.4 Begriffe, Symbole, Worte, Redewendungen und Inhalte, die im Allgemeinen von Angehörigen einer ethnischen Gruppe, eines Geschlechts oder einer anderen Gruppe als beleidigend empfunden werden, zu eliminieren, sofern diese nicht für den Test unverzichtbar sind. Die Standards betrachten auch die möglichen Auswirkungen der Publikation von Testergebnissen, z. B. auf die Zukunftschancen und Ausbildungsmöglichkeiten der Testteilnehmer aus spezifischen Gruppen, und mahnen zu einer differenzierten und zurückhaltenden Darstellung (z. B. Standard 7.10).

Der achte Abschnitt widmet sich dem Testprozess aus der individuellen Perspektive der getesteten Person. Neben Rechten (aufgrund berufsständischer, ethischer Verpflichtungen), insbesondere dem Recht auf umfassende Informationen, werden den Testteilnehmern auch Pflichten zugeschrieben, etwa die Pflicht, den Testinstruktionen Folge zu leisten. Diese Pflichten kommen allerdings nur in den vorangestellten Erläuterungen, nicht aber in den Standards selbst zum Ausdruck. Zu den Rechten des Teilnehmers zählt beispielsweise auch das Recht, eine Kopie der Testergebnisse bzw. -interpretationen zu erhalten, falls die Testergebnisse dazu genutzt werden, für den Testteilnehmer relevante Entscheidungen zu fällen oder vorzubereiten. Ausnahmen gelten dann, wenn der Teilnehmer explizit auf dieses Recht verzichtet hat oder eine derartige Weitergabe gesetzlich verboten ist (Standard 8.9). Die Standards beschäftigen sich auch mit dem Fall, dass der Testanwender den Testteilnehmer verdächtigt, bei der Testung ein Fehlverhalten gezeigt zu haben (z. B. Täuschung). Diesbezüglich sehen die Standards vor, den Teilnehmer rechtzeitig über den gegen ihn bestehenden Verdacht zu informieren und ihm Gelegenheit zur Stellungnahme einzuräumen (Standards 8.10 bis 8.12).

In Abschnitt 9 geht es um den Zusammenhang zwischen der Sprache der getesteten Person und der Sprache, in der Test und Testung durchgeführt werden. Ziel der Standards ist es, die Testung so zu gestalten, dass Reliabilität und Validität der aus dem Test abgeleiteten Schlussfolgerungen möglichst unabhängig von potenziellen sprachlichen Unterschieden der getesteten Person sind (Standard 9.1). Besteht ein begründeter Verdacht, dass diese Unabhängigkeit nicht gegeben ist, fordern die Standards beispielsweise, soweit möglich, für jede untersuchte sprachliche Untergruppe dieselbe Form des Validitätsnachweises, wie für die Gesamtgruppe (Standard 9.2). Die Standards beschäftigen sich auch mit der Frage, wie vorzugehen ist, wenn ein Test in andere Sprachen übersetzt wird (Standard 9.7), wenn bei einem Testverfahren ein Dolmetscher eingesetzt wird (Standard 9.11) oder wie ein Test auszuwählen ist, wenn ein Test in mehreren Sprachen zur Verfügung steht und der Teilnehmer mehrere dieser Sprachen

beherrscht (Standard 9.3). Für den Fall, dass ein Test in mehreren verschiedenen sprachigen Versionen zur Verfügung steht, die als vergleichbar behandelt werden, fordern die Standards den empirischen Nachweis dieser Vergleichbarkeit (Standard 9.9):

Die Untersuchung von Menschen mit Behinderungen sowie insbesondere die in diesem Kontext ggf. notwendigen Modifikationen (wie z. B. veränderte Zeitvorgaben) der standardisierten Vorgehensweisen sind Gegenstand des zehnten Abschnitts der Standards. Durch die Einhaltung der Standards des zehnten Abschnitts soll sichergestellt werden, dass die aus der Testung abgeleiteten Interpretationen das intendierte vorgesehene Konstrukt widerspiegeln und nicht etwa Behinderungen und damit verbundene Eigenschaften, die außerhalb der beabsichtigten Messung liegen (Standard 10.1). Wenn immer möglich, sollen empirische Studien zur Validität der Schlussfolgerungen, die aufgrund von Ergebnissen in den modifizierten Testversionen getroffen wurden, mit Teilnehmern mit unterschiedlichen Behinderungen durchgeführt werden (Standard 10.7). Grundsätzlich dürfen Diagnosen oder Interventionsentscheidungen über behinderte Menschen den Standards zufolge nicht allein aufgrund eines Tests getroffen werden, sondern es sind mehrere verschiedene Informationsquellen zu verwenden (Standard 10.12). In mehreren Standards des Abschnitts 10 wird gefordert, dass Personen, die für die Testung von Menschen mit Behinderungen verantwortlich sind, zusätzlich zu den Kenntnissen über Tests auch über Kenntnisse der jeweiligen Behinderungen und ihre Auswirkungen verfügen (z. B. Standards 10.2 und 10.10).

Im abschließenden dritten Teil werden in fünf Abschnitten mit weiteren 93 Standards Forderungen zur Testanwendung formuliert, wobei hier auf spezielle Anwendungsfelder wie z. B. Tests für pädagogische Zwecke oder für die Eignungsdiagnostik eingegangen wird.

Abschnitt 11 thematisiert die Verantwortlichkeiten der Testanwender, denen unter bestimmten Umständen Nachweise der Reliabilität und Validität abverlangt werden, beispielsweise dann, wenn der Test für einen Zweck eingesetzt werden soll, für den wenig oder keine Dokumentation verfügbar ist (Standard 11.2) oder wenn ein Test bei Personen eingesetzt wird, von denen anzunehmen ist, dass sie nur im begrenzten Umfang mit der Testsprache vertraut sind (Standard 11.22). Testanwender sollen über die Qualifikation und Erfahrung für die Testdurchführung und -interpretation verfügen, die im Testmanual gefordert wird (Standard 11.3) und grundsätzlich mit den Möglichkeiten und Grenzen testbasierter Aussagen vertraut sein (z. B. Standard 11.10 und 11.15). Die Standards verpflichten den Testanwender, berechnete Informationsinteressen zu befriedigen (Standard 11.5), Testergebnisse innerhalb eines angemessenen Zeitraums auf verständliche Weise mitzuteilen (Standard 11.6), die Urheberrechte (Standard 11.8) und die

Vertraulichkeit (Standard 11.7) von Tests sowie den Schutz der Testdaten (Standard 11.14) zu wahren. Die Informationspflicht geht so weit, dass z. B. die Testentwickler Teilnehmer über die Anfechtbarkeit der Korrektheit der von einem Testanwender ermittelten Testergebnisse informieren müssen (Standard 11.1).

Abschnitt 12 behandelt u. a. die Kombination von Tests (Standards 12.4 und 12.5) und erneut die notwendige Qualifikation und die Erfahrung der für die Testdurchführung, -auswertung und/oder -interpretation verantwortlichen Personen (z. B. Standards 12.8, 12.9, 12.13) sowie den Schutz der Testmaterialien (Standard 12.11). Ein multimethodales Vorgehen (Standard 12.8) wird zum Standard erklärt, darüber hinaus wird empfohlen, auch alternative Hypothesen zur Erklärung der Testergebnisse in Betracht zu ziehen (Standard 12.19) und die Interpretation der Testwerte z. B. durch Beobachtungen bei Vorgesprächen oder der Testung zu stützen (Standard 12.15).

Standards zu von Schulen, Bezirken, Ländern oder Organisationen im pädagogischen Umfeld durchgeführten Tests oder Testprogrammen werden im 13. Abschnitt formuliert, wobei beispielsweise angemahnt wird, die Folgen solcher Programme im Allgemeinen zu untersuchen und potenzielle negative Folgen solcher Programme im Besonderen zu erkennen und zu minimieren (Standard 13.1). Sofern der Test für die Schüler oder Studenten schwerwiegende Entscheidungen zeitigt (z. B. Versetzung), muss die Gelegenheit zu einer Testwiederholung (z. B. mit einer gleichwertigen Form des Tests) bestehen. Dabei sollte den getesteten Personen in der Regel die Gelegenheit eingeräumt werden, vor der erneuten Testung die notwendigen Lernerfahrungen zu sammeln (S. 111).

Um die Eignungsdiagnostik geht es im 14. Abschnitt der Standards. So wird beispielsweise für Tests gefordert, die zum Zwecke der Auswahl, Klassifizierung und Beförderung genutzt werden und für deren Einsatz das Argument der Inhaltsvalidität angeführt wird, auf eine Arbeits- und Anforderungsanalyse zurückzugreifen (Standard 14.8). Bei der Kriteriumsvalidität findet sich beispielsweise ein Hinweis, unter welchen Bedingungen die Ergebnisse von Studien zur Kriteriumsvalidität auf die aktuelle Situation generalisiert werden können (Standard 14.6). Bezüglich der Konstruktvalidität unterscheiden die Standards zwischen der Beziehung zwischen dem Konstrukt, das dem Prädiktor zugrunde liegt, sowie dem Konstrukt, das dem Kriterium zugrunde liegt, einerseits, und der Beziehung zwischen dem Prädiktorkonstrukt und seiner Operationalisierung durch den Test andererseits (Standard 14.12). Zur Eignungsdiagnostik zählen auch Tests zur Lizenzierung im Kontext beruflicher Qualifikationen, die in verschiedenen Standards thematisiert werden.

Die Verwendung von Tests für (Lehr-)Programmevaluationen sowie zur Unterstützung (bildungs-)politischer Entscheidungen aufgrund von Testergebnissen

ist Gegenstand des letzten, 15. Abschnitts. Auch in diesem Kontext wird beispielsweise eine umsichtige, aussagekräftige und um Kontextinformationen anreicherte Darstellung der Testergebnisse angemahnt, die die Wahrscheinlichkeit von Fehlinterpretationen minimiert (Standards 15.11 und 15.12).

Insgesamt stellen die Standards eine über Jahrzehnte gewachsene, einzigartige, prozessorientierte Sammlung von Qualitätsaspekten der entscheidungsorientierten Diagnostik dar. Sie repräsentieren den Konsens einer sehr großen Gruppe hervorragender Fachexperten. Die – entgegen der Behauptung eines sehr weit gefassten Testbegriffes – spürbare Konzentration auf psychometrische Tests ist bedauerlich, da zahlreiche Standards von der Intention her verfahrensübergreifenden Charakter haben und auch z. B. auf Interviews oder Assessment-Center angewendet werden können. Die sprachliche Ausrichtung auf psychometrische Tests erschwert diese Analogie aber. Die Konzeption der Standards, derzufolge die Diagnostik in den 15 Abschnitten unter verschiedenen Perspektiven betrachtet wird (z. B. Eignungsdiagnostik), bringt zahlreiche Redundanzen mit sich, bis hin zu sich wiederholenden, identischen Standards in unterschiedlichen Kapiteln (Standards 1.22 und 15.08). Die Forderungen nach Transparenz, nach Dokumentation und nach einer zurückhaltenden und umsichtigen Darstellung der Testergebnisse kommen in zahlreichen Standards wiederholt vor. Der Vorteil dieser Konzeption liegt in der Möglichkeit einer selektiven Lektüre einzelner Abschnitte. Unklar ist, welchen Stellenwert die Erläuterungen und Kommentare haben, häufig werden hier Qualitätsforderungen formuliert, die in den Standards selbst nicht vorkommen. Als Beispiel wurde bereits ausgeführt, dass in der Erläuterung zum achten Abschnitt, nicht aber in den Standards des achten Abschnitts, auch die Testteilnehmer in die Pflicht genommen werden, aktiv und bewusst zur Qualität beizutragen.

2.1.2 DIN 33430

Im Jahre 2002 wurde die DIN 33430 (DIN, 2002) verabschiedet, die „Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen“ formuliert. Sie bezieht sich auf (a) die Planung von berufsbezogenen Eignungsbeurteilungen, (b) die Auswahl, Zusammenstellung, Durchführung und Auswertung von diagnostischen Verfahren, (c) die Interpretation der Verfahrensergebnisse und die Urteilsbildung sowie (d) die Anforderungen an die Qualifikation der an der Eignungsbeurteilung beteiligten Personen. Der Text umfasst somit zum einen Qualitätskriterien für berufsbezogene Eignungsbeurteilungen und zum anderen Qualifikationsforderungen an die an der Eignungsbeurteilung beteiligten Personen. Die Norm betrachtet den gesamten Prozess der Eignungsbeurteilung als eine Einheit, sie ist eine Prozess- und keine Produktnorm. Die normativen Anforderungen der DIN 33430 sind im Detail im

Abschnitt 3 wiedergegeben (und den in Abschnitt 2.1.1 beschriebenen Standards gegenübergestellt), der hier vorliegende Abschnitt informiert über wenige ausgewählte Aspekte.

Die DIN 33430 gliedert sich in sieben Abschnitte:

1. Anwendungsbereich
2. Normative Verweisungen
3. Begriffe
4. Qualitätskriterien und -standards für Verfahren zur berufsbezogenen Eignungsbeurteilung (Auswahl, Zusammenstellung, Durchführung und Auswertung)
5. Verantwortlichkeiten
6. Qualitätsanforderungen an den Auftragnehmer und die Mitwirkenden
7. Leitsätze für die Vorgehensweise bei berufsbezogenen Eignungsbeurteilungen

Der Text umfasst außerdem einen normativen Anhang A (Anforderungen an Verfahrenshinweise, wie Beschreibungen und Manual) und einen informativen Anhang B (Glossar).

Während die ersten drei Abschnitte einleitenden, informativen Charakter haben, beginnt der normative Teil der DIN 33430 mit Abschnitt 4, der Qualitätskriterien und -standards für Verfahren zur Eignungsbeurteilung benennt und Rahmenbedingungen zur Auswahl und Zusammenstellung der diagnostischen Verfahren sowie zu ihrer Durchführung und Auswertung formuliert. Den Grundsätzen zufolge dürfen beispielsweise nur solche Verfahren für die Eignungsbeurteilung eingeplant werden, die nachweislich einen Bezug zu den Anforderungen haben. Für jedes Verfahren der Eignungsbeurteilung ist darzulegen, wie es zu handhaben ist etc. Hinsichtlich der Planung der Untersuchungssituation ist u. a. sicherzustellen, dass alle Aspekte der Durchführung und Auswertung der Verfahren zur Eignungsbeurteilung sowie die Ergebnisvermittlung vorab in Form von Regeln festgelegt werden. Insgesamt fordert die DIN 33430 umfassende Dokumentationen, die Grundlage jedweder kontrollierten Praxis sind.

Die Verantwortlichkeiten für die Eignungsdiagnostik werden in der DIN 33430 durch Abschnitt 5 geregelt. Diesbezüglich führt die DIN 33430 zwei Begriffe für Beteiligte ein: den „Auftragnehmer“ (eine organisationsinterne oder -externe Person, die für die gesamte, infragestehende Eignungsbeurteilung der Organisation verantwortlich ist) und den „Mitwirkenden“ an der praktisch durchgeführten Eignungsbeurteilung. Jeder Beteiligte muss für seinen Bereich qualifiziert sein.

Im Abschnitt 6 der DIN 33430 werden Qualitätsanforderungen an diejenigen formuliert, die an der Eignungsbeurteilung beteiligt sind. Allerdings sieht die

DIN 33430 keine Regelungen vor, wie die notwendigen Qualifikationen erworben und nachgewiesen werden können. Als Ergänzung zur DIN 33430 hat die Föderation Deutscher Psychologinnen im März 2004 eine „Fortbildungs- und Prüfungsordnung zur Personenlizenzierung für berufsbezogene Eignungsbeurteilungen nach DIN 33430“ erlassen, das entsprechende Fortbildungsprogramm wird im vorliegenden Text im Abschnitt 2.2 beschrieben.

Der Abschnitt 7 der DIN 33430, in dem Leitsätze für die Vorgehensweisen bei berufsbezogenen Eignungsbeurteilungen formuliert werden, hat lediglich informativen Charakter. Hier werden im Sinne eines „Best-practice-Ansatzes“ Hinweise zur Gestaltung der Praxis gegeben.

Im normativen Anhang A der DIN 33430 werden vor allem Anforderungen an Verfahrenshinweise genannt. Der informative Anhang B erläutert in Form eines Glossars die in der Norm genutzten Begriffe.

Die normativen Aussagen der DIN 33430 werden im Abschnitt 3 wiedergegeben und den Standards gegenübergestellt. Grundlage der Darstellung ist hier allerdings eine modifizierte Version des Textes der DIN 33430, nämlich die gesondert erarbeitete „DIN-Screen“-Checkliste. Kersting (2008a) hat den normativen Text (mit Ausnahme der normativen Verweisungen) der DIN 33430 sowie einige wenige nicht normative Aussagen in 318 Einzelaussagen zerlegt, in Aussageform gebracht und in sechs Kategorien (Einzelchecklisten) neu gegliedert. Diese Gliederung ist prozessorientiert und somit nach Ansicht von Kersting (2008a) leichter nachvollziehbar als die Originalgliederung der DIN 33430. Anhand der Checklisten, die auch in elektronischer Form vorliegen, kann man in Form von Selbst- oder Fremdbeurteilungen rasch kontrollieren, ob die Anforderungen der DIN 33430 für einen konkreten Fall erfüllt sind. Bei einer solchen Prüfung wird rasch sichtbar, was u. U. sinnvollerweise noch getan werden kann. Tabelle 2 zeigt beispielhaft einen Ausschnitt aus der Checkliste. In 69 Fällen sind den zu bewertenden Aussagen „Verzweigungsfragen“ vorangestellt um sicherzustellen, dass nur die für den jeweiligen Anwendungszweck relevanten DIN-Anforderungen geprüft werden. In diesem Fall werden einzelne, nachfolgende Aussagen (bei der elektronischen Version „automatisch“) übersprungen. Durch die Verzweigungen und die Hinweise auf Verknüpfungen wird die „DIN Screen“ zu einer „intelligenten“ Checkliste. Zusätzlich zu den DIN-Aussagen finden sich in der „DIN Screen“ in den Fußnoten weitergehende Erläuterungen und/oder Beispiele zu einigen DIN-Regelungen. Die Erläuterungen sind teilweise der DIN 33430 entlehnt, teilweise aber auch als Interpretation der DIN 33430 ergänzt.

Der DIN-Ansatz unterscheidet sich von den übrigen Qualitätsstandards vor allem dadurch, dass die DIN 33430 unter dem Dach eines etablierten Normungsinsti-

Tabelle 2:
Auszug aus der Checkliste zur DIN 33430 (Kersting, 2008a)

	Beispielaussagen aus Checkliste 4: Durchführung, Auswertung und Interpretation der Verfahrensergebnisse sowie Urteilsbildung Normative Aussagen	ja	nein	nicht zu bewerten	Anmerkungen	Quelle (Seite)
208	Es werden nur Informationen zu anforderungsrelevanten Eignungsmerkmalen ausgewertet. (DIN, S. 8)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
209	Die Bewertung richtet sich nach den Beurteilungskategorien <u>und</u> den diesen Beurteilungskategorien vorab zugeordneten Beispielaussagen und/oder Beispielverhaltensweisen. (DIN, S. 8).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
V54	<i>Verzweigungsfrage: Beurteilen mehrere Personen den Ausprägungsgrad von Personenmerkmalen aufgrund von mündlich gewonnenen Informationen bzw. Verhaltensbeobachtungen (z.B. Teamfähigkeit aufgrund eines Interviews oder einer Assessment-Center-Übung)? Falls „nein“ → Bitte weiter bei Aussage 213.</i>	<input type="radio"/>	<input type="checkbox"/>			
210	Der Grad der Übereinstimmung zwischen den beurteilenden Personen wurde berechnet. (DIN, S. 7)	<input type="radio"/>	<input type="checkbox"/>	<input type="radio"/>	<input type="radio"/>	

V: Verzweigungsfrage

tutes außerhalb der psychologischen Berufsverbände formuliert wurde. Die DIN 33430 ist arbeitsfeldbezogen und nicht berufsgruppenorientiert. Dies sollte den Wirkungsgrad des Qualitätsstandards erhöhen. Während sich Nichtpsychologen von Berufsstandesregeln für Psychologen nicht angesprochen fühlen, setzt die DIN 33430 allen Verantwortlichen – ungeachtet ihrer Berufsgruppenzugehörigkeit – Qualitätsstandards. Einer DIN wohnt bei der Formulierung von Qualitätsstandards eine autoritative Wirkung inne, die eine berufsständisch formulierte Richtlinie nicht erlangen kann. Hinzu kommt eine gewisse rechtliche Bedeutung der DIN 33430, sie ist zwar keine Rechtsnorm, indem sie über den aktuellen Stand von Wissenschaft und Praxis informiert, kann sie aber gleichwohl Bedeutung bei einer richterlichen Entscheidungsfindung erlangen. Normen finden bei Gericht Beachtung, um den herrschenden Standard in einem Gebiet zu ermitteln. Die DIN 33430 kann bei Rechtsstreitigkeiten im Kontext der Eignungsdiagnostik (z. B. vor dem Hintergrund des Allgemeinen Gleichbehandlungsgesetzes) als eine Art Beweismittel herangezogen werden (zur rechtlichen Bedeutung der DIN 33430 vgl. u. a. Abeln & Reimann, 2004, Kersting, 2008a, Kersting & Püttner, 2006, sowie Wegener, 2003).

Nicht nur in juristischer, sondern auch in marktwirtschaftlicher Hinsicht kann die Wirkung einer DIN weit über die Wirkung der bisherigen berufsgruppen-intern formulierten Regelungen hinausgehen: Zwar ist die Anwendung einer Norm freiwillig, sie übt aber Zwang aus, wenn sie von relevanten Marktteilnehmern als Standard angesehen und akzeptiert wird. Dadurch kommt dem Nachweis der Konformität mit der Norm eine hohe Bedeutung zu. Ein Novum des DIN-Ansatzes ist die Möglichkeit von Ist-Soll-Vergleichen: Während mit anderen Qualitätsstandards oft lediglich Soll-Werte formuliert wurden, ist die DIN 33430 Grundlage von Personenlizenzierungen sowie von Organisations-zertifizierungen. Die Personenlizenzierung wird in Abschnitt 2.2 erörtert. Bei der Organisationszertifizierung geht es um die Möglichkeit, im Rahmen der ISO-9000-Reihe durch ein aufgrund einer externen Auditierung erteiltes Zertifikat die Übereinstimmung des in der Organisation geübten Qualitätsmanagementsystems mit der Norm zu bestätigen. Eine weniger aufwendige, aber auch weniger überzeugende Variante besteht in einer Selbsterklärung, nach den Regeln der DIN 33430 Dienste anzubieten bzw. zu arbeiten. Eine nicht wahrheitsgemäße Selbsterklärung kann allerdings im Rahmen des Wettbewerbsrechts angegangen werden.

Durch verschiedene Initiativen wurde die Idee der DIN 33430 internationalisiert. Zu nennen ist die Publikation der englischen Fassung der DIN 33430 (Hornke, 2005) sowie die Tatsache, dass Österreich mit der ÖNorm D 4000 (Österreichisches Normungsinstitut, 2005) explizit die DIN 33430 in Anspruch nimmt. Besonders hervorzuheben ist die in 2006 begonnene Entwicklung einer ISO-Norm „Assessment Service Delivery: Procedures and methods to assess people in work and organizational settings“, bei der die DIN 33430 zu Beratungszwecken herangezogen wird.

Insgesamt hat die DIN 33430 eine Vielzahl von Publikationen, Begleit- und Folgeinitiativen nach sich gezogen, eine Übersicht über diese Initiativen gibt Kersting (2008a). Eine Sammlung von Literaturhinweisen zur DIN 33430 kann unter <http://www.kersting-internet.de/DIN-Büch> abgerufen werden.

2.1.3 SIOP-Principles for the validation and use of personnel selection procedures

Die im Jahre 2004 publizierten „Principles for the validation and use of personnel selection procedures“ der Society for Industrial and Organizational Psychology (SIOP, vgl. http://www.siop.org/_Principles/principles.pdf) beziehen sich explizit auf die Standards, konzentrieren sich aber ausschließlich auf die Erfassung von Eignung und Leistung im beruflichen Kontext und auf die in diesem Kontext relevanten Themen wie z. B. Anforderungsorientierung oder Selektions-

und Entscheidungsstrategien. Sie beschäftigen sich damit noch intensiver mit den speziell für die Arbeits- und Organisationspsychologie geltenden Normen, Qualitätsstandards und Richtlinien, als dies seitens der APA mit den „Specialty guidelines for the delivery of services by industrial-organizational Psychologists“ (APA, 1981) geleistet wurde.

Die von der Sektion Arbeits-, Betriebs- und Organisationspsychologie im BDP herausgegebenen „Grundsätze für die Anwendung psychologischer Eignungsuntersuchungen in Wirtschaft und Verwaltung“ (Sektion Arbeits-, Betriebs- und Organisationspsychologie im Berufsverband Deutscher Psychologen, 1980) entsprechen in ihrer Konzentration auf die Eignungsdiagnostik den Qualitätsstandards der SIOP, bleiben aber inhaltlich weit hinter den US-amerikanischen Standards zurück.

2.1.4 Uniform Guidelines on Employee Selection Procedures

Im Jahre 1978 wurden die „Uniform Guidelines on Employee Selection Procedures“ (<http://www.uniformguidelines.com>) publiziert, die von der Equal Employment Opportunity Commission, den Departments of Labor sowie of Justice und der Civil Service Commission herausgegeben wurden. Die Einhaltung der Richtlinien soll gewährleisten, dass bei der Eignungsdiagnostik die staatlichen Anforderungen zum Schutz gegen Diskriminierung wegen Rasse, ethnischer und nationaler Herkunft, Religion und Geschlecht erfüllt werden. Die uniform guidelines sind laut Jeanneret und Zedeck (2010) fachlich umstritten und widersprechen in einigen Punkten den Standards (vgl. Abschnitt 2.1.1) sowie den Principles for the validation and use of personnel selection procedures (vgl. Abschnitt 2.1.3)

2.1.5 Richtlinien für den diagnostischen Prozess

Einen nationenübergreifenden Ansatz stellen die „Richtlinien für den diagnostischen Prozess“ dar, die von der European Association of Psychological Assessment (EAPA) entwickelt werden (Fernández-Ballesteros et al., 2001, bzw. deutschsprachig Westhoff, Hornke & Westmeyer, 2003). Diagnostizieren wird hier als ein komplexer Entscheidungsprozess begriffen, der in einem allgemeinen Rahmenmodell abgebildet werden kann, das vier Phasen umfasst. Zunächst formuliert der Diagnostiker – an den Anforderungen eines Auftraggebers orientiert – Fragen (Phase 1: Analysieren des Falls), bevor er dann zur Beantwortung dieser Fragen gezielt Instrumente auswählt, Daten sammelt und die Ergebnisse analysiert, interpretiert, kombiniert und die Diagnostik nachbereitet (Phase 2:

Organisieren und Berichten der Ergebnisse; Phase 3: Planen der Intervention und Phase 4: Evaluation und Nachuntersuchung). Für jede Phase des diagnostischen Prozesses (z. B. Datensammlung, Interpretation) werden (unterschiedlich viele) Teilschritte formuliert, die letztlich zu 96 einzelnen Richtlinien führen.

2.2 Kompetenzorientierter Qualitätsansatz

Während die bislang referierten Qualitätsstandards den gesamten diagnostischen Prozess thematisieren, konzentriert sich der im Folgenden dargestellte kompetenzorientierte Qualitätsansatz auf den Erwerb und die Vermittlung der Kompetenzen sowie teilweise auch auf die Überprüfung/Lizenzierung der Kenntnisse der qualifizierten Personen. Der Beitrag thematisiert die postgraduale Qualifizierung. Die universitäre Qualifizierung, z. B. die im Rahmen des Psychologiestudiums geleistete Ausbildung in psychologischer Diagnostik und der damit in einigen Ländern (z. B. Australien, Kanada) verbundene „Testschutz“, der auf die universitäre Ausbildung der Anwender vertraut, indem bestimmte Verfahren nur von universitär gebildeten Personen bezogen werden sollen, wird nicht behandelt. Die Darstellung untergliedert sich in verfahrensunspezifische und verfahrensspezifische Varianten des kompetenzorientierten Qualitätsansatzes. Es werden also einerseits Ansätze dargestellt, die sich auf die gesamte Diagnostik beziehen und andererseits Ansätze, die sich auf spezifische Verfahren (z. B. Tests, Assessment-Center) beziehen.

2.2.1 Verfahrensspezifische Variante des kompetenzorientierten Qualitätsansatzes: Fortbildung und Personenlizenzierung zur DIN 33430

Die in Abschnitt 2.1.2 beschriebene DIN 33430 zur Eignungsdiagnostik stellt nicht nur Anforderungen an die Verfahren, sondern auch an die Personen, die den Einsatz der Verfahren planen, die Verfahren durchführen und die Ergebnisse auswerten sowie interpretieren. Dabei wird unterschieden zwischen den Anforderungen an den sogenannten Auftragnehmer einerseits und an die sogenannten Mitwirkenden andererseits: (1) Der Auftragnehmer (A) trägt die operative Verantwortung für die gesamte Eignungsbeurteilung, er plant und gestaltet die Eignungsbeurteilung. Mitwirkende sind an der Eignungsbeurteilung beteiligt, z. B. (2) als Mitwirkende an Verhaltensbeobachtungen (MV) und/oder (3) als Mitwirkende an Eignungsinterviews (ME). Die DIN 33430 definiert, welche Kenntnisse und Erfahrungen für jede der drei Qualifikationsstufen erforderlich sind. Allerdings sieht die Norm keine Regelungen vor, wie die notwendigen Qualifikationen erworben und nachgewiesen werden können. Als Ergänzung

zur DIN 33430 hat die Föderation Deutscher Psychologinnenvereinigungen im März 2004 eine „Fortbildungs- und Prüfungsordnung zur Personenlizenzierung für berufsbezogene Eignungsbeurteilungen nach DIN 33430“ veröffentlicht (http://www.bdp-verband.org/bdp/politik/2004/40920_ordnung.pdf) sowie ein Angebot für Fortbildungen und Lizenzprüfungen erarbeitet. Durch eine gebührenpflichtige, schriftliche Prüfung kann eine Personenlizenz für die berufsbezogene Eignungsbeurteilung nach DIN 33430 erworben werden. Dies ist für jede der drei genannten Funktionen (A, ME und MV) möglich. Grundlage der Fortbildungen sowie der im Multiple-Choice-Format dargebotenen Prüfungsfragen ist das vom Testkuratorium herausgegebene Buch „Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33430“ (Westhoff et al., 2010). Die Gültigkeit der Lizenzen ist auf 5 Jahre befristet. Für die dann notwendige Relizensierung ist es ausreichend, eine Selbsterklärung über Praxiserfahrungen, über die Durchführung von oder die Teilnahme an einschlägigen Fort- oder Weiterbildungsveranstaltungen oder über das Verfassen einschlägiger Vorträge oder Publikationen einzureichen, wobei diese Selbsterklärungen stichprobenartig überprüft werden sollen. Die lizenzierten Personen werden auf Wunsch in einem Register geführt (<http://www.din33430portal.de>). Die Zulassung zur Lizenzprüfung ist nicht an bestimmte Ausbildungsabschlüsse gebunden und auch der vorherige Besuch von Fortbildungsveranstaltungen ist nicht vorgeschrieben. Die Teilnahme an Fortbildungen erfolgt allein auf freiwilliger Basis. Eine ausführliche Darstellung der Ableitung der Trainings- und Prüfungsinhalte aus der DIN 33430 findet sich bei Kersting (2008a), eine Darstellung des Bedarfs nach dieser Fortbildung sowie des Inhalts und des didaktischen Konzeptes der Fortbildung sowie den Prüfungsmodalitäten bei Kersting (2010).

Auch wenn sich die Fortbildung explizit nicht nur an Psychologen wendet, sondern an alle Personen, die in der Eignungsdiagnostik tätig sind, so ist es dennoch zu begrüßen, dass die Inhalte der DIN 33430 an einigen Hochschulinstituten explizit im Psychologiestudium thematisiert werden und dass Universitäten den Studenten die Möglichkeit bieten, die Personenlizenzen zur DIN 33430 im Rahmen des Psychologiestudiums zu erwerben (vgl. Dormann, Moosbrugger, Stemmeler & Maier, 2009). Diesbezüglich ist eine Ausdehnung auf andere Studiengänge wünschenswert.

2.2.2 Verfahrensspezifische Variante des kompetenzorientierten Qualitätsansatzes

Der kompetenzorientierte Qualitätsansatz wurde insbesondere in Bezug auf psychodiagnostische Tests realisiert. Ansätze zur Kompetenzvermittlung und -steigerung sowie ggf. Überprüfung im Bereich anderer Verfahren sind demgegenüber deutlich seltener.

2.2.2.1 Testbezogene Qualifizierung

Wie bereits oben erwähnt, ist die testbezogene Qualifizierung in einigen Ländern über den Mechanismus des „Testschutzes“ an eine universitäre Qualifizierung gekoppelt, indem psychodiagnostische Tests nur von universitär vorgebildeten Personen bezogen werden dürfen. Die an den Universitäten somit geleistete indirekte testbezogene Qualifizierung ist nicht Gegenstand dieses Beitrags. In Deutschland ist die Anwendung von Testverfahren entgegen der weit verbreiteten Meinung nicht gesetzlich geregelt. Wenn Tests in Deutschland von einigen Verlagen nur an bestimmte Nutzergruppen verkauft werden, so handelt es sich hierbei um eine freiwillige vertragliche Regelung der privaten Anbieter.

Eine Qualifizierung für die Anwendung von Tests ist gängigerweise sowohl in Form von allgemeinen Richtlinien als auch in Form spezifischer Qualifizierungs- und Zertifizierungsprogramme vorgesehen. Im vorliegenden Abschnitt werden diejenigen Qualitätsstandards und Qualifizierungsprogramme beschrieben, die sich explizit auf Tests beziehen. Das in Abschnitt 2.2.1 dargestellte verfahrensunspezifische diagnostische Qualifizierungsprogramm zur DIN 33430 bezieht die Anwendung von Tests mit ein.

Eine Beschreibung der für die Testanwendung als notwendig erachteten Kompetenzen (in Abhängigkeit vom Anwendungskontext) wurde von der „Task Force on Test User Qualifications“ geleistet (vgl. z. B. Fremer, 1996; Turner et al., 2001). Während hier die notwendigen Kompetenzen der für die Testung verantwortlichen Personen lediglich beschrieben werden, thematisieren Qualifizierungsprogramme auch den Erwerb und die Vermittlung der Kompetenzen sowie die Überprüfung/Lizenzierung der qualifizierten Personen.

In etlichen Ländern (z. B. Großbritannien, Kanada, USA) variieren die Anforderungen an die Kompetenzen von z. B. Testanwendern in Abhängigkeit von einer Klassifikation der Tests (z. B. Level A, B und C, wobei als „Level C“ solche Tests klassifiziert werden, die als Grundlage besonders weitreichender Entscheidungen herangezogen werden). Für die Nutzung dieser Tests werden dann besonders gut ausgebildete Anwender verlangt, während die weniger anspruchsvollen Tests auch von weniger kompetentem Personal administriert und interpretiert werden dürfen (vgl. beispielhaft das weiter unten beschriebene System zum „Certificate of Competence in Occupational Testing“).

Ein Beispiel für postgraduale Qualifizierungsangebote zur Nutzung von Tests geben die Niederlande. Hier können sich akademisch ausgebildete Psychologen mit psychodiagnostischem Tätigkeitsschwerpunkt von der Psychologenvereinigung NIP (Nederlands Instituut van Psychologen, vgl. <http://www.psynip.nl>)

registrieren lassen. Nachzuweisen sind eine universitäre Ausbildung, ein Training in psychologischer Diagnostik sowie mindestens (maximal über 2 Jahre verteilt) 200 Stunden psychodiagnostischer Tätigkeit und drei Fallbearbeitungen unter Supervision eines akkreditierten Psychologen (Evers, 1996). Der registrierte Psychologe erwirbt die Berechtigung, den Titel *Psycholoog NIP* zu führen. Die unter diesem Zusatz arbeitenden Psychologen sollen professionelle Berufsstandards einhalten. Über die Anerkennung (und die damit verbundene Möglichkeit der (gebührenpflichtigen) Registrierung) entscheidet eine mehrköpfige Kommission. Angedacht war, dass die Nutzung von Tests, die Kenntnisse sowie Weiterbildung in psychometrischer Theorie und Forschung erfordern (Intelligenz-, Persönlichkeits-, Leistungstests), registrierten Psychologen vorbehalten ist, während andere (weniger anspruchsvolle) Tests wie z. B. Interessensinventare auch von anderen Gruppen (z. B. Lehrern) genutzt werden können.

In Großbritannien konnte sich ein Zertifizierungssystem für Anwender von Tests aus den Anwendungsbereichen der Arbeits- und Organisationspsychologie einerseits und der Pädagogischen Psychologie andererseits etablieren. Im Gegensatz zu dem Ausbildungsangebot der niederländischen Psychologenvereinigung wird die Zertifizierung hier nicht nur für Psychologen, sondern auch für Angehörige anderer Berufsgruppen angeboten.

Das System für die Tests aus dem Bereich der Arbeits- und Organisationspsychologie differenziert zwischen zwei Qualifizierungsgraden, dem *Level A* und dem *Level B* des „Certificate of Competence in Occupational Testing“. Der kostenpflichtige Zertifikatserwerb setzt den Besuch spezifischer, in der Regel fünf-tägiger Kurse voraus, die mit einer Prüfung enden (vgl. Bartram, 1996, 2001). Trainings und Prüfung werden von Personen durchgeführt, die von der British Psychological Society (BPS) akkreditiert sind. In jüngster Zeit wurde noch ein Zertifikat für „Test Administration“ vorgeschaltet, wobei die Zertifikate des Levels A oder B dieses Zertifikat bereits umfassen.

Level A des Zertifikats verlangt allgemeines Hintergrundwissen bezüglich wichtiger psychometrischer Konzepte wie Validität, Reliabilität oder Standardisierung und die Kompetenzen zum adäquaten Umgang mit Fähigkeits- und Leistungstests (vgl. Tab. 3). Dazu zählen beispielsweise das Wissen, wann ein Test angewendet werden sollte, wie mit Rückfragen der Testteilnehmer während der Testung umzugehen ist, wie man die Testergebnisse angemessen nutzt und Feedback gibt.

Level B konzentriert sich auf die Nutzung von Persönlichkeitstests und setzt das Zertifikat für Level A voraus. Innerhalb der Level-B-Zertifizierung wird noch

Tabelle 3:
Das britische System „Certificate of Competence in Occupational Testing“

Zertifikat	Einheit
Test Administration	<ol style="list-style-type: none"> 1 Grundlagen des psychologischen Testens 2 Planung der Testanwendung 3 Testanwendung 4 Testbewertung und Datenerfassung 5 Gewährleistung des Test- sowie Datenschutzes
Level A	<ol style="list-style-type: none"> 1 Psychologisches Testen: Bedarfsanalysen/Auftragsklärung 2 Grundprinzipien der Skalierung und Standardisierung 3 Die Bedeutung von Reliabilität und Validität 4 Die Entscheidung, ob psychologische Tests im Rahmen eines eignungsdiagnostischen Prozesses eingesetzt werden sollten oder nicht 5 Die Testdurchführung mit einem oder mehreren Kandidaten sowie der Umgang mit Testbewertungsprozeduren 6 Der angemessene Gebrauch von Testergebnissen sowie die Bereitstellung akkurater schriftlicher und mündlicher Feedbacks für den Auftraggeber und den Kandidaten 7 Gewährleistung des Test- sowie Datenschutzes
Level B	<ol style="list-style-type: none"> 1 Grundlagen der Persönlichkeitstheorie 2 Grundlagen der Persönlichkeitsdiagnostik 3 Testanwendung: Testdurchführung 4 Testanwendung: Aspekte der Interpretation 5 Testanwendung: Feedback 6 Testauswahl und -bewertung: Ansätze zur Persönlichkeitsdiagnostik und Testkonstruktion 7 Testauswahl und -bewertung: Aspekte der Validität und Reliabilität 8 Testauswahl und -bewertung: Computergestützte Diagnostik und computer-generierte Ergebnisdarstellungen 9 Testauswahl und -bewertung: In welchen Fällen sollten Instrumente der Persönlichkeits- und Interessensdiagnostik wie eingesetzt werden?

einmal unterschieden zwischen einem *Intermediate*, einem *Intermediate Plus* oder einem *Full Level B Certificate*. Ersteres wird Personen mit ausreichendem Basiswissen und Expertise bezüglich eines spezifischen Persönlichkeitstests erteilt. Die zweite Variante erfordert dieselben Wissensgrundlagen für ein weiteres Verfahren und das „Full Level B Certificate“ schließlich steht für die allgemeine Kenntnis von Persönlichkeitstests. Die dem Level B zugeordneten Standards untergliedern sich in neun Untereinheiten (*Units*), die sich wiederum nach thematischen Aspekten zu drei übergeordneten Qualifikationsclustern zusammenfassen lassen (vgl. Tab. 3): Theoretische Wissensbasis in Bezug auf Persönlichkeitstheorien und -beurteilung (*Units 1 und 2*), praktische Fähigkeiten zur

Testanwendung (*Units 3 bis 5*) und zur Auswahl passender Tests sowie deren Evaluation (*Units 6 bis 9*).

Laut der Website <http://www.bps.org.uk> hatten bis Anfang 2009 knapp 22.000 Personen ein Zertifikat des Levels A erworben, für Level B wurde zu diesem Zeitpunkt eine Anzahl von 7.682 Zertifikaten genannt.

Die Möglichkeit ein „Certificate of Competence in Educational Testing“ zu erwerben ist relativ neu. Der Aufbau folgt dem System für Tests aus dem Bereich der Arbeits- und Organisationspsychologie, wobei hier nur ein Level-A-Zertifikat angeboten wird. 1.257 Personen hatten dieses „Level A Certificate of Competence in Educational Testing“ erworben (Stand: Anfang 2009). Unter der URL <http://www.psychtesting.org> können ausführliche Informationen zu den Lerninhalten sowie zu den Modalitäten des Kompetenznachweises abgerufen werden.

Darüber hinaus gibt es zahlreiche testbezogene Qualifizierungen, die weniger formalisiert sind, da die Vermittlungsmethode in der Regel nicht curricular geregelt ist und häufig keine Überprüfungen/Lizenzierungen stattfinden. Die entsprechenden Trainings werden zumeist von Einzelpersonen angeboten, das Trainingskonzept ist nicht auf einer organisationalen Ebene abgestimmt. Die entsprechende Trainingslandschaft muss als intransparent bezeichnet werden. Derartige Angebote sind gelegentlich optional (vgl. beispielsweise <http://www.hogrefe.de> oder <http://www.schuhfried.at>). Insbesondere bei kommerziell vertriebenen Tests ist eine schulungsbasierte Akkreditierung, Autorisierung oder Lizenzierung für Nichtpsychologen oder für alle Anwender aber auch verpflichtende Voraussetzung für den Testerwerb und/oder die Testinterpretation (dies gilt beispielsweise für die Verfahren DISG-Persönlichkeits-Profil®; Golden Profiler of Personality; Hermann-Dominanz-Instrument, INSIGHTS®, LIFO-Methode, Myers-Briggs-Type-Indikator, Team-Management-Profil-Fragebogen®, alle Verfahren zitiert nach Sarges und Wottawa (2004)). Eine derartige Praxis ist problematisch, wenn diese Schulungen transparente schriftliche Information über das Verfahren ersetzen.

2.2.2.2 Qualifizierung für Interviews und Assessment-Center

Weit verbreitet, aber wenig formalisiert sind Interviewtrainings sowie Beobachtertrainings für Assessoren im Assessment-Center. Für Trainingskonzepte zu Interviews vgl. u. a. Kanning (2007) und Schuler (2002), für Beobachtertrainings vgl. u. a. Höft und Funke (2006). Zur Effektivität vgl. u. a. Conway, Jako und Goodman (1995) für Interviewtrainings und Woehr und Huffcutt (1994) für Beobachtertrainings für Assessoren.

2.3 Testbezogene Qualitätsstandards sowie Beurteilungssysteme für Tests

2.3.1 Testbezogene Qualitätsstandards des Joint Committee on Testing Practices sowie der British Psychological Society

Zahlreiche Qualitätsstandards beziehen sich auf die Testpraxis. So hat das Joint Committee on Testing Practices (JCTP) 1988 den „Code of fair testing practices in education“ entwickelt (vgl. <http://www.apa.org/science/programs/testing/fair-code.aspx>) und später durch die Herausgabe von „case studies“ veranschaulicht (JCTP, 1993). Der „Code“, der inhaltlich mit den Prinzipien der Standards übereinstimmt, thematisiert die Fairness der Testung gegenüber Alter, Geschlecht, Behinderungen, ethnischer, nationaler und sprachlicher Herkunft sowie religiöser und/oder sexueller Orientierungen. Eine Arbeitsgruppe der JCTP hat sich speziell mit der Perspektive der getesteten Personen beschäftigt und die Richtlinie „The Rights and Responsibilities of Test Takers: Guidelines and Expectations“ verfasst (JCTP Working Group, 2000, vgl. <http://www.apa.org/science/programs/testing/rights.aspx>). Ein eigener Abschnitt dieser Richtlinien richtet sich an die für die Testung verantwortlichen Organisationen und Personen und verdeutlicht, was bei der Testung beachtet werden muss, um die Rechte der getesteten Personen zu wahren. Teilweise vergleichbar zu der Information der Arbeitsgruppe des JCTP ist die vom BDP herausgegebene Broschüre „Psychologische Testverfahren bei Einstellungsuntersuchungen, Sieben Tipps für Bewerber“ (BDP, n. d.).

Die British Psychological Society (BPS) hat einen „Code of Good Practice in Psychological Testing“ erarbeitet, den sie u. a. im Rahmen der Trainings zum „Certificate of Competence in Occupational Testing“ (vgl. Abschnitt 2.2.2.1) zugrunde legt.

2.3.2 Testbezogene Qualitätsstandards der International Test Commission

Qualitätsstandards zum Testeinsatz wurden nationenübergreifend von der International Test Commission (ITC, vgl. <http://www.intestcom.org>) erarbeitet. In der ITC sind eine Vielzahl nationaler psychologischer Vereinigungen und mit verschiedenen Aspekten des Testens befasste Kommissionen sowie Herausgeber von Tests vertreten. Die von der ITC erarbeiteten International Guidelines for Test Use (ITC, 2001) wurden bislang in 13 Sprachen übertragen, darunter auch Deutsch (vgl. <http://www.intestcom.org>). Die Richtlinien schreiben eine fachlich kompetente Praxis in der Testanwendung sowie eine Verantwortung für eine ethisch korrekte Testanwendung vor, wobei sie allerdings weniger beschreiben, was kompetente Anwender auszeichnet als vielmehr, was kompetente An-

wender tun sollten. Im Anhang der Richtlinien wird ein Beispiel für einen Entwurf von Grundsätzen in der Testanwendung sowie ein Beispiel für eine Vereinbarung zwischen Testanwendern und den anhand der Tests beurteilten Personen dargestellt. Außerdem werden im Anhang Aspekte gelistet, die bei Testungen mit behinderten oder in anderer Weise beeinträchtigten Personen beachtet werden sollten.

Aktuell arbeitet auch eine „Joint EAWOP-EFPA Working Group“ (<http://www.eawop.org/web/?cat=48>) an der Formulierung von „Test User Standards“.

Durch die Internationalisierung wird es notwendig, vorhandene monolinguale bzw. monokulturelle Testformate an andere Sprach- und Kulturräume zu adaptieren. Dabei sind nach Ansicht des ITC die von ihr erarbeiteten Guidelines for Test Translation and Adaptation zu beachten (vgl. z. B. Hambleton, 2001, 2005). Diesen Richtlinien zufolge ist bei Übersetzungen und Adaptionen von Verfahren auf eine eventuell bestehende „kulturelle Distanz“ zu achten, die sich u. a. in Bereichen wie Sprache, Familienstruktur, Religion, Lebensstil und Wertekanon manifestieren kann. Die Richtlinien empfehlen, bei Übersetzungen und Adaptionen erfahrene Übersetzer einzubinden. Der Anwendung adaptierter Tests in anderen kulturellen Kontexten ist ein eigener Abschnitt gewidmet, der u. a. für den möglichen Einfluss unterschiedlicher Situationsbedingungen sensibilisiert.

Eine weitere Richtlinie der ITC widmet sich der zunehmend an Bedeutung gewinnenden computer- und internetbasierten Testung (vgl. auch Bartram & Hambleton, 2006; Naglieri et al., 2004). Die ITC International Guidelines on Computer-Based and Internet-Delivered Testing (ITC, 2006) thematisieren hier u. a. Anforderungen an den Datenschutz und die Qualität der Tests, die technischen Voraussetzungen und die Frage der Verantwortlichkeiten für diese Art von Testungen. Die einzelnen Aspekte werden aus der Sicht der Testentwickler, Testverlage und Testnutzer betrachtet. In diesem Zusammenhang sei auch auf die „Guidelines for computer-based testing“ verwiesen, die von der Association of Test Publishers (2002) herausgegeben wurden.

2.3.3 Kriterien für die Testbeurteilung sowie Testbeurteilungssysteme

Bezüglich der Beurteilung von Tests muss zwischen mehr oder minder standardisierten Beschreibungssystemen einerseits und Beurteilungssystemen im engeren Sinne andererseits unterschieden werden. Beschreibungssysteme, wie z. B. die Kriterien für die Testbeurteilung der Föderation Deutscher Psychologenv Verbände (1986) zielen zwar auf die Beurteilung von Verfahren, indem sie die für die Beurteilung relevanten Bereiche kennzeichnen und Beurteilungshinweise

geben, sie stellen aber kein Beurteilungssystem im engeren Sinne dar, weil es ihnen an eindeutigen Bewertungsvorschriften mangelt. Die alleinige Vorgabe eines Test-Beschreibungsrasters, wie z. B. von dem ZPID (<http://www.zpid.de/index.php?wahl=products&uwahl=printed&uuwahl=testverzeichnis>) oder der ZUMA (ZUMA-Informationssystem, ZIS, vgl. <http://www.gesis.org/methodenberatung/zis/>) kann nicht als Testbeurteilungssystem gewertet werden.

Wie eingangs erwähnt, sollen im vorliegenden Beitrag nur Ansätze näher erläutert werden, die als Konventionen zwischen verschiedenen Personen unter dem Dach einer oder mehrerer Organisationen ausgehandelt wurden. Entsprechend fallen Überblickspublikationen zu Tests aus dem Raster, auch wenn sie, wie im Falle des „Handbuch psychologischer und pädagogischer Tests“ von Brickenkamp (Erstauflage 1975; Neuauflage von Brähler, Holling, Leutner & Petermann, 2002) oder den mehrbändigen „Tests and Test Critiques“ (erste Ausgabe Keyser & Sweetland, 1984) übergeordnete Bedeutung erlangt haben. Ausgegrenzt aus der Erläuterung werden auch Publikationen, die zwar von Organisationen verantwortet werden, denen aber kein explizites Testinformations- und oder -beurteilungssystem zugrunde liegt, beispielsweise der von der National Career Development Association (NCDA) verantwortete Counselor's Guide to Career Assessment Instruments (z. B. Whitfield, Feller & Wood, 2009).

Der Übergang von einer Richtlinie zur Beurteilung von Verfahren hin zu einem vollständigen Beurteilungssystem ist aber fließend, zumal die Beurteilungssysteme in der Regel nicht deterministisch sind und Freiheitsgrade der Beurteilung vorsehen. So ist beispielsweise das nordamerikanische Buros-System (vgl. z. B. Plake & Impara, 2001; <http://www.unl.edu/buros>) an der Grenze zwischen Richtlinie und Beurteilungssystem anzusiedeln. Das Buros-Institut informiert pro Test u. a. über den Verlag, die verfügbaren Testformen, den Preis und führt, soweit vorhanden, eine Bibliografie über Publikationen zum Test an. Vor allem gibt das Institut ein bestimmtes Schema für die Testrezensionen vor, die regelmäßig im Mental Measurement Yearbook (MMY) publiziert werden. Im Jahre 2010 wurde die 18. Fassung dieses umfassenden Nachschlagewerks publiziert, in der über 200 Tests rezensiert werden (Spies, Carlson & Geisinger, 2010).

Einen standardisierten Aufbau von Testrezensionen sehen auch die sogenannten „Labels“ (Prüfsiegel, Prüfzeichen, Gütesiegel) der Fachgruppe Diagnostik des Schweizerischen Dienstleistungszentrums Berufsbildung/Berufs-, Studien- und Laufbahnberatung vor (vormals Diagnostikkommission des Schweizerischen Verbandes für Berufsberatung) vor. Mittels der Labels werden die wesentlichen Aspekte von diagnostischen Instrumenten beschrieben und beurteilt und auf diese Weise Berufsberatern vorgestellt. Die aktuell über 50 vorhandenen Labels können auf der Homepage der Diagnostikkommission eingesehen werden (vgl. <http://www.diagnostik.sdbb.ch/labels>).

2.3.3.1 Das COTAN-Testbeurteilungssystem

Ein Beurteilungssystem für Tests stellt das niederländische COTAN-System (Evers, 2001a) dar, das vom Committee On Test Affairs Netherlands erarbeitet wurde. Bereits im Jahre 2000 umfasste die „Documentation of Test and Test Research“ (Evers, van Vliet-Mulder & Groot, 2000) laut Evers (Evers et al., 2000) 372 Tests, die einheitlich nach dem COTAN-System analysiert und hinsichtlich vorgegebener Kriterien von jeweils zwei Gutachtern beurteilt worden waren. Das System sieht die Beurteilung von sieben Kriterien aus fünf Kategorien vor: (1) Testkonstruktion (Transparenz des Anwendungsbereiches, des theoretischen Hintergrunds und der Operationalisierung), (2) Qualität, (2a) des Testmaterials sowie (2b) der Verfahrenshinweise, (3) Normen, (4) Reliabilität, (5) Validität, (5a) Kriteriumsvalidität und (5b) Konstruktvalidität. Zu den Bereichen werden insgesamt fast 50 Beurteilungsaspekte in Form von konkreten Fragen formuliert, zu jedem Aspekt müssen die Gutachter ein Urteil auf einer dreistufigen Skala abgeben. Die den Beurteilungen pro Frage zugrunde gelegten Kriterien sind ausführlich operationalisiert (Evers, 2001b). Für die Testgütekriterien sowie für weitere Aspekte wie z. B. Normen gibt es Beurteilungsregeln, die für verschiedene Anwendungsbereiche unterschiedlich streng ausfallen. Auch die Kombination der Beurteilungen von Einzelfragen zu einer Gesamtbeurteilung pro Bereich ist geregelt. Auf diese Art und Weise wird jeder Test von zwei unabhängigen Gutachtern als „gut“, „ausreichend“ oder „unzureichend“ beurteilt. Bei Abweichungen in der Beurteilung wird eine Konsensbildung durch Diskussion zwischen den Gutachtern und in Ausnahmefällen durch die Berufung eines dritten Gutachters erzielt.

2.3.3.2 Das EFPA-Testbeurteilungssystem

Mittlerweile hat die EFPA eine Initiative zur Entwicklung eines „common set of European criteria for test reviews“ gestartet (Bartram, 2001, S. 180). Dieser Ansatz basiert einerseits auf dem britischen System (vgl. Bartram, Lindley & Foster, 1990, weist andererseits aber auch deutliche Ähnlichkeiten mit dem niederländischen COTAN-System (vgl. Abschnitt 2.3.3.1) auf. Der aktuelle Stand des „EFPA Review Model for the Description and Evaluation of Psychological Tests“ kann im Internet eingesehen werden (vgl. <http://www.efpa.eu>). Auch wenn die Entwicklung noch nicht abgeschlossen ist, wird sich an den nachfolgend dargestellten Grundzügen voraussichtlich nicht mehr viel ändern. Jeder Test soll demnach von zwei unabhängigen Gutachtern beurteilt werden. Im Vergleich zu dem COTAN-System legt das EFPA-System mehr Wert auf die zu Beginn der Rezension abverlangte, möglichst wertfreie Beschreibung des zu beurteilenden Tests. Es existiert eine umfangreiche Liste der dabei zu berücksichtigenden Merkmale. Hierzu zählt beispielsweise der Anwendungsbereich des Verfahrens und das Test-, Auswertungs- und Interpretationskonzept. Viel Wert

wird außerdem auf die möglicherweise angebotenen computergenerierten Berichte gelegt. Im nächsten Schritt erfolgt dann eine Bewertung einzelner Verfahrensmerkmale. Das System sieht ein Formblatt zur Beschreibung und Bewertung der Verfahren vor und bietet den Reviewern darüber hinaus eine ausführliche Instruktion, die eine einheitliche Bewertung verschiedener Aspekte sichern soll. Für die Bewertungen ist eine vierstufige Skala (von „inadequate“ bis „excellent“) vorgesehen, außerdem kann die Bewertung ausbleiben, wenn das Bewertungsmerkmal nicht sinnvoll auf das Verfahren angewendet werden kann oder wenn keine (ausreichenden) Informationen zu dem Merkmal vorliegen. Fehlende Informationen führen – ebenso wie die Einstufung als „inadequate“ – im Fall von besonders relevanten Merkmalen (Normen und Bezugsgruppen, Konstruktvalidität, Kriteriumsvalidität, Reliabilität und computergenerierte Berichte) allerdings dazu, dass von einer Anwendung des Verfahrens abgeraten wird. Damit soll sichergestellt werden, dass problematischen Verfahren keine Empfehlung ausgesprochen wird.

Das EFPA-System orientiert seine Beurteilungsrichtlinien, genauso wie das COTAN-System, an der numerischen Ausprägung von Gütekriterien und legt je nach Anwendungszweck (z. B. Entscheidungen über Individuen wie Auswahlentscheidungen, oder Gruppen) unterschiedlich strenge Maßstäbe an. Beispielsweise werden die Kennwerte für die interne Konsistenz auf einer vierstufigen Skala entsprechend ihrer Höhe als „inadequate“ bis „excellent“ klassifiziert. Der EFPA-Rezensionsprozess endet mit einer eindeutigen Anwendungsempfehlung, die sechs Kategorien unterscheidet. Die Empfehlung lässt sich nicht arithmetisch aus den Einzelurteilen herleiten.

2.3.3.3 Das Testbeurteilungssystem des Testkuratoriums

Das in Deutschland vorhandene Angebot an Informationen über und Beurteilungen von Tests kann weder quantitativ noch qualitativ überzeugen (vgl. Kersting, 2006). Das Testkuratorium (TK) der Föderation der Deutschen Psychologengvereinigungen hat daher ein System zur Information über und Beurteilung von Tests entwickelt (Testbeurteilungssystem des Testkuratoriums, TBS-TK; vgl. <http://www.zpid.de/Testkuratorium>), welches die DIN 33430 (vgl. Abschnitt 2.1.2) berücksichtigt. 2009 wurde eine geringfügig überarbeitete zweite Fassung erstellt (Testkuratorium, 2010). Das TBS-TK ersetzt die Kriterien für die Testbeurteilung der Föderation Deutscher Psychologenverbände (1986).

Das System umfasst die folgenden drei Stufen:

- Stufe 1: Prüfung der Informationsgrundlage nach DIN 33430,
- Stufe 2: Testbeschreibung nach einem vorgegebenen Raster und unter Berücksichtigung der Datenbankkompatibilität

– Stufe 3: Testbeurteilung und -rezension durch zwei unabhängige Rezensenten auf der Grundlage einer Beurteilungsrichtlinie.

Für die erste Stufe sind nur diejenigen Aussagen der DIN 33430 relevant, die sich auf die Verfahrenshinweise beziehen und die über den Anwendungsbereich Eignungsbeurteilung hinweg Geltung beanspruchen können. Grundlage der Prüfung auf Stufe 1 ist die „Checkliste 1“ der Publikation „DIN Screen“ (Kersting, 2008a). Sie gilt offiziell als „Standard zur Information und Dokumentation von Instrumenten zur Erfassung menschlichen Erlebens und Verhaltens des Testkuratoriums der Föderation Deutscher Psychologengvereinigungen“. Die erste Prüfstufe endet mit der Feststellung, ob der Test nach Ansicht der Rezensenten „prüffähig“ ist. Ein Test, der in diesem Sinne nicht prüffähig ist, erhält ohne weitere Begutachtungen eine negative Gesamtbeurteilung.

Auf der Stufe 2 erfolgt eine Testkategorisierung und es werden formale Datenbankangaben getroffen. Hierzu wird das Kategorisierungssystem des ZPID (<http://www.zpid.de/index.php?wahl=products&uwahl=fee&uwahl=prf1>) genutzt. Zur Wahrung internationaler Kompatibilität werden zusätzlich noch datenbankfähige Informationen im Sinne des EFPA-Systems verlangt.

Die eigentliche Testbeurteilung stellt die dritte Stufe des Systems dar. Hierzu hat das Testkuratorium eine Beurteilungsrichtlinie (Testkuratorium, 2010) verfasst. Durch die Richtlinie soll die Objektivität und Vergleichbarkeit der Rezensionen erhöht werden, ungeachtet davon, dass die Beurteilung ein subjektives Werturteil der Rezensenten ist und bleibt. Eine Bewertung allein aufgrund der numerischen Ausprägung der Testgütekriterien, wie sie in den bereits dargestellten COTAN- und EFPA-Systemen praktiziert wird (so formuliert das COTAN-System beispielsweise die Regel, dass eine Reliabilität kleiner als $r = .80$ für bestimmte Tests als „unzureichend“ zu bewerten sei) ist im TBS-TK-System nicht vorgesehen. Eine derartige Bewertung der Güte eines Tests aufgrund der numerischen Ausprägung von Kennwerten wäre nur dann sinnvoll, wenn die Kennwerte vollständig vom Messinstrument selbst dominiert würden. Testgütekennwerte charakterisieren nach der klassischen Testtheorie, wie Fischer bereits (1968, S. 133) ausführt, jedoch die jeweils realisierte Kombination aus einem Test einerseits und einer Untersuchungsgruppe andererseits. Die Ausprägung der Gütekriterien hängt somit nicht nur vom Test, sondern auch von der Verteilung der Parameter in der Referenzpopulation ab. Darüber hinaus handelt es sich bei Kennwerten lediglich um Einzelschätzungen (vgl. z. B. Schmidt, 1992, für ein Beispiel zur Schätzung der Kriteriumsvalidität). Eine ausführliche Darstellung der Argumente findet sich bei Kersting (2006, 2008a).

Das neue TBS-TK-System sieht die Beurteilung jeden Tests durch zwei zunächst unabhängig voneinander arbeitende Reviewer vor. In der Richtlinie werden sieben Beurteilungskategorien vorgegeben:

- Allgemeine Informationen über den Test, Beschreibung des Tests und seiner diagnostischen Zielsetzung
- Theoretische Grundlagen als Ausgangspunkt der Testkonstruktion
- Objektivität
- Normierung (Eichung)
- Zuverlässigkeit (Reliabilität, Messgenauigkeit)
- Gültigkeit (Validität)
- Weitere Gütekriterien (Störanfälligkeit, Unverfälschbarkeit und Skalierung)

Darüber hinaus soll eine Abschlussbewertung bzw. Empfehlung formuliert werden. Für die Kategorien 1, 3, 5 und 6 müssen die Rezensenten zusätzlich zu einem Freitext ihre Beurteilung auf einer vierstufigen Skala ausdrücken [„Der Test erfüllt die Anforderungen (...) (a) ‚voll‘, (b) ‚weitgehend‘, (c) ‚teilweise‘ und (d) ‚nicht‘“]. Tabelle 4 zeigt beispielhaft die Bewertungen zum Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R, Liepmann, Beauducel, Brocke & Amt-hauer, 2007; Rezension von Schmidt-Atzert & Rauch, 2008).

Tabelle 4:
Rezension nach dem Testbeurteilungssystem des Testkuratoriums (TBS-TK)

Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R; 2., erweiterte und überarbeitete Auflage)	Die TBS-TK-Anforderungen sind erfüllt			
	voll	weitgehend	teilweise	nicht
Allgemeine Informationen, Beschreibung und diagnostische Zielsetzung	x			
Objektivität	x			
Zuverlässigkeit		x		
Validität		x		

Über erste Erfahrungen mit dem System berichten Moosbrugger, Stemmler und Kersting (2008a); die nach diesem Muster erstellten Rezensionen werden frei zugänglich veröffentlicht (vgl. <http://www.zpid.de/Testkuratorium>).

2.4 Qualitätsstandards zu weiteren Verfahren (Assessment-Center, Interviews und 360-Grad-Feedbacks)

Speziell auf die Assessment-Center(AC)-Methode zugeschnittene Richtlinien wurden im deutschsprachigen Raum vom Arbeitskreis Assessment Center (AKAC) entwickelt. Die 1992 erstmals veröffentlichten deutschen „Standards

der „Assessment Center Technik“ (abgedruckt u. a. in Fisseni & Fennekels, 1995) liegen seit 2004 in der zweiten, überarbeiteten Auflage vor (AKAC, 2004, vgl. <http://www.arbeitskreis-ac.de>) Insgesamt werden neun Qualitätsstandards formuliert, die sich am Prozess der Assessment-Center-Konstruktion und -Durchführung orientieren (vgl. Neubauer & Höft, 2006). Jeder Standard zielt auf einen Prozessabschnitt, beispielsweise auf die Arbeits- und Anforderungsanalyse, die Übungskonstruktion oder die Beobachterausswahl und -vorbereitung. Der Inhalt und die Nützlichkeit jedes einzelnen Qualitätsstandards werden erläutert. Anschließend werden Hinweise zur konkreten Umsetzung des Qualitätsstandards gegeben. Während die in den Abschnitten 2.3.1. und 2.3.2 thematisierten Qualitätsstandards fast ausschließlich positiv formuliert sind, nennen die AKAC-Standards unter der Überschrift „Verstöße“ auch Beispiele für Vorgehensweisen, die dem jeweiligen Standard zuwiderhandeln. Eine weitere Besonderheit der AKAC-Standards ist die Systemperspektive. Es wird viel Wert darauf gelegt, Assessment-Center systematisch zu initiieren, in den jeweils gegebenen organisatorischen sowie personalpolitischen Rahmen zu integrieren und beispielsweise mit anderen Instrumenten der Personalentwicklung zu verknüpfen.

Die Qualitätsstandards des Arbeitskreises Assessment Center Schweiz entsprechen weitgehend den deutschen Standards (http://www.akac.ch/files/ac_standards_akac.pdf). Weitere nationale Assessment-Center-Qualitätsstandards liegen nach Auskunft der internationalen „Task force on Assessment Center“ (2009, S. 252) für Südamerika sowie Indonesien vor.

Die von der Task Force on Assessment Center erarbeiteten internationalen Guidelines (2009) für Assessment-Center liegen nun, nach Publikationen in 1975, 1979, 1989 und 2000 in einer fünften Fassung vor. Zusätzlich zu den bisherigen Aspekten wird in der aktuellen Version auch die Entwicklung und der Einsatz von Assessment-Centern in unterschiedlichen Sprach- und Kulturräumen thematisiert. Ein Vergleich dieser Guidelines mit den AKAC-Qualitätsstandards zeigt, dass auch innerhalb der Expertenkreise keine einheitlichen Vorstellungen über Qualität existieren. Während in den deutschen AKAC-Qualitätsstandards beispielsweise die Simulation von mindestens drei verschiedenen Arbeitssituationen gefordert wird, begnügen sich die US-Amerikaner, zumindest für einfache Berufstätigkeiten, mit nur einer Simulationsaufgabe (Task Force on Assessment Center, 2009, S. 246). In den deutschen AKAC-Qualitätsstandards werden psychometrische Tests und Interviews als „Nichtsimulationen“ bezeichnet und tendenziell als zweite Wahl klassifiziert. Dem steht die folgende Aussage der amerikanischen Guidelines gegenüber: „Multiple assessment techniques must be used“ (Task Force on Assessment Center, 2009, S. 245). Als Beispiele für solche Techniken werden u. a. Tests, Interviews, Fragebogen und Simulationen genannt, d. h. für den von den deutschen AKAC-Qualitätsstandards eher empfohlenen

alleinigen Einsatz von Simulationen lässt sich im amerikanischen Standard kein Äquivalent finden. Andere Unterschiede beziehen sich beispielsweise auf den Einsatz von technischen Hilfsmitteln (z. B. videobasiertes oder virtuelles Stimulusmaterial oder EDV Unterstützung für die Datenintegration), die in den nordamerikanischen, nicht aber in den deutschen AKAC-Qualitätsstandards thematisiert werden. Teilnehmermerkmale wie Alter, ethnischer und kultureller Hintergrund und Geschlecht werden in den Guidelines viel deutlicher thematisiert als in den AKAC-Qualitätsstandards.

Im Jahre 2008 hat der Arbeitskreis Assessment Center erstmalig auch „Interview-Standards“ publiziert, die ähnlich aufgebaut sind wie die Standards der Assessment-Center-Technik (vgl. ebenfalls <http://www.arbeitskreis-ac.de>).

Darüber hinaus finden sich Qualitätsstandards zu vereinzelt Verfahren, so haben beispielsweise verschiedene Institutionen aus Großbritannien Richtlinien für 360-Grad-Feedback erarbeitet (Chartered Institute of Personnel and Development et al., 2004, siehe <http://www.psychtesting.org.uk>).

3 Die DIN 33430 und ihr Bezug zu den Standards für pädagogisches und psychologisches Testen

Die Standards für pädagogisches und psychologisches Testen sind das Bedeutendste internationale Regelwerk der Diagnostik, in Deutschland stellt die DIN 33430 das am weitesten ausgearbeitete Qualitätssystem der Diagnostik dar. Im Folgenden sollen die zentralen Aussagen der DIN 33430 wiedergegeben und so weit möglich den Standards zugeordnet werden. Daran schließt sich im vierten Kapitel ein Vergleich der beiden Regelwerke an. Ziel ist es, die Gemeinsamkeiten und Eigenständigkeiten der beiden Qualitätsstandards herauszuarbeiten, um einerseits Basisforderungen der Qualität zu identifizieren und andererseits aus der Reflexion der Unterschiede zu lernen. Die Gegenüberstellung und der Vergleich erscheinen zunächst als unstimmig, da sich die DIN 33430 nur auf Eignungsbeurteilungen bezieht, während die Standards für alle diagnostischen Anwendungsbereiche gelten. Die folgenden Gesichtspunkte sprechen allerdings für die Gegenüberstellung: Wie oben ausgeführt hat Kersting (2008a) eine Checkliste zur DIN 33430 erstellt und dafür den Text der DIN 33430 in Einzelaussagen zerlegt sowie thematisch geordnet. Die Checkliste umfasst sechs thematische Einzellisten mit 311 eigenständigen Aussagen, von denen sechs Aussagen in zwei Listen abgedruckt sind und vier andere Aussagen (Nr. 56 und 57 sowie Nr. 295 und 296) zusätzlich in zwei weiteren (Nr. 142 sowie Nr. 190) wiederholt werden, sodass sich insgesamt 318 Aussagen ergeben. Bei der Erstel-

lung der Checklisten zur DIN 33430 stellte sich heraus, dass 140 Aussagen (44 % aller Aussagen) der DIN 33430 Anforderungen an Verfahrenshinweise (auch „Handanweisungen“ oder in Bezug auf Tests „Testmanuale“ genannt) formulieren. Diese Teilmenge der DIN 33430 wurde von Kersting (2008a) zu einer eigenständigen Checkliste „Anforderungen an Verfahrenshinweise“ zusammengefasst, diese Teilmenge kann auf Tests aus allen Anwendungsbereichen angewendet werden und hat somit einen den Standards entsprechenden Geltungsanspruch. Entsprechend ist sie Bestandteil des Testbeurteilungssystems des Testkuratoriums (vgl. Abschnitt 2.3.3.3). Auch andere Forderungen der DIN 33430, z. B. zur Durchführung, Auswertung und Interpretation diagnostischer Verfahren haben allgemeinen Charakter und beschränken sich nicht auf die Eignungsdiagnostik.

Nachfolgend werden die normativen Aussagen der DIN 33430 in Form der „DIN-Screen“-Checkliste (Kersting, 2008a) wiedergegeben. Sofern möglich werden den DIN-Aussagen die Nummern der korrespondierenden Standards zugeordnet. In der mittleren Spalte der Tabelle ist mit einem doppelten oder einfachen Pluszeichen eine hohe bzw. mittlere Korrespondenz der deutschen mit den nordamerikanischen Standards gekennzeichnet, ein fehlender Eintrag verweist auf die fehlende Übereinstimmung. Eine Auswertung der Gegenüberstellung liefert Kapitel 4.

Kasten 1:

Beispiel für eine hohe Korrespondenz zwischen Aussagen der DIN 33430 und einem Standard

Aussagen der DIN 33430 (nach Kersting, 2008a, S. 161)

- V27 Verzweigungsfrage: Wurden zur Bestimmung der Gültigkeit Methoden der statistischen Adjustierung angewandt (z. B. Minderungskorrektur oder Variabilitätskorrektur)? Falls „nein“ → Bitte weiter bei Verzweigungsfrage V28.
- 92 Bei den Analysen zur Gültigkeit sind sowohl die ursprünglich erhaltenen als auch die korrigierten Kennwerte aufgeführt. (DIN, 2002, S. 16)
- 93 Alle im Zusammenhang mit der Adjustierung verwendeten Statistiken sind genannt. (DIN, 2002, S. 16)
- 94 Neben den statistisch optimierten Schätzungen (z. B. multiple Regression) sind auch die einfachen Schätzungen (z. B. einfache Korrelationen) angegeben. (DIN, 2002, S. 16)

Korrespondierender Standard

Standard 1.18:

When statistical adjustments, such as those for restriction of range or attenuation, are made, both adjusted and unadjusted coefficients, as well as the specific procedure used, and all statistics used in the adjustment, should be reported. (AERA et al., 1999, S. 21 f.)

Häufig werden einem Standard, wie im Beispiel in Kasten 1, mehrere DIN-Aussagen zugeordnet. Die DIN-Screen-Aussagen sind feiner aufgelöst und beinhalten nach Möglichkeit jeweils nur eine Forderung, während die Standards umfassendere Themen repräsentieren. Gelegentlich kann aber auch ein Standard mehreren DIN-Aussagen zugeordnet werden.

Kasten 2:

Beispiel für eine mittlere Korrespondenz zwischen einer Aussage der DIN 33430 und einem Standard

<i>DIN-Aussage (nach Kersting, 2008a, S. 161)</i>	
Die Kandidaten werden zu Beginn der Untersuchung aufgeklärt ... 241 ... über mögliche Folgen mangelnder Kooperation (DIN, 2002, S. 13)	
<i>Korrespondierender Standard: 8.7</i>	
Test takers should be made aware that having someone else take the test for them, disclosing confidential test material, or any other form of cheating is inappropriate and that such behavior may result in sanctions. (AERA et al., 1999, S. 88)	

Kasten 2 gibt ein Beispiel für eine mittlere Korrespondenz zwischen einem Standard und einer Aussage der DIN 33430. Beide Regelwerke thematisieren die mögliche mangelhafte Kooperation der Teilnehmer, die Standards wählen dafür aber ein konkretes unkooperatives Verhalten, während die DIN allgemein von „mangelnder Kooperation“ spricht. Daher wird die Korrespondenz der beiden Aussagen als „mittel“ eingestuft.

Sofern eine DIN-Aussage mehreren Standards zugeordnet wird, wird die Korrespondenz in der Regel nur als „mittel“ eingeschätzt, die DIN-Aussage kann zu einigen der genannten Standards eine sehr hohe Korrespondenz aufweisen, in der Regel aber nicht zu allen. Aus diesem Grunde verzeichnet die nachfolgende Aufstellung kaum hohe Korrespondenzen, gleichwohl diese auf der Ebene der Beziehung einzelner DIN-Aussagen zu einzelnen Standards sehr wohl vorkommen. Es gibt eine weitere, hier nicht dargestellte Zuordnungstabelle, bei der als Ausgangspunkt der Zuordnung die Standards gewählt werden. Diese Tabelle war Ausgangspunkt für die hier dargestellte (Vergleichs-)Tabelle 5, die beim Erstautor angefordert werden kann.

In Tabelle 5 werden die 318 Aussagen der DIN in Form der DIN-Screen-Aufbereitung (Kersting, 2008a) wiedergegeben und um eine Zuordnung der Standards ergänzt.

Tabelle 5:

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zu Anforderungen an Verfahrenshinweise (Checkliste 1)			
	<i>DIN 33430, Aussagen zu Grundsätzen/zur Informationspolitik</i>	<i>Korrespondenz</i>	<i>Standards</i>
1	Für jedes eingesetzte Verfahren liegen Verfahrenshinweise (Handhabungshinweise) vor. (DIN, S. 6)	+	6.1; 11.5
2	Die Verfahrenshinweise sind dem Anwender der Verfahren <u>und</u> in Sonderfällen auch Außenstehenden zugänglich. (DIN, S. 14)	+	6.1
3	Alle Quellen, die vom Autor bzw. Vertreter des Verfahrens in den Verfahrenshinweisen zitiert werden, sind zugänglich. (DIN, S. 14)	+	6.9
V1	<i>Verzweigungsfrage: Gibt es allgemein nicht zugängliche zusätzliche Informationen, die erst nach der Herausgabe der Verfahrenshinweise bekannt geworden sind <u>und</u> die für eine Beurteilung der Tauglichkeit des Verfahrens herangezogen werden können? Falls „nein“ → Bitte weiter bei Verzweigungsfrage V3.</i>		
V2	<i>Verzweigungsfrage: Wurde der Autor bzw. wurden Herausgeber bzw. die Vertreter des Verfahrens gebeten, die allgemein nicht zugänglichen zusätzlichen Informationen, die erst nach der Herausgabe der Verfahrenshinweise bekannt geworden sind <u>und</u> die für eine Beurteilung der Tauglichkeit des Verfahrens herangezogen werden können, zur Verfügung zu stellen? Falls „nein“ → Bitte weiter bei Verzweigungsfrage V3.</i>		
4	Zusätzliche, allgemein nicht zugängliche Informationen, die erst nach der Herausgabe der Verfahrenshinweise bekannt geworden sind <u>und</u> die für eine Beurteilung der Tauglichkeit des Verfahrens herangezogen werden können, werden auf Nachfrage von mindestens einer der drei folgenden Gruppen zur Verfügung gestellt: Entwickler, Herausgeber und/oder Verleger des Verfahrens. (DIN, S. 14)	+	6.15
V3	<i>Verzweigungsfrage: Werden empirische Belege für die Leistungsfähigkeit des Verfahrens zitiert, die noch nicht veröffentlicht sind? Falls „nein“ → Bitte weiter bei Verzweigungsfrage V5.</i>		
V4	<i>Verzweigungsfrage: Wurde der Autor bzw. wurden Herausgeber bzw. die Vertreter des Verfahrens gebeten, die noch nicht veröffentlichten empirischen Belege für die Leistungsfähigkeit des Verfahrens zur Verfügung zu stellen? Falls „nein“ → Bitte weiter bei Verzweigungsfrage V5.</i>		

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zu Anforderungen an Verfahrenshinweise (Checkliste 1)			
	DIN 33430, Aussagen zu Grundsätzen/zur Informationspolitik	Korrespondenz	Standards
5	Die nicht veröffentlichten empirischen Belege für die Leistungsfähigkeit des Verfahrens werden auf Anfrage zur Verfügung gestellt. (DIN, S. 14)	+	6.15
V5	<i>Verzweigungsfrage: Sind dem Prüfer Studien oder Rezensionen anderer Autoren zum Verfahren bekannt? Falls „nein“ → Bitte weiter bei Aussage 8.</i>		
6	In den Verfahrenshinweisen werden Studien, die von anderen Autoren zum Verfahren oder mit dem Verfahren durchgeführt wurden, berichtet. (DIN, S. 6)	+	6.9
7	In den Verfahrenshinweisen wird auf Rezensionen, die von anderen Autoren zum Verfahren geschrieben wurden, hingewiesen. (DIN, S. 6)	+	6.9
8	In den Verfahrenshinweisen werden Angaben zur notwendigen weiteren Verfahrenspflege getroffen. (DIN, S. 6)		
V6	<i>Verzweigungsfrage: Gibt es Bedarf, die Verfahrenshinweise zu berichtigen und/oder zu ergänzen und/oder zu überarbeiten? Falls „nein“ → Bitte weiter bei Aussage 11.</i>		
9	Es existiert eine berichtigte oder ergänzte oder überarbeitete Version der Verfahrenshinweise. (DIN, S. 14)	+	5.14; 6.13
10	In der überarbeiteten Version der Verfahrenshinweise ist angegeben, was an der Ausgangsversion verändert wurde <u>und</u> warum. (DIN, S. 14)	+	6.13
11	Alle in jeglicher Form der Publikation (z. B. Artikel, Kurzdarstellungen, Werbematerialien) dargebotenen Informationen zum Verfahren sind wahrheitsgetreu <u>und</u> auf Anforderung belegbar. (DIN, S. 14)		
12	Die Vertreter <u>und</u> Anwender des Verfahrens vermeiden Werbestrategien, die nahe legen, dass ein Verfahren mehr oder anderes leistet, als aufgrund der empirischen bzw. theoretischen Grundlagen des Verfahrens belegt werden konnte. (DIN, S. 14)		

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zu Anforderungen an Verfahrenshinweise (Checkliste 1)			
	DIN 33430, Aussagen zu Grundsätzen/zur Informationspolitik	Korrespondenz	Standards
13	Durch den Vertrieb des Verfahrens bzw. der dazugehörigen detaillierten Informationen ist sichergestellt, dass die Kandidaten in gleichem Maße Zugang oder keinen Zugang zu Vorinformationen, Kenntnis einzelner Items u. a. haben (Verfahrensschutz). (DIN, S. 14)	+	5.6; 15.9
V7	<i>Verzweigungsfrage: Erfordert die Handhabung des Verfahrens besondere Qualifikationen? Falls „nein“ → Bitte weiter bei Aussage 15.</i>		
14	Die für die Handhabung erforderlichen besonderen Qualifikationen werden genannt. (DIN, S. 14)	+	3.23; 6.7
	DIN 33430, Aussagen zu spezifischen Informationsanforderungen	Korrespondenz	Standards
15	In den Verfahrenshinweisen sind die Anwendungsbereiche des Verfahrens verständlich benannt. (DIN, S. 6)	+	3.2; 6.3; 11.5; 14.1
16	In den Verfahrenshinweisen ist die Zielsetzung des Verfahrens verständlich benannt. (DIN, S. 6)	+	3.2; 6.4; 11.5; 14.1
17	Die Verfahrenshinweise enthalten Angaben zu den Zielsetzungen des Verfahrens, die es dem Anwender ermöglichen, den Beitrag des Verfahrens zum Ziel der Verfahrensanwendung zu erkennen. (DIN, S. 15)	+	3.2; 11.5; 14.1
V8	<i>Verzweigungsfrage: Sind missbräuchliche Anwendungen des Verfahrens naheliegend? Falls „nein“ → Bitte weiter bei Verzweigungsfrage V9.</i>		
18	Die Verfahrenshinweise geben warnende Hinweise auf missbräuchliche Anwendungen des Verfahrens. (DIN, S. 15)	+	1.3; 4.3; 6.3
V9	<i>Verzweigungsfrage: Gibt es Verfahrensanwendungen, die nicht bzw. nicht mehr gerechtfertigt sind? Falls „nein“ → Bitte weiter bei Verzweigungsfrage VI1.</i>		
19	Die Fachöffentlichkeit wird auf Verfahrensanwendungen hingewiesen, die nicht bzw. nicht mehr gerechtfertigt sind. (DIN, S. 15)	+	3.25

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zu Anforderungen an Verfahrenshinweise (Checkliste 1)			
	DIN 33430, Aussagen zu spezifischen Informationsanforderungen	Korrespondenz	Standards
V10	<i>Verzweigungsfrage: Wurde der Autor bzw. wurden Herausgeber bzw. die Verreiber des Verfahrens gefragt, ob es Verfahrensanwendungen gibt, die nicht bzw. nicht mehr gerechtfertigt sind? Falls „nein“ → Bitte weiter bei Verzweigungsfrage V11</i>		
20	Auf Anfrage hin wurde deutlich gemacht, welche Verfahrensanwendungen nicht bzw. nicht mehr gerechtfertigt sind. (DIN, S. 15)		
V11	<i>Verzweigungsfrage: Kann das Verfahren lediglich für Erprobungs- und/oder für Forschungszwecke verwendet werden? Falls „nein“ → Bitte weiter bei Aussage 22</i>		
21	In den Verfahrenshinweisen wird explizit darauf hingewiesen, dass das Verfahren lediglich für Erprobungs- und/oder für Forschungszwecke verwendet werden kann. (DIN, S. 15)		
22	In den Verfahrenshinweisen ist angemessen (im Sinne von ausführlich <u>und</u> verständlich <u>und</u> nachvollziehbar) dargestellt, wie das standardisierte Verfahren konstruiert wurde. (DIN, S. 6)	+	3.3; 3.7; 6.4
	Durch die Verfahrenshinweise wird deutlich, wie Fehler in den nachfolgend genannten Bereichen vermieden werden können:		
23	– Durchführung des Verfahrens (DIN, S. 6)	+	3.3; 11.5
24	– Auswertung des Verfahrens (DIN, S. 6)	+	3.3; 11.5
25	– Interpretation des Verfahrens. (DIN, S. 6)	+	3.4; 11.5
	Die Verfahrenshinweise liefern Informationen, aus denen der Anwender den hinsichtlich der folgenden Aspekte entstehenden Aufwand abschätzen kann:		
26	– Material (DIN, S. 14)	+	3.3
27	– Personal (DIN, S. 14)	+	3.3
28	– Räumlichkeiten. (DIN, S. 14)	+	3.3

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zu Anforderungen an Verfahrenshinweise (Checkliste 1)			
	DIN 33430, Aussagen zu spezifischen Informationsanforderungen	Korrespondenz	Standards
	Die Verfahrenshinweise liefern Informationen, aus denen der Anwender den hinsichtlich der folgenden Aspekte entstehenden zeitlichen Aufwand abschätzen kann:		
29	zeitliche Belastung für den Kandidaten (DIN, S. 15)	+	3.3
30	zeitliche Belastung für den Anwender bei der Vorbereitung (DIN, S. 15)	+	3.3
31	zeitliche Belastung für den Anwender bei der Durchführung (DIN, S. 15)	+	3.3
32	zeitliche Belastung für den Anwender bei der Auswertung. (DIN, S. 15)	+	3.3
33	In den Verfahrenshinweisen werden die Ergebnisse einer oder mehrerer empirischen/empirischer Untersuchung(en) berichtet. (DIN, S. 6)	+	3.6; 6.9
	<i>DIN 33430, Aussagen zu erforderlichen Hinweise zur Verfahrensdurchführung, -auswertung und -interpretation</i>	<i>Korrespondenz</i>	<i>Standards</i>
34	In den Verfahrenshinweisen werden Regeln aufgestellt, wie mit vorhersehbaren Nachfragen der Verfahrensteilnehmer umgegangen wird. (DIN, S. 8)	+	3.19
V12	<i>Verzweigungsfrage: Handelt es sich um ein computergestütztes Verfahren? Falls „nein“ → Bitte weiter bei Aussage 36</i>		
V13	<i>Verzweigungsfrage: Ist es notwendig, bei der Durchführung spezifische technische Vorgaben (z. B. Anforderungen an Hardware wie Bildschirm, Hilfsmittel) zu berücksichtigen? Falls „nein“ → Bitte weiter bei Aussage 36</i>		
35	In den Verfahrenshinweisen werden technische Vorgaben (z. B. Anforderungen an Hardware wie Bildschirm, Hilfsmittel) genannt. (DIN, S. 8)		
36	Es wird beschrieben, welche situativen Rahmenbedingungen für den störungsfreien Einsatz des Verfahrens notwendig sind. (DIN, S. 14)	+	5.4

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zu Anforderungen an Verfahrenshinweise (Checkliste 1)			
	DIN 33430, Aussagen zu erforderlichen Hinweisen zur Verfahrensdurchführung, -auswertung und -interpretation	Korrespondenz	Standards
37	Die zu befürchtenden/möglichen personenbedingten Störungen des Verfahrenseinsatzes <u>und</u> deren Auswirkungen werden benannt. (DIN, S. 14)		
V14	Verzweigungsfrage: Werden apparativ gestützte Verfahren eingesetzt? Falls „nein“ → Bitte weiter bei Aussage 39.		
38	Die zu befürchtenden/möglichen apparativen Störungen des Verfahrenseinsatzes <u>und</u> deren Auswirkungen werden benannt. (DIN, S. 14)		
39	Es ist angegeben, wie <u>und</u> in welchem Ausmaß potenzielle Störungen des Verfahrenseinsatzes kompensiert werden können. (DIN, S. 14)		
40	Die Maßnahmen, mit denen potenzielle Störungen des Verfahrenseinsatzes kompensiert werden können, sind beschrieben. (DIN, S. 14)		
41	Es wird thematisiert, ob Faktoren auf die Verfahrensbearbeitung und/oder auf das Verfahrensergebnis Einfluss nehmen können, die nicht zu den personenbedingten und/oder apparativen Störungen zählen. (DIN, S. 14)		
42	Es ist angegeben, mit welchen Verfälschungsmöglichkeiten zu rechnen ist. (DIN, S. 14)	+	5.6; 15.9
43	Es ist ausgeführt, wie einer Verfälschung durch die Art der Verfahrensvorgabe <u>und</u> -durchführung – sowie ggf. auch bei der Auswertung – entgegengewirkt werden kann. (DIN, S. 14)	+	5.6; 15.9
V15	Verzweigungsfrage: Handelt es sich um Verfahrenshinweise für schriftliche Testverfahren? Falls „nein“ → Bitte weiter bei Aussage 45.		
44	In den Verfahrenshinweisen werden Regeln aufgestellt, wie bei der Auswertung mit nicht bearbeiteten Fragen bzw. (Teil-)Aufgaben umgegangen wird. (DIN, S. 8)		

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zu Anforderungen an Verfahrenshinweise (Checkliste 1)			
	DIN 33430, Aussagen zu Anforderungen an die Darstellung empirischer Untersuchungen	Korrespondenz	Standards
45	Alle in den Verfahrenshinweisen aufgeführten relevanten empirischen Untersuchungen sind nachvollziehbar beschrieben. (DIN, S. 6)	+	1.5; 1.13; 2.4; 3.8; 4.6; 6.4
	Der Bericht über empirische Untersuchungen enthält:		
46	– eine Angabe über das Jahr der Datenerhebung (DIN, S. 14)	+	1.5; 1.13; 2.4; 3.8; 4.6; 6.4
47	– deskriptive Statistiken über die Merkmale der Untersuchungsteilnehmer (DIN, S. 14)	+	1.5; 1.13; 2.4; 3.8; 4.6; 6.4
48	– eine Information über den Stichprobenplan (DIN, S. 14)	+	1.5; 1.13; 2.4; 3.8; 4.6; 6.4
49	– eine Information zu den Teilnehmerquoten. (DIN, S. 14)	+	1.5; 1.13; 2.4; 3.8; 4.6; 6.4
	Die empirische Arbeit ist in den Verfahrenshinweisen so dargestellt, dass eine kritische Würdigung der Ergebnisse dieser empirischen Arbeit hinsichtlich ihrer ...		
50	– theoretischen Grundlagen möglich ist (DIN, S. 15)		
51	– methodischen Grundlagen möglich ist. (DIN, S. 15)		
52	Die Dokumentation der empirischen Arbeit entspricht den üblichen Kriterien für wissenschaftliche Publikationen. (DIN, S. 15)		
	DIN 33430, Aussagen zu theoretischen Grundlagen psychometrischer Verfahren	Korrespondenz	Standards
V16	Verzweigungsfrage: Handelt es sich bei dem eingesetzten Verfahren um ein psychometrisches Verfahren? Falls „nein“ → Bitte weiter bei Verzweigungsfrage V18.		

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zu Anforderungen an Verfahrenshinweise (Checkliste 1)			
	DIN 33430, Aussagen zu Anforderungen an die Darstellung empirischer Untersuchungen	Korrespondenz	Standards
53	Die theoretischen Grundlagen des Verfahrens sind hinreichend ausführlich beschrieben. (DIN, S. 15)	+	6.3
54	Die Grundkonzeption ist ohne zusätzliche Sekundärliteratur anhand der Verfahrenshinweise erkennbar. (DIN, S. 15)		
V17	Verzweigungsfrage: Hat der Autor Modifikationen an etablierten theoretischen Vorstellungen vorgenommen? Falls „nein“ → Bitte weiter bei Verzweigungsfrage V18.		
55	In den Verfahrenshinweisen wird verdeutlicht, welche Modifikationen an etablierten theoretischen Vorstellungen der Autor vorgenommen hat. (DIN, S. 15)		
	DIN 33430, Aussagen zu Normwerten/Referenzkennwerten	Korrespondenz	Standards
V18	Verzweigungsfrage: Handelt es sich um Verfahrenshinweise zu einem Verfahren, das einen Vergleich mit Normwerten vorsieht? Falls „nein“ → Bitte weiter bei Aussage 61.		
56	Die herangezogenen Normwerte entsprechen der Fragestellung. (DIN, S. 7)	+	4.5
57	Die herangezogenen Normwerte entsprechen der Referenzgruppe der Zielgruppe. (DIN, S. 7)	+	4.5
58	Die Angemessenheit der Normwerte wurde in den letzten 8 Jahren überprüft. (DIN, S. 7)	+	4.18
V19	Verzweigungsfrage: Zielt der Verfahrenseinsatz auf die Erfassung eines Merkmals ab, dessen Ausprägung in der Referenzgruppe möglicherweise relativ kurzfristigen Veränderungen unterliegt? Falls „nein“ → Bitte weiter bei Verzweigungsfrage 21.		
V20	Verzweigungsfrage: Ist das Verfahren mindestens 8 Jahre alt? Falls „nein“ → Bitte weiter bei Verzweigungsfrage 21.		
59	Noch vor Ablauf der 8-Jahres-Frist wurde eine Neunormierung vorgenommen. (DIN, S. 7)	+	4.18

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zu Anforderungen an Verfahrenshinweise (Checkliste 1)			
	DIN 33430, Aussagen zu Normwerten/Referenzkennwerten	Korrespondenz	Standards
V21	Verzweigungsfrage: Gibt es unterschiedliche „Vorgabearten“ des Verfahrens, also z. B. eine Papier-Bleistift-Version und eine Computerversion? Falls „nein“ → Bitte weiter bei Aussage 61.		
V22	Verzweigungsfrage: Wurden Normwerte von einer Vorgabeart auf die andere übertragen? Falls „nein“ → Bitte weiter bei Aussage 61.		
60	Die Übertragbarkeit der Normwerte von einer Vorgabeart auf die andere wurde nachgewiesen. (DIN, S. 7)	+	4.10; 6.11
	DIN 33430, Aussagen zur Objektivität	Korrespondenz	Standards
61	In den Verfahrenshinweisen werden Angaben zur Objektivität des Verfahrens gemacht. (DIN, S. 6)		
62	In den Verfahrenshinweisen werden die eingesetzten Analysemethoden zur Gewinnung der Angaben zur Objektivität nachvollziehbar dokumentiert. (DIN, S. 6)		
	Die Verfahrenshinweise sind so gestaltet, dass verschiedene Personen in der Lage sind, das Verfahren allein aufgrund dieser Verfahrenshinweise in vergleichbarer Weise ...		
63	– durchzuführen (DIN, S. 6)	+	3.19
64	– auszuwerten (DIN, S. 6)		
65	– zu interpretieren. (DIN, S. 6)	+	1.2; 6.3
66	Die Instruktionen des Verfahrens sind so gestaltet, dass die Ergebnisse so wenig wie möglich durch den Kandidaten selbst verfälscht werden können. (DIN, S. 6/7)	+	5.6; 15.9
67	Die Vorgehensweise bei der Beurteilung ist so gestaltet, dass die Ergebnisse so wenig wie möglich durch den Kandidaten selbst verfälscht werden können. (DIN, S. 6/7)	+	5.6; 15.9

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zu Anforderungen an Verfahrenshinweise (Checkliste 1)			
	DIN 33430, Aussagen zur Zuverlässigkeit (Reliabilität)	Korrespondenz	Standards
68	In den Verfahrenshinweisen werden Angaben zur Zuverlässigkeit des Verfahrens gemacht. (DIN, S. 6)	+	2.1; 6.5
69	In den Verfahrenshinweisen werden die eingesetzten Analysemethoden zur Gewinnung der Angaben zur Zuverlässigkeit nachvollziehbar dokumentiert. (DIN, S. 6)	+	2.4; 6.5
70	Es ist angegeben, nach welcher/welchen Methode(n) die Zuverlässigkeit bestimmt wurde. (DIN, S. 15)	++	2.4
71	Die Angemessenheit der für die Zuverlässigkeitsbestimmung genutzten Methode(n) ist für verschiedene Typen von Beurteilungen beispielhaft erläutert. (DIN, S. 15)	++	2.4
V23	<i>Verzweigungsfrage: Gibt es Erkenntnisse darüber, dass sich die Zuverlässigkeitswerte bzw. Standardmessfehler für verschiedene (sozio-)demografische Gruppen (z. B. nach Alter, Geschlecht, Ausbildung, Nationalität) maßgeblich voneinander unterscheiden? Falls „nein“ → Bitte weiter bei Aussage 73.</i>		
72	Für alle Gruppen, für die das Verfahren empfohlen wird, werden verschiedene Zuverlässigkeitswerte angegeben. (DIN, S. 15)	++	2.11; 2.12
73	In den Verfahrenshinweisen wird beschrieben, wie die zur Zuverlässigkeitsbestimmung herangezogenen Untersuchungsgruppen zusammengesetzt waren. (DIN, S. 15)	++	2.4
V24	<i>Verzweigungsfrage: Sollen mit dem Verfahren Merkmale erfasst werden, für die eine zumindest relative Zeit- und Situationsstabilität angenommen wird? Falls „nein“ → Bitte weiter bei Aussage 76.</i>		
74	Die Zuverlässigkeit wurde über die Retest-Methode bestimmt oder die Retest-Reliabilität wurde durch einen geeigneten Untersuchungsplan geschätzt. (DIN, S. 15)		
75	Das zur Bestimmung der Retest-Reliabilität gewählte Vorgehen wird begründet. (DIN, S. 15)		
76	Der aktuellste Nachweis der Geltung der Zuverlässigkeitskennwerte ist jünger als 8 Jahre. (DIN, S. 15)		

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zu Anforderungen an Verfahrenshinweise (Checkliste 1)			
	DIN 33430, Aussagen zur Gültigkeit (Validität)	Korrespondenz	Standards
77	In den Verfahrenshinweisen werden Angaben zur Gültigkeit des Verfahrens gemacht. (DIN, S. 6)	+	6.5
78	In den Verfahrenshinweisen werden die eingesetzten Analysemethoden zur Gewinnung der Angaben zur Gültigkeit nachvollziehbar beschrieben. (DIN, S. 6 und S. 16)	+	6.5
79	In den Verfahrenshinweisen wird nachvollziehbar begründet (bzw. es ist evident), dass die gewählte Art der Gültigkeitsbestimmung dem Zweck des Verfahrens <u>und</u> der vorliegenden Fragestellung angemessen ist. (DIN, S. 7)		
80	Aus den Verfahrenshinweisen wird deutlich, welche empirischen Nachweise der Inhalts- und/oder Kriteriums- und/oder Konstruktgültigkeit eine Anwendung des Verfahrens für den intendierten Anwendungszweck rechtfertigen. (DIN, S. 16)	+	1.4; 14.1
81	Die Ausführungen zur Gültigkeit, die die Anwendung des Verfahrens für den intendierten Anwendungszweck rechtfertigen, berücksichtigen die diagnostische(n) Fragestellung(en) <u>und</u> die Zielgruppe(n), für die das Verfahren empfohlen wird. (DIN, S. 16)	+	1.4; 14.1
82	Die Wahl der Methoden, die bei der Gültigkeitsanalyse zum Einsatz gekommen sind, wird begründet. (DIN, S. 16)		
	In den Verfahrenshinweisen ist angegeben, welche Gültigkeitswerte...		
83	– für welches Verfahrensergebnis erzielt wurden (DIN, S. 16)		
84	– in Bezug zu welchem Bewährungskriterium erzielt wurden (DIN, S. 16)		
85	– für welche Referenzgruppe erzielt wurden (DIN, S. 16)		
86	– in welcher Situation erzielt wurden (DIN, S. 16)		
87	– in welcher Untersuchung erzielt wurden (DIN, S. 16)		
88	– zu welchem Zeitpunkt erzielt wurden. (DIN, S. 16)		

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zu Anforderungen an Verfahrenshinweise (Checkliste 1)			
	DIN 33430, Aussagen zur Gültigkeit (Validität)	Korrespondenz	Standards
V25	Verzweigungsfrage: Ist aus den Verfahrenshinweisen ersichtlich oder aus anderen Quellen bekannt, dass für einige Untersuchungsteilnehmer Ergebnisse zur Verfahrensgültigkeit fehlen? Falls „nein“ → Bitte weiter bei Aussage 90.		
89	Es wird erläutert, warum für einige Untersuchungsteilnehmer keine Ergebnisse zur Verfahrensgültigkeit vorliegen. (DIN, S. 16)		
90	Der aktuellste Nachweis über die Gültigkeit des Verfahrens für den intendierten Anwendungsbereich ist jünger als 8 Jahre. (DIN, S. 15 und S. 16)		
V26	Verzweigungsfrage: Wurde das Verfahren nach der Gültigkeitsanalyse, deren Ergebnisse in den Verfahrenshinweisen dargestellt sind, wesentlich geändert? (DIN, S. 16) Falls „nein“ → Bitte weiter bei Verzweigungsfrage V27.		
91	Für das geänderte Verfahren wurden neue Gültigkeitsbelege erbracht. (DIN, S. 16)		
V27	Verzweigungsfrage: Wurden zur Bestimmung der Gültigkeit Methoden der statistischen Adjustierung angewandt (z. B. Minderungskorrektur oder Variabilitätskorrektur)? Falls „nein“ → Bitte weiter bei Verzweigungsfrage V28.		
92	Bei den Analysen zur Gültigkeit sind sowohl die ursprünglich erhaltenen als auch die korrigierten Kennwerte aufgeführt. (DIN, S. 16)	++	1.18
93	Alle im Zusammenhang mit der Adjustierung verwendeten Statistiken sind genannt. (DIN, S. 16)	++	1.18
94	Neben den statistisch optimierten Schätzungen (z. B. multiple Regression) sind auch die einfachen Schätzungen (z. B. einfache Korrelationen) angegeben. (DIN, S. 16)	++	1.18
95	Die optimierten Schätzungen konnten an einer anderen Personengruppe aus dem Geltungsbereich des Verfahrens repliziert werden. (DIN, S. 16)		

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zu Anforderungen an Verfahrenshinweise (Checkliste 1)			
	DIN 33430, Aussagen zur Gültigkeit (Validität)	Korrespondenz	Standards
96	Die statistischen Optimierungsprozeduren wurden in handlungsleitende Beurteilungsregeln umgesetzt. (DIN, S. 16)		
V28	Verzweigungsfrage: Wird der Gültigkeitsanspruch damit begründet, dass Gültigkeitshinweise aus anderen Untersuchungen in Anspruch genommen werden (Validitätsgeneralisierung)? Falls „nein“ → Bitte weiter bei Verzweigungsfrage V29.		
	In den Verfahrenshinweisen ist ausführlich dargestellt <u>und</u> begründet, ...		
97	welche Befunde generalisiert werden können (DIN, S. 16)	++	1.20; 14.6
98	weshalb (<u>und</u> in welchem Ausmaß) sich die Gültigkeitshinweise übertragen lassen, die sich aus anderen Studien ergeben. (DIN, S. 16)	++	1.20; 14.6
99	Auf die Ähnlichkeit zwischen den vorliegenden Studien und der vorliegenden Fragestellung wird eingegangen. (DIN, S. 16)	++	1.20; 14.6
V29	Verzweigungsfrage: Gibt es Hinweise darauf, dass das Verfahren gruppenspezifische (z. B. geschlechtsspezifische) Ergebnisse liefert? Falls „nein“ → Bitte weiter bei Verzweigungsfrage V30.		
100	Für jede der Gruppen, für die das Verfahren gruppenspezifische (z. B. geschlechtsspezifische) Ergebnisse liefert, wurde empirisch untersucht, wie sich die Gruppenspezifität der Verfahrensergebnisse auf mögliche Entscheidungen auswirkt. (DIN, S. 16)	+	7.1
V30	Verzweigungsfrage: Wird in den Verfahrenshinweisen die Fairness-Frage thematisiert und/oder analysiert? Falls „nein“ → Bitte weiter bei Verzweigungsfrage V31.		
101	Die der Argumentation und/oder den empirischen Analysen zugrunde gelegte Fairness-Auffassung wird expliziert. (DIN, S. 16)		
102	Das der Argumentation und/oder den empirischen Analysen zugrunde gelegte Fairness-Modell wird begründet. (DIN, S. 16)		

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zu Anforderungen an Verfahrenshinweise (Checkliste 1)			
	DIN 33430, Aussagen zur Gültigkeit (Validität)	Korrespondenz	Standards
V31	Verzweigungsfrage: Gibt es Gruppen, bei denen das Verfahren nicht als Entscheidungsgrundlage genutzt werden darf? Falls „nein“ → Bitte weiter bei Verzweigungsfrage V32.		
103	Es ist angegeben, bei welchen Gruppen das Verfahren nicht als Entscheidungsgrundlage genutzt werden darf. (DIN, S. 16)	+	7.2
V32	Verzweigungsfrage: Sieht das Verfahren gruppenspezifische Normierungen vor (z. B. Alters-, Geschlechts- und Bildungsnormen)? Falls „nein“ → Bitte weiter bei Aussage 106.		
104	Die gruppenspezifischen Normierungen werden begründet. (DIN, S. 16)		
105	Die gruppenspezifischen Normierungen sind in ihren Effekten beschrieben. (DIN, S. 16)		
	DIN 33430, Aussagen zur Konstruktgültigkeit (Konstruktvalidität)	Korrespondenz	Standards
106	Jedes für das Verfahren relevante Konstrukt ist von anderen Konstrukten klar abgrenzbar. (DIN, S. 17)		
107	Jedes für das Verfahren relevante Konstrukt ist in einen theoretischen Rahmen eingebettet. (DIN, S. 17)		
108	Jedes für das Verfahren relevante Konstrukt ist in den Verfahrenshinweisen so beschrieben, dass es ohne Sekundärliteratur verständlich ist. (DIN, S. 17)	+	1.2
109	Die auf das Konstrukt/die Konstrukte bezogenen empirisch-psychologischen Forschungsergebnisse sind in den Verfahrenshinweisen so beschrieben, dass sie ohne Sekundärliteratur verständlich sind. (DIN, S. 17)		
V33	Verzweigungsfrage: Gibt es verfahrensrelevante theoretische Alternativen? Falls „nein“ → Bitte weiter bei Verzweigungsfrage V34.		
110	Die verfahrensrelevanten theoretischen Alternativen sind dargestellt. (DIN, S. 17)		

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zu Anforderungen an Verfahrenshinweise (Checkliste 1)			
	DIN 33430, Aussagen zur Konstruktgültigkeit (Konstruktvalidität)	Korrespondenz	Standards
V34	Verzweigungsfrage: Gibt es empirische Ergebnisse, die den zugrunde gelegten Annahmen widersprechen? Falls „nein“ → Bitte weiter bei Aussage 113.		
111	Empirische Ergebnisse, die den zugrunde gelegten Annahmen widersprechen, sind dargestellt. (DIN, S. 17)		
112	Trotz der Darstellung von empirischen Ergebnissen, die den zugrunde gelegten Annahmen widersprechen, bleiben das Konstrukt und die diesbezüglichen empirisch-psychologischen Forschungsergebnisse ohne Sekundärliteratur verstehbar. (DIN, S. 17)		
113	Aufgrund von inhaltlichen Überlegungen wird dargelegt, wie sich das fragliche Konstrukt zu ähnlichen Konstrukten verhält (konvergente Gültigkeit). (DIN, S. 17)	+	1.14
114	Aufgrund von empirischen Ergebnissen wird dargelegt, wie sich das fragliche Konstrukt zu ähnlichen Konstrukten verhält (konvergente Gültigkeit). (DIN, S. 17)	+	1.14
115	Aufgrund von inhaltlichen Überlegungen wird dargelegt, wie sich das fragliche Konstrukt zu unähnlichen Konstrukten verhält (diskriminante Gültigkeit). (DIN, S. 17)	+	1.14
116	Aufgrund von empirischen Ergebnissen wird dargelegt, wie sich das fragliche Konstrukt zu unähnlichen Konstrukten verhält (diskriminante Gültigkeit). (DIN, S. 17)	+	1.14
	DIN 33430, Aussagen zur Kriteriumsgültigkeit (Kriteriumsvalidität)	Korrespondenz	Standards
117	Bei der Analyse der Kriteriumsgültigkeit des Verfahrens wird beschrieben, warum das/die in der Analyse verwendete(n) Kriterium/Kriterien angemessen ist/sind. (DIN, S. 17)	+	1.16; 14.4
118	Bei der Analyse der Kriteriumsgültigkeit des Verfahrens wird beschrieben, warum die in der Analyse verwendeten Operationalisierung(en) des Kriteriums/der Kriterien angemessen ist/sind. (DIN, S. 17)	+	1.16; 14.4

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zu Anforderungen an Verfahrenshinweise (Checkliste 1)			
	DIN 33430, Aussagen zur Kriteriumsgültigkeit (Kriteriumsvalidität)	Korrespondenz	Standards
119	Die inhaltliche Qualität jedes verwendeten Kriterienmaßes ist ausführlich dargestellt. (DIN, S. 17)	+	1.16; 14.4
120	Die technische Qualität jedes verwendeten Kriterienmaßes ist ausführlich dargestellt. (DIN, S. 17)	+	1.16
121	Die Angemessenheit des Designs (retrograd und/oder konkurrenz und/oder prädiktiv) der Analyse der Kriteriumsgültigkeit wird erläutert. (DIN, S. 17)		
122	Die Angemessenheit der für die Analyse der Kriteriumsgültigkeit herangezogenen Untersuchungsgruppe wird erläutert. (DIN, S. 17)		
V35	<i>Verzweigungsfrage: Wird die Kriteriumsgültigkeit anhand einer Studie aufgezeigt, in der aus dem Verfahren Beurteilungen abgeleitet werden, die Grundlage für Auswahl- und/oder Klassifikationsentscheidungen sind? Falls „nein“ → Bitte weiter bei Verzweigungsfrage V36.</i>		
123	Die Kriteriumsgültigkeit wird auch unter entscheidungstheoretischen Gesichtspunkten diskutiert. (DIN, S. 17)	+	12.17
V36	<i>Verzweigungsfrage: Liegen Kriteriumsgültigkeiten vergleichbarer Verfahren für gleiche oder ähnliche Anwendungsbereiche vor? Falls „nein“ → Bitte weiter bei Verzweigungsfrage V37.</i>		
124	In den Verfahrenshinweisen werden auch Kriteriumsgültigkeiten vergleichbarer Verfahren für gleiche oder ähnliche Anwendungsbereiche berichtet. (DIN, S. 17)		
	DIN 33430, Aussagen zur Inhaltsgültigkeit (Inhaltsvalidität)	Korrespondenz	Standards
V37	<i>Verzweigungsfrage: Wird für das Verfahren Inhaltsgültigkeit in Anspruch genommen? Falls „nein“ → Ende der normativen Aussagen der vorliegenden Checkliste, bitte weiter bei der nicht normativen Aussage 133.</i>		
125	Der im Verfahren abgebildete Inhaltsbereich ist beschrieben. (DIN, S. 17)	+	1.6; 3.2; 3.3; 3.11; 14.8

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zu Anforderungen an Verfahrenshinweise (Checkliste 1)			
	DIN 33430, Aussagen zur Inhaltsgültigkeit (Inhaltsvalidität)	Korrespondenz	Standards
126	Die Bedeutung des im Verfahren abgebildeten Inhaltsbereiches für die vorgesehene Beurteilung ist beschrieben. (DIN, S. 17)	+	1.6; 3.2; 3.11; 14.8
127	Aus der Beschreibung des Inhaltsbereiches <u>und</u> seiner Bedeutung für die vorgesehene Beurteilung geht hervor, dass die den Inhaltsbereich definierenden Merkmale wesentliche Determinanten für das in Frage stehende erwünschte Zielverhalten der untersuchten Personen darstellen. (DIN, S. 17)	+	1.6; 3.11; 14.8
128	Die Regeln für die Erzeugung des dem Verfahren zugrunde liegenden Itemuniversums sind angegeben. (DIN, S. 17)	++	1.6
129	Die Regeln, nach denen das Verfahren als systematisch zusammengestellte Itemstichprobe aus dem Itemuniversum abgeleitet wurde, sind dargestellt. (DIN, S. 17)	++	1.6
V38	<i>Verzweigungsfrage: Wurde von Experten beurteilt, ob das Verfahren den definierten Inhaltsbereich repräsentiert? Falls „nein“ → Ende der normativen Aussagen der vorliegenden Checkliste, bitte weiter bei der nicht normativen Aussage 133.</i>		
130	Der fachbezogene Ausbildungsstand <u>und</u> die Erfahrung <u>und</u> die Qualifikation der genutzten Experten sind beschrieben. (DIN, S. 17)	+ ¹	1.7
131	Es ist erläutert, wie die Experten zu einer Einschätzung gekommen sind. (DIN, S. 17)	+ ¹	1.7
132	Es ist angegeben, inwieweit die Expertenbeurteilungen übereinstimmen. (DIN, S. 17)	+ ¹	1.7
	DIN 33430, Aussagen zu Studien zur Zuverlässigkeit (Reliabilität). Nicht normative Aussagen	Korrespondenz	Standards
133	Die Untersuchung(en), mit denen die Zuverlässigkeit des Verfahrens bestimmt wurde, sind aktuell. (DIN, S. 6).		

¹ Der Text der DIN-Aussagen bezieht sich nur auf die Kontextvalidität, der Text des Standards 1.7 bezieht sich allgemein auf Validität.

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zu Anforderungen an Verfahrenshinweise (Checkliste 1)			
	<i>DIN 33430, Aussagen zu Studien zur Zuverlässigkeit (Reliabilität). Nicht normative Aussagen</i>	Korrespondenz	Standards
	Die Teilnehmergruppe der Untersuchung(en), mit denen die Zuverlässigkeit des Verfahrens bestimmt wurde ...		
134	– zeichnet sich durch einen angemessenen Umfang aus (DIN, S. 6)		
135	– ist repräsentativ für die Zielgruppe. (DIN, S. 6)		
136	Es liegen mehrere unabhängige Vergleichs- und/oder Wiederholungsuntersuchungen vor, mit denen die Zuverlässigkeit des Verfahrens bestimmt wurde. (DIN, S. 6)		
	<i>DIN 33430, Aussagen zu Studien zur Gültigkeit (Validität) – Nicht normative Aussagen</i>	Korrespondenz	Standards
137	Die Untersuchung(en), mit denen die Gültigkeit des Verfahrens bestimmt wurde(n), ist/sind aktuell. (DIN, S. 5).		
	Die Teilnehmergruppe der Untersuchung(en), mit denen die Gültigkeit des Verfahrens bestimmt wurde ...		
138	– zeichnet sich durch einen angemessenen Umfang aus (DIN, S. 5)	+	1.13
139	– ist repräsentativ für die Zielgruppe. (DIN, S. 5)	+	1.13
140	Der Ansatz der Untersuchung(en), mit denen die Gültigkeit des Verfahrens bestimmt wurde, ist für das zu messende Merkmal angemessen. (DIN, S. 5)	+	1.13
DIN 33430, Aussagen zur Planung von berufsbezogenen Eignungsbeurteilungen (Checkliste 2)			
	<i>DIN 33430, Aussagen zur Auftragsklärung</i>	Korrespondenz	Standards
141	Dem Auftragnehmer wurde ein Auftrag zur Eignungsbeurteilung erteilt und im Rahmen des erteilten Auftrags wurden dem Auftragnehmer Fragen zur Beantwortung aufgegeben. (DIN, S. 9)		

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zur Planung von berufsbezogenen Eignungsbeurteilungen (Checkliste 2)			
	<i>DIN 33430, Aussagen zur Auftragsklärung</i>	Korrespondenz	Standards
142	Der Auftragnehmer hat dem Auftraggeber seine Qualifikation (im Sinne der Aussagen 271 bis 294 der Checkliste 6) <u>und</u> seine regelmäßige fachliche Fortbildung nachgewiesen.	+	12.1
	<i>DIN 33430, Aussagen zu Verantwortlichkeiten</i>	Korrespondenz	Standards
143	Der Auftragnehmer übernimmt die Hauptverantwortung für die Planung <u>und</u> Durchführung der gesamten Eignungsbeurteilung sowie für die Auswertung <u>und</u> Interpretation der Ergebnisse <u>und</u> für den Bericht an den Auftraggeber. (DIN, S. 9)		
V39	<i>Verzweigungsfrage: Wird ein Teil der Verantwortung vom Auftraggeber auf Mitwirkende übertragen? Falls „nein“ → Bitte weiter bei Aussage 151.</i>		
	Die folgenden Verantwortungen und/oder Aufgaben wurden nicht delegiert, sondern vom Auftragnehmer persönlich wahrgenommen:		
144	– die Auswahl <u>und</u> Zusammenstellung von Verfahren (DIN, S. 9)		
145	– die Planung der Untersuchungssituation (DIN, S. 9)		
146	– die Festlegung der Beurteilungsregeln (DIN, S. 9)		
147	– die Festlegung der Interpretationsregeln. (DIN, S. 9)		
148	Der Auftragnehmer nimmt seine Fachaufsicht über die Tätigkeit der Mitwirkenden aktiv <u>und</u> nachweislich wahr. (DIN, S. 9)	+	11.3
149	Der Auftragnehmer stellt sicher, dass alle Mitwirkenden in ihren jeweiligen Aufgabenbereich entsprechend eingewiesen <u>und</u> für ihre Aufgaben spezifisch geschult sind. (DIN, S. 9)	+	11.3; 12.1; 13.10
150	Der Auftragnehmer gewährleistet, dass die Mitwirkenden so qualifiziert sind, wie es zur Durchführung des Verfahrens erforderlich ist. (DIN, S. 9)	+	11.3; 13.10

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zur Planung von berufsbezogenen Eignungsbeurteilungen (Checkliste 2)			
	<i>DIN 33430, Aussagen zur Anforderungsanalyse I</i>	<i>Korrespondenz</i>	<i>Standards</i>
151	Es wurde eine Anforderungsanalyse durchgeführt. (DIN, S. 6)		
152	Jeder Anforderungsdimension, die erhoben werden soll, kann mindestens ein Verfahren zugeordnet werden, mit dem diese Dimension erfasst wird. (DIN, S. 6)		
153	Jedem der eingesetzten Verfahren kann mindestens eine Anforderungsdimension zugeordnet werden, zu deren Erfassung es eingesetzt wird. (DIN, S. 6)		
	<i>DIN 33430, Aussagen zur Berücksichtigung von Besonderheiten der Eignungsbeurteilung</i>	<i>Korrespondenz</i>	<i>Standards</i>
	Es wurde eine Schätzung vorgenommen über ... (DIN, S. 6)		
154	– das Angebot an qualifizierten Bewerbern (DIN, S. 6)		
155	– den Bedarf an qualifizierten Bewerbern (DIN, S. 6)		
156	– den Anteil der geeigneten Personen in der Gruppe der Bewerber (DIN, S. 6)		
157	– den Nutzen, der entsteht, wenn es mithilfe des Verfahrens besser gelingt, leistungsstärkere Personen auszuwählen (DIN, S. 6)		
158	– die direkten und indirekten Verfahrenskosten. (DIN, S. 6)		
	<i>DIN 33430, Aussagen zur Festlegung der Regeln zur Durchführung, Auswertung und Interpretation der Verfahrensergebnisse sowie Urteilsbildung und Ergebnismrückmeldung</i>	<i>Korrespondenz</i>	<i>Standards</i>
159	Es wurden im Vorhinein Regeln aufgestellt, die alle Aspekte der Durchführung der Verfahren festlegen. (DIN, S. 8)	+	3.19
160	Es wurden im Vorhinein Regeln aufgestellt, die alle Aspekte der Auswertung der Verfahren festlegen. (DIN, S. 8)	+	3.14; 3.22
161	Es wurden im Vorhinein Regeln aufgestellt, wie die über einen Kandidaten erhobenen Informationen zu einem Eignungsurteil zusammengefügt werden. (DIN, S. 8)		

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zur Planung von berufsbezogenen Eignungsbeurteilungen (Checkliste 2)			
	<i>DIN 33430, Aussagen zur Festlegung der Regeln zur Durchführung, Auswertung und Interpretation der Verfahrensergebnisse sowie Urteilsbildung und Ergebnismrückmeldung</i>	<i>Korrespondenz</i>	<i>Standards</i>
V40	<i>Verzweigungsfrage: Werden Verfahren eingesetzt, die auf mündlich gewonnenen Informationen bzw. Verhaltensbeobachtungen basieren (z. B. Interviews, Arbeitsproben, Assessment Center)? Falls „nein“ → Bitte weiter bei Aussage 163.</i>		
162	Den Beurteilungskategorien wurden vorab Beispielaussagen und/oder Beispielverhaltensweisen zugeordnet (DIN, S. 8).		
163	Es wurde vorab vereinbart, wie den Kandidaten die Ergebnisse vermittelt werden. (DIN, S. 8)		
V41	<i>Verzweigungsfrage: Wird das Auswahlprogramm in gleicher Art und Weise wiederholt durchgeführt? Falls „nein“ → Bitte weiter bei Verzweigungsfrage V42.</i>		
164	Die Bewährung der Regeln zur Durchführung und Auswertung der Verfahren und der Regeln zur Zusammenführung der Informationen zu einem Eignungsurteil wurde innerhalb der letzten 3 Jahre überprüft. (DIN, S. 8)		
V42	<i>Verzweigungsfrage: Bearbeiten einzelne Kandidaten die Verfahren oder Teile der Verfahren in unterschiedlicher Reihenfolge? Falls „nein“ → Bitte weiter bei Aussage 166.</i>		
165	Es kann mit großer Wahrscheinlichkeit davon ausgegangen werden, dass trotz der unterschiedlichen Durchführungsabfolge der Verfahren alle Kandidaten die gleichen Chancen haben. (DIN, S. 8)		
	<i>DIN 33430, Aussagen zur Anforderungsanalyse II – Nicht normative Aussagen</i>	<i>Korrespondenz</i>	<i>Standards</i>
	Mit der Anforderungsanalyse wurden die Merkmale eines Arbeitsplatzes und/oder einer Ausbildung bzw. eines Studiums und/oder eines Berufs und/oder einer beruflichen Tätigkeit ermittelt, die ...		
166	– für den beruflichen Erfolg bedeutsam sind (DIN, S. 12)		
167	– für die berufliche Zufriedenheit bedeutsam sind. (DIN, S. 12)		

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zur Planung von berufsbezogenen Eignungsbeurteilungen (Checkliste 2)			
	<i>DIN 33430, Aussagen zur Anforderungsanalyse II – Nicht normative Aussagen</i>	<i>Korrespondenz</i>	<i>Standards</i>
168	Aus der Anforderungsanalyse wurden diejenigen Eignungsmerkmale abgeleitet, die zur Erfüllung der Anforderungen nötig sind. (DIN, S. 12)		
169	Aus der Anforderungsanalyse wurden diejenigen Eignungsmerkmale mitsamt ihren Ausprägungsgraden abgeleitet, die zur Erfüllung der Anforderungen nötig sind. (DIN, S. 12)		
170	Es wird verdeutlicht, aufgrund welcher Überlegungen die Eignungsmerkmale (und ihre Ausprägungsgrade) aus der Anforderungsanalyse abgeleitet wurden. (DIN, S. 12)		
V43	<i>Verzweigungsfrage: Wurde auf vorhandene Tätigkeits-, Stellen-, Aufgaben- oder Funktions-Beschreibungen zurückgegriffen? Falls „nein“ → Bitte weiter bei Verzweigungsfrage V44.</i>		
171	Es wurde sichergestellt, dass sich seit der Erstellung der für die Anforderungsanalyse genutzten Tätigkeits-, Stellen-, Aufgaben- oder Funktions-Beschreibungen keine bedeutsamen Veränderungen der Anforderungen mehr ergeben haben. (DIN, S. 12)		
V44	<i>Verzweigungsfrage: Wird die Eignungsbeurteilung (auch) zum Zweck der Berufsberatung vorgenommen? Falls „nein“ → Bitte weiter bei Aussage 173.</i>		
172	Es werden auch die Ergebnisse von Berufsanalysen berücksichtigt. (DIN, S. 12)		
	<i>DIN 33430, Aussagen zu Planungen zur Verfahrensdurchführung. Nicht normative Aussagen.</i>	<i>Korrespondenz</i>	<i>Standards</i>
173	Pausen und Wartezeiten und deren Mindestdauer werden vorab festgelegt. (DIN, S. 13)		
DIN 33430, Aussagen zur Auswahl und Zusammenstellung von Verfahren (Checkliste 3)			
	<i>DIN 33430, Aussagen zu grundsätzlichen Gesichtspunkten der Verfahrensauswahl</i>	<i>Korrespondenz</i>	<i>Standards</i>
174	Es werden nur Verfahren eingesetzt, für die Verfahrenshinweise (Handhabungshinweise) vorliegen. (DIN, S. 6)		

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zur Auswahl und Zusammenstellung von Verfahren (Checkliste 3)			
	<i>DIN 33430, Aussagen zu grundsätzlichen Gesichtspunkten der Verfahrensauswahl</i>	<i>Korrespondenz</i>	<i>Standards</i>
175	Die Verfahrenshinweise zu dem ausgewählten Verfahren entsprechen den Anforderungen an Verfahrenshinweise (laut Checkliste 1). (DIN, S. 6)		
	<i>DIN 33430, Aussagen zum Auswahlgesichtspunkt Objektivität</i>	<i>Korrespondenz</i>	<i>Standards</i>
176	Das Verfahren besitzt eine größtmögliche Durchführungsobjektivität. (DIN, S. 6)		
177	Das Verfahren besitzt eine größtmögliche Auswertungsobjektivität. (DIN, S. 6)		
178	Das Verfahren besitzt eine größtmögliche Interpretationsobjektivität. (DIN, S. 6)		
179	Das Material des Verfahrens ist so gestaltet, dass die Ergebnisse so wenig wie möglich durch den Kandidaten selbst verfälscht werden können. (DIN, S. 6/7)	+	5.6; 15.9
180	Die Instruktionen des Verfahrens sind so gestaltet, dass die Ergebnisse so wenig wie möglich durch den Kandidaten selbst verfälscht werden können. (DIN, S. 6/7)	+	5.6; 15.9
181	Die Vorgehensweise bei der Beurteilung ist so gestaltet, dass die Ergebnisse so wenig wie möglich durch den Kandidaten selbst verfälscht werden können. (DIN, S. 6/7)	+	5.6; 15.9
	<i>DIN 33430, Aussagen zum Auswahlgesichtspunkt Zuverlässigkeit</i>	<i>Korrespondenz</i>	<i>Standards</i>
182	Bei der Auswahl der Verfahren wurde deren Zuverlässigkeit berücksichtigt. (DIN, S. 7)	+	2.1; 12.13
183	Das Verfahren besitzt eine der jeweiligen Art des Verfahrens <u>und</u> der angestrebten Aussage entsprechende, möglichst hohe Zuverlässigkeit. (DIN, S. 7)		

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zur Auswahl und Zusammenstellung von Verfahren (Checkliste 3)			
	DIN 33430, Aussagen zum Auswahlgesichtspunkt Zuverlässigkeit	Korrespondenz	Standards
V45	Verzweigungsfrage: Beurteilen mehrere Personen den Ausprägungsgrad von Personmerkmalen aufgrund von mündlich gewonnenen Informationen bzw. Verhaltensbeobachtungen (z. B. Teamfähigkeit aufgrund eines Interviews oder einer Assessment-Center-Übung)? Falls „nein“ → Bitte weiter bei Aussage 185.		
184	Es wurden konkrete Maßnahmen ergriffen, um sicherzustellen, dass Interviewer bzw. Beobachter in ihren Ergebnissen übereinstimmen. (DIN, S. 7)		
DIN 33430, Aussagen zum Auswahlgesichtspunkt Gültigkeit			
	DIN 33430, Aussagen zum Auswahlgesichtspunkt Gültigkeit	Korrespondenz	Standards
185	Bei der Auswahl der Verfahren wurde nachweislich (Dokumentation liegt vor) deren Gültigkeit berücksichtigt. (DIN, S. 7)	+	11.4; 12.13
186	Die Gültigkeit des Verfahrens wurde aufgrund von empirischen Analysen zur Konstrukt- und/oder Kriteriums- und/oder Inhaltsvalidität nachgewiesen. (DIN, S. 7)		
187	Die Ergebnisse von unabhängigen Vergleichs- und/oder Wiederholungsuntersuchungen zur Gültigkeit wurden berücksichtigt. (DIN, S. 7)		
188	Das Verfahren besitzt eine für den intendierten Anwendungsbereich möglichst hohe Gültigkeit. (DIN, S. 7)		
V46	Verzweigungsfrage: Wird das Verfahren vom Auftragnehmer wiederholt bei gleichartigen Fragestellungen eingesetzt? Falls „nein“ → Bitte weiter bei Verzweigungsfrage V47.		
189	Der Auftragnehmer hat die Gültigkeit des bei gleichartigen Fragestellungen eingesetzten Verfahrens nachweislich ermittelt. (DIN, S. 7)		
DIN 33430, Aussagen zum Auswahlgesichtspunkt Normwert/Referenzkennwerte			
	DIN 33430, Aussagen zum Auswahlgesichtspunkt Normwert/Referenzkennwerte	Korrespondenz	Standards
V47	Verzweigungsfrage: Wird ein Verfahren eingesetzt, das einen Vergleich mit Normwerten vorsieht? Falls „nein“ → Ende der Checkliste 3.		

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zur Auswahl und Zusammenstellung von Verfahren (Checkliste 3)			
	DIN 33430, Aussagen zum Auswahlgesichtspunkt Normwert/Referenzkennwerte	Korrespondenz	Standards
190	Die herangezogenen Normwerte entsprechen der Fragestellung und der Referenzgruppe der Zielgruppe. (DIN, S. 7)	+	12.15
191	Die Angemessenheit der Normwerte wurde in den letzten 8 Jahren überprüft. (DIN, S. 7)	+	4.18
V48	Verzweigungsfrage: Zielt der Verfahrenseinsatz auf die Erfassung eines Merkmals, dessen Ausprägung in der Referenzgruppe möglicherweise relativ kurzfristigen Veränderungen unterliegt? Falls „nein“ → Ende der Checkliste 3.		
V49	Verzweigungsfrage: Ist das Verfahren mindestens 8 Jahre alt? Falls „nein“ → Ende der Checkliste 3.		
192	Noch vor Ablauf der 8-Jahres-Frist wurde eine Neunormierung vorgenommen. (DIN, S. 7)	+	4.18
V50	Verzweigungsfrage: Gibt es unterschiedliche „Versionen“ des Verfahrens, also z. B. eine Papier-Bleistift-Version und eine Computerversion? Falls „nein“ → Ende der Checkliste 3.		
V51	Verzweigungsfrage: Gibt es gesonderte Normen pro Version? Falls „nein“ → Ende der Checkliste 3.		
193	Die Übertragbarkeit der Normwerte von einer Vorgabeart auf die andere wurde nachgewiesen. (DIN, S. 7)	+	4.10; 6.11
DIN 33430, Aussagen zur Durchführung, Auswertung und Interpretation der Verfahrensergebnisse sowie Urteilsbildung (Checkliste 4)			
	DIN 33430, Aussagen zur Durchführung I	Korrespondenz	Standards
	Bei der Verfahrensdurchführung werden die in den Verfahrenshinweisen enthaltenen Vorgaben und/oder Empfehlungen zu den nachfolgend genannten Aspekten eingehalten:		
194	– Vorbereitung (DIN, S. 8)	+	5.1; 12.12
195	– Material und dessen Einsatz (DIN, S. 8)	+	5.1; 12.12

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zur Durchführung, Auswertung und Interpretation der Verfahrensergebnisse sowie Urteilsbildung (Checkliste 4)			
	DIN 33430, Aussagen zur Durchführung I	Korrespondenz	Standards
196	– Mündliche Aufgabeninstruktionen (DIN, S. 8)	+	5.1; 12.12
197	– Vorgeschriebene Protokollierungen (DIN, S. 8)	+	5.1; 12.12
198	– Vorgeschriebene Zeiten (DIN, S. 8)	+	5.1; 12.12
199	– Umgang mit Nachfragen. (DIN, S. 8)	+	5.1; 12.12
200	Es werden die Originalmaterialien verwendet. (DIN, S. 8)	+	11.8
201	Die Anweisungen bzw. Erläuterungen an die Kandidaten sind verständlich und eindeutig. (DIN, S. 8)		
202	Die Anweisungen bzw. Erläuterungen an die Kandidaten sind so weit wie möglich standardisiert. (DIN, S. 8)	+	3.19
203	Es wird – so weit wie möglich – dafür gesorgt, dass niemand durch Betrug und/oder Täuschung verfälschte Verfahrensergebnisse erzielen kann. (DIN, S. 8)	+	5.6; 15.9
204	Durch die Art der Durchführung des Verfahrens bzw. durch den Umgang mit den dazugehörigen detaillierten Informationen ist sichergestellt, dass die Kandidaten in gleichem Maße Zugang oder keinen Zugang zu Vorinformationen, Kenntnis einzelner Items u. a. haben (Verfahrensschutz). (DIN, S. 14).	+	5.7; 8.1; 11.7; 12.11
V52	<i>Verzweigungsfrage: Werden computergestützte Verfahren eingesetzt? Falls „nein“ → Bitte weiter bei Aussage 206.</i>		
V53	<i>Verzweigungsfrage: Werden in den Verfahrenshinweisen technische Vorgaben (z. B. Hardware wie Bildschirm, Hilfsmittel) genannt? Falls „nein“ → Bitte weiter bei Aussage 206.</i>		
205	Die in den Verfahrenshinweisen genannten technischen Vorgaben (z. B. Hardware wie Bildschirm, Hilfsmittel) werden eingehalten. (DIN, S. 8)	+	12.12

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zur Durchführung, Auswertung und Interpretation der Verfahrensergebnisse sowie Urteilsbildung (Checkliste 4)			
	DIN 33430, Aussagen zur zur Auswertung, Interpretation und Urteilsbildung I	Korrespondenz	Standards
206	Die Auswertung folgt den vorab festgelegten Vorschriften. (DIN, S. 8)	+	12.11
207	Es ist sichergestellt, dass in dem Fall, in dem aufgrund von Störungen oder Verfälschungen von den Verfahrenshinweisen und/oder der geplanten Durchführung abgewichen wird, diese Abweichung bei der Auswertung berücksichtigt wird. (DIN, S. 8)		
208	Es werden nur Informationen zu anforderungsrelevanten Eignungsmerkmalen ausgewertet. (DIN, S. 8)		
209	Die Bewertung richtet sich nach den Beurteilungskategorien und den diesen Beurteilungskategorien vorab zugeordneten Beispielaussagen und/oder Beispielverhaltensweisen. (DIN, S. 8).		
V54	<i>Verzweigungsfrage: Beurteilen mehrere Personen den Ausprägungsgrad von Personmerkmalen aufgrund von mündlich gewonnenen Informationen bzw. Verhaltensbeobachtungen (z. B. Teamfähigkeit aufgrund eines Interviews oder einer Assessment-Center-Übung)? Falls „nein“ → Bitte weiter bei Aussage 213.</i>		
210	Der Grad der Übereinstimmung zwischen den beurteilenden Personen wurde berechnet. (DIN, S. 7)		
211	Der Grad der Übereinstimmung zwischen den beurteilenden Personen ist zufriedenstellend. (DIN, S. 7)		
212	Aufgrund der Urteile mehrerer Beurteiler wird ein Gesamtergebnis erstellt, dabei wird die Streubreite der zugrundeliegenden Einzelurteile bestimmt. (DIN, S. 8)		
	DIN 33430, Aussagen zur Interpretation/Urteilsbildung	Korrespondenz	Standards
213	Die Regeln für die Interpretation der Verfahrensergebnisse und zur abschließenden Eignungsbeurteilung wurden vom verantwortlichen Auftragnehmer festgelegt. (DIN, S. 9)		

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zur Durchführung, Auswertung und Interpretation der Verfahrensergebnisse sowie Urteilsbildung (Checkliste 4)			
	<i>DIN 33430, Aussagen zur Interpretation/Urteilsbildung</i>	<i>Korrespondenz</i>	<i>Standards</i>
214	Die Ausprägung(en) des/der interessierenden Merkmals/Merkmale eines Kandidaten wird/werden mit Blick auf die Referenzgruppe bewertet. (DIN, S. 9)	+	3.4
215	Die Interpretation und Eignungsbeurteilung richtet sich nach den Grundsätzen der Objektivität und der Unparteilichkeit und Unabhängigkeit in Bezug auf die Kandidaten. (DIN, S. 9)	+	7.12
V55	<i>Verzweigungsfrage: Wurden mehrere Informationen über einen Kandidaten erhoben? Falls „nein“ → Bitte weiter bei Verzweigungsfrage V56.</i>		
216	Es wird festgestellt, ob und falls ja: welche Ergebnisse zu gleichsinnigen oder sich widersprechenden Interpretationen führen. (DIN, S. 9)		
V56	<i>Verzweigungsfrage: Wird ein standardisiertes Verfahren eingesetzt? Falls „nein“ → Bitte weiter bei Aussage 218.</i>		
V57	<i>Verzweigungsfrage: Werden Interpretationen von Subtestwerten und/oder Messwertdifferenzen und/oder Profilen und/oder Reaktionen auf Itemebene und/oder einzelne Teile von Verfahren (z. B. Subtests u. Ä.) vorgenommen? Falls „nein“ → Bitte weiter bei Aussage 218.</i>		
217	Für alle Werte, die interpretiert werden (z. B. Subtestwerte und/oder Messwertdifferenzen und/oder Profilen und/oder Reaktionen auf Itemebene und/oder einzelne Teile von Verfahren), ist die Gültigkeit gesondert nachgewiesen. (DIN, S. 9)	+	1.1; 1.10; 1.12
218	Es ist sichergestellt, dass in dem Fall, in dem aus bestimmten Erwägungen heraus von den Verfahrenshinweisen und/oder den Durchführungsvorgaben abgewichen wird, diese Abweichung bei der Interpretation berücksichtigt wird. (DIN, S. 8)		
219	Die Eignungsbeurteilung gibt Antwort auf alle im erteilten Auftrag gestellten Fragen. (DIN, S. 9).		

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zur Durchführung, Auswertung und Interpretation der Verfahrensergebnisse sowie Urteilsbildung (Checkliste 4)			
	<i>DIN 33430, Aussagen zur Interpretation/Urteilsbildung</i>	<i>Korrespondenz</i>	<i>Standards</i>
220	Bei der Darstellung der Eignungsbeurteilung wird darauf eingegangen, auf welche Verfahrensergebnisse sich die Eignungsbeurteilung stützt. (DIN, S. 9)		
221	Alle Aussagen der Eignungsbeurteilung sind belegt. (DIN, S. 9)		
222	Die Darstellung der Eignungsbeurteilung ist sprachlich für den Auftraggeber verständlich. (DIN, S. 9)		
V58	<i>Verzweigungsfrage: Wurde ein computergestütztes Verfahren mit automatischer Klassifikation und/oder Textbausteinen für die abschließende Eignungsbeurteilung/Befundinterpretation genutzt? Falls „nein“ → Bitte weiter bei Verzweigungsfrage V59.</i>		
223	Der Auftragnehmer übernimmt die Verantwortung für die Richtigkeit des übermittelten Befundes. (DIN, S. 9)	+	11.21; 12.15
224	Die Kandidaten werden darauf hingewiesen, dass der Befund automatisiert erstellt wurde. (DIN, S. 9)		
	<i>DIN 33430, Aussagen Vorauswahl. Nicht normative Aussagen.</i>	<i>Korrespondenz</i>	<i>Standards</i>
V59	<i>Verzweigungsfrage: Wird eine Vorauswahl vorgenommen? Falls „nein“ → Bitte weiter bei Aussage 229.</i>		
225	Die Vorgehensweise der Vorauswahl wurde vorab festgelegt. (DIN, S. 12)		
226	Die für die Vorauswahl definierten Auswahlkriterien wurden vorab festgelegt. (DIN, S. 12)		
227	Die Vorauswahlkriterien wurden aus dem Anforderungsprofil abgeleitet. (DIN, S. 12)		
228	Die Vorauswahl wurde von mindestens zwei unabhängigen Beurteilern vorgenommen. (DIN, S. 12)		

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zur Durchführung, Auswertung und Interpretation der Verfahrensergebnisse sowie Urteilsbildung (Checkliste 4)			
	<i>DIN 33430, Aussagen zur Durchführung II. Nicht normative Aussagen.</i>	<i>Korrespondenz</i>	<i>Standards</i>
229	Den Kandidaten werden vor und/oder während deren Teilnahme am Verfahren Informationen über den Arbeitsplatz und die Aufgaben angeboten, für die sie sich beworben haben. (DIN, S. 12)		
230	Alle gesetzlichen Vorgaben für Eignungsbeurteilungen werden eingehalten. (DIN, S. 12)		
231	Alle Beteiligten sind informiert, dass die Teilnahme der Kandidaten an dem Verfahren prinzipiell freiwillig ist. (DIN, S. 12)	+	8.4
232	Das Verfahren beginnt erst, nachdem die Kandidaten ihre freiwillige Teilnahme ausdrücklich erklärt haben. (DIN, S. 12)	+	8.4
V60	<i>Verzweigungsfrage: Werden zusätzlich zu dem Verfahren zur Eignungsbeurteilung weitere Verfahren zum Zweck der Überarbeitung bzw. Neu- und/oder Weiterentwicklung eingesetzt, die nicht für den Beurteilungszweck herangezogen werden? Falls „nein“ → Bitte weiter bei Aussage 235.</i>		
233	Die Kandidaten werden darauf hingewiesen, dass die Untersuchungssituation auch der Überarbeitung bzw. Neu- und/oder Weiterentwicklung eines Verfahrens dient. (DIN, S. 13)	+	3.20
234	Das Verfahren beginnt erst, nachdem die Kandidaten ihre freiwillige Teilnahme an Zusatzverfahren, die zum Zweck der Überarbeitung bzw. Neu- und/oder Weiterentwicklung eingesetzt werden und die nicht für den Beurteilungszweck herangezogen werden, ausdrücklich erklärt haben. (DIN, S. 12)		
235	Die Kandidaten haben bereits bei der Einladung Hinweise zum Ablauf der Eignungsuntersuchung erhalten. (DIN, S. 13)	+	8.2; 12.10
236	Die Kandidaten haben bereits bei der Einladung Hinweise zur Freiwilligkeit der Teilnahme erhalten. (DIN, S. 13)		

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zur Durchführung, Auswertung und Interpretation der Verfahrensergebnisse sowie Urteilsbildung (Checkliste 4)			
	<i>DIN 33430, Aussagen zur Durchführung II. Nicht normative Aussagen.</i>	<i>Korrespondenz</i>	<i>Standards</i>
	Die Kandidaten werden zu Beginn der Untersuchung aufgeklärt ...		
237	– über die Ziele und Funktion der Untersuchung (DIN, S. 13)	+	8.2
238	– über den Ablauf und die Dauer der Untersuchung (DIN, S. 13)	+	8.2
239	– über die an der Untersuchung mitwirkenden Personen und deren Funktion im Verfahren (DIN, S. 13)	+	8.2
240	– über die Berufsausbildung und Qualifikation der an der Untersuchung mitwirkenden Personen (DIN, S. 13)	+	8.2
241	– über mögliche Folgen mangelnder Kooperation (DIN, S. 13)	+	8.7
242	– über die Art der zu erhebenden Daten und ihre Verwendung und über Personen, die von den Ergebnissen Kenntnis erhalten (DIN, S. 13)	+	8.2
243	– über den Ort und die Form und die Dauer der Aufbewahrung der erhobenen Daten. (DIN, S. 13)	+	8.2
244	Den Kandidaten wird ermöglicht, sich mit dem Auftragnehmer über die Untersuchung auszutauschen. (DIN, S. 13)		
245	Die Kandidaten werden zeitlich und psychisch und körperlich nicht mehr beansprucht, als für den Untersuchungszweck erforderlich ist. (DIN, S. 13)		
246	Pausen und Wartezeiten und deren Mindestdauer werden den Kandidaten vorab mitgeteilt. (DIN, S. 13)		
247	Die Untersuchungsbedingungen werden so gestaltet, dass es den Kandidaten möglich ist, ihr anforderungsbezogenes Potenzial zu zeigen. (DIN, S. 13)	+	5.4
248	Die Untersuchungsbedingungen werden so gestaltet, dass die Kandidaten daran gehindert werden, Verfahrensergebnisse durch Betrug und/oder Täuschung zu erzielen. (DIN, S. 13)	+	5.6; 15.9

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zur Dokumentation (Checkliste 5)			
	DIN 33430, Aussagen zur Dokumentation.	Korrespondenz	Standards
249	Es ist dokumentiert, nach welchen Gesichtspunkten in Bezug auf die Zuverlässigkeit die Verfahren zur Eignungsbeurteilung ausgewählt wurden. (DIN, S. 7)		
250	Es ist dokumentiert, nach welchen Gesichtspunkten in Bezug auf die Gültigkeit die Verfahren zur Eignungsbeurteilung ausgewählt wurden. (DIN, S. 7)		
251	Die Auswahl des Verfahrens ist so dargestellt, dass die Eignungsbeurteilung von dem Auftraggeber nachvollzogen werden kann. (DIN, S. 8)		
V61	<i>Verzweigungsfrage: Werden im Kontext der Eignungsbeurteilung mehrere Verfahren genutzt? Falls „nein“ → Bitte weiter bei Aussage 253.</i>		
252	Die Zusammenstellung der Verfahren ist so dargestellt, dass die Eignungsbeurteilung von dem Auftraggeber nachvollzogen werden kann. (DIN, S. 8)		
253	Die Regeln, mit denen alle Aspekte der Durchführung der Verfahren festgelegt sind, wurden dokumentiert. (DIN, S. 8)		
V62	<i>Verzweigungsfrage: Wurden mehrere Informationen über einen Kandidaten erhoben? Falls „nein“ → Bitte weiter bei Aussage 255.</i>		
254	Es ist dokumentiert, ob und falls ja: welche Ergebnisse zu gleichsinnigen oder sich widersprechenden Interpretationen führen. (DIN, S. 9)		
	Das Vorgehen bei der Eignungsbeurteilung ist so dargestellt, dass die Eignungsbeurteilung von dem Auftraggeber nachvollzogen werden kann. Hierzu gehört die Dokumentation der nachfolgenden Aspekte:		
255	– Instruktionen (DIN, S. 8)		
256	– Verfahrenselemente (DIN, S. 8)		
257	– Einstufungshilfen/Regeln für die Ableitung der Eignungsbeurteilung. (DIN, S. 8)		

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zur Dokumentation (Checkliste 5)			
	DIN 33430, Aussagen zur Dokumentation.	Korrespondenz	Standards
258	Die Regeln, mit denen alle Aspekte der Auswertung der Verfahren festgelegt sind, wurden dokumentiert. (DIN, S. 8)		
259	Die Regeln, mit denen die Zusammenfügung aller über einen Kandidaten erhobenen Informationen zu einem Eignungsurteil festgelegt wurde, wurden dokumentiert. (DIN, S. 8)		
260	Es ist Vorsorge getroffen, dass in dem Fall, in dem aus bestimmten Erwägungen heraus von den Verfahrenshinweisen und/oder den Durchführungsvorgaben abgewichen wird, diese Abweichung dokumentiert wird. (DIN, S. 8)	+	5.2
V63	<i>Verzweigungsfrage: Gibt es Hinweise darauf, dass das Verfahren gruppenspezifische (z. B. geschlechtsspezifische) Ergebnisse liefert? Falls „nein“ → Bitte weiter bei Verzweigungsfrage V64.</i>		
261	Für jede der Gruppen, für die das Verfahren gruppenspezifische (z. B. geschlechtsspezifische) Ergebnisse liefert, wurde dokumentiert, wie sich die Gruppenspezifität der Verfahrensergebnisse auf mögliche Entscheidungen auswirkt. (DIN, S. 16)	+	7.1
V64	<i>Verzweigungsfrage: Werden Verfahren eingesetzt, die auf mündlich gewonnenen Informationen bzw. Verhaltensbeobachtungen basieren (z. B. Interviews, Arbeitsproben, Assessment Center)? Falls „nein“ → Bitte weiter bei Aussage V66.</i>		
262	Die Zuordnung der Beispielaussagen und/oder Beispielverhaltensweisen zu den Beurteilungskategorien wurde dokumentiert. (DIN, S. 8)		
V65	<i>Verzweigungsfrage: Beurteilen mehrere Personen den Ausprägungsgrad von Personmerkmalen aufgrund von mündlich gewonnenen Informationen bzw. Verhaltensbeobachtungen (z. B. Teamfähigkeit) aufgrund eines Interviews oder einer Assessment-Center-Übung? Falls „nein“ → Bitte weiter bei Verzweigungsfrage V66.</i>		
263	Der Grad der Übereinstimmung zwischen den beurteilenden Personen wurde dokumentiert. (DIN, S. 7)		
264	Das aufgrund der Urteile mehrerer Beurteiler erstellte Gesamtergebnis und die Streubreite sind dokumentiert. (DIN, S. 8)		

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zur Dokumentation (Checkliste 5)			
	DIN 33430, Aussagen zur Dokumentation.	Korrespondenz	Standards
V66	Verzweigungsfrage: Wird das Verfahren vom Auftragnehmer wiederholt bei gleichartigen Fragestellungen eingesetzt? Falls „nein“ → Bitte weiter bei Aussage 266.		
265	Der Auftragnehmer hat die Gültigkeit des bei gleichartigen Fragestellungen eingesetzten Verfahrens dokumentiert. (DIN, S. 7)		
266	Die Ergebnisse der Anforderungsanalyse sind dokumentiert. (DIN, S. 6)		
	DIN 33430, Aussagen zur Dokumentation. Nicht normative Aussagen.	Korrespondenz	Standards
	Die Anforderungsanalyse ist nachvollziehbar dokumentiert, indem die folgenden Aspekte beschrieben sind:		
267	– wer an der Anforderungsanalyse beteiligt war <u>und</u> über welche Qualifikationen diese Personen verfügen (DIN, S. 12)		
268	– welche Quellen für die Anforderungsanalyse genutzt wurden (DIN, S. 12)		
269	– welche Analyse- <u>und</u> Auswertungsverfahren genutzt wurden (DIN, S. 12)		
270	– welche Ergebnisse mit der Anforderungsanalyse erzielt wurden. (DIN, S. 12)		
DIN 33430, Aussagen zu Anforderungen an die Qualifikation der an der Eignungsbeurteilung beteiligten Personen (Checkliste 6)			
	DIN 33430, Aussagen zu Qualitätsanforderungen an den Auftragnehmer	Korrespondenz	Standards
	Der Auftragnehmer verfügt über		
271	– angeleitete Praxiserfahrungen in Entwicklung <u>und</u> Planung <u>und</u> Gestaltung <u>und</u> kontrollierter Durchführung von Verfahren zur Eignungsbeurteilung sowie deren Evaluation (DIN, S. 10)		
272	– Kenntnisse über eignungsrelevante Konstrukte (DIN, S. 10)		

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zu Anforderungen an die Qualifikation der an der Eignungsbeurteilung beteiligten Personen (Checkliste 6)			
	DIN 33430, Aussagen zu Qualitätsanforderungen an den Auftragnehmer	Korrespondenz	Standards
273	– Kenntnisse über Qualitätsstandards (DIN, S. 10)		
274	– Kenntnisse über qualitätssichernde Maßnahmen (DIN, S. 10)		
275	– Kenntnisse über rechtliche Rahmenbedingungen (DIN, S. 10)		
276	– Kenntnisse der Arbeits-/Anforderungsanalyse (DIN, S. 10)		
277	– Kenntnisse von Methoden zur Analyse von Arbeitsanforderungen (DIN, S. 10)		
278	– Kenntnisse von Verfahren zur Darstellung der Ergebnisse in Form eines Anforderungsprofils (DIN, S. 10)		
279	– Kenntnisse über Methoden zur Operationalisierung von Eignungsmerkmalen (DIN, S. 10)		
280	– Grundkenntnisse über Verfahren der Eignungsbeurteilung <u>und</u> ihre Einsatzvoraussetzungen (DIN, S. 10)		
281	– Statistisch-methodische Grundlagen (DIN, S. 10)		
282	– Kenntnisse über Testtheorien (Klassische Testtheorie <u>und</u> Item-Response-Theorien) <u>und</u> Messtheorien (DIN, S. 10)		
283	– Kenntnisse über Evaluationsmethodik einschließlich Kosten-Nutzen-Aspekten (DIN, S. 10)		
284	– Kenntnisse über Konstruktionsgrundlagen von Verfahren (DIN, S. 10)		
285	– Kenntnisse über Einsatzmöglichkeiten von Verfahren (DIN, S. 10)		
286	– Kenntnisse über Durchführungsbedingungen von Verfahren (DIN, S. 10)		
287	– Kenntnisse der Gütekriterien/Qualitätsgesichtspunkte von Verfahren (DIN, S. 10)		

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zu Anforderungen an die Qualifikation der an der Eignungsbeurteilung beteiligten Personen (Checkliste 6)			
	DIN 33430, Aussagen zu Qualitätsanforderungen an den Auftragnehmer	Korrespondenz	Standards
288	- Kenntnisse über die Gutachtenerstellung (abschließende Eignungsbeurteilung) (DIN, S. 10)		
289	- Kenntnisse der Vorgehensweisen/Prozesse in der Eignungsbeurteilung (DIN, S. 10)		
290	- Kenntnisse über verschiedene Strategien der Eignungsbeurteilung (DIN, S. 10)		
291	- Kenntnisse über Beurteilungsprozeduren (verfahrens- und prozessbezogen) (DIN, S. 10)		
292	- Kenntnisse über die Abschätzung der Prognosegüte von berufsbezogenen Eignungsbeurteilungen und darauf aufbauenden Entscheidungen (DIN, S. 11)		
293	- Kenntnisse der Ergebnisse einschlägiger Evaluationsstudien (DIN, S. 11)		
294	- Kenntnisse über Geltungsbereiche von Eignungsbeurteilungen. (DIN, S. 11)		
295	Der Auftragnehmer hat dem Auftraggeber die Qualifikation im Sinne der Aussagen 271 bis 294 nachgewiesen. (DIN, S. 11)		
296	Der Auftragnehmer hat dem Auftraggeber seine regelmäßige fachliche Fortbildung nachgewiesen. (DIN, S. 11)		
	DIN 33430, Aussagen zu Qualitätsanforderungen zur Durchführung und Auswertung von Verhaltensbeobachtungen und -beurteilungen	Korrespondenz	Standards
V67	Verzweigungsfrage: Sind der Auftragnehmer (und ggf. Mitwirkende) an der Durchführung und Auswertung von Verhaltensbeobachtungen und/oder -beurteilungen beteiligt? Falls „nein“ → Bitte weiter bei Verzweigungsfrage V68.		

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zu Anforderungen an die Qualifikation der an der Eignungsbeurteilung beteiligten Personen (Checkliste 6)			
	DIN 33430, Aussagen zu Qualitätsanforderungen an den Auftragnehmer	Korrespondenz	Standards
	Der Auftragnehmer <u>und</u> (sofern beteiligt) die Mitwirkenden verfügt/verfügen über ...		
297	- Kenntnisse über einschlägige Evaluationen von Verfahren zur Verhaltensbeobachtung <u>und</u> -beurteilung (DIN, S. 11)		
298	- Kenntnisse der Rahmenbedingungen von Verhaltensbeobachtung <u>und</u> -beurteilung		
299	- Kenntnisse zum Thema „Beobachtung“: Begriff <u>und</u> Verständnis (DIN, S. 11)		
300	- Kenntnisse über die Systematik der Beobachtung (DIN, S. 11)		
301	- Kenntnisse über Operationalisierungen von Eignungsmerkmalen (DIN, S. 11)		
302	- Kenntnisse über die Definition <u>und</u> Abgrenzung von Beobachtungseinheiten (DIN, S. 11)		
303	- Kenntnisse über die Registrierung <u>und</u> Dokumentation der Beobachtungen (DIN, S. 11)		
304	- Kenntnisse über die Auswertung/Bewertung der Beobachtungen (DIN, S. 11)		
305	- Kenntnisse über Bezugsmaßstäbe (DIN, S. 11)		
306	- Kenntnisse über Rating-/Skalierungsverfahren (DIN, S. 11)		
307	- Kenntnisse über verschiedene Formen (statistische <u>und</u> nicht statistische) der Urteilsbildung (DIN, S. 11)		
308	- Kenntnisse über Beobachtungsfehler/-verzerrungen (DIN, S. 11)		
309	- Kenntnisse über Gütekriterien (Objektivität <u>und</u> Zuverlässigkeit (auch Übereinstimmung) <u>und</u> Gültigkeit). (DIN, S. 11)		

Tabelle 5 (Fortsetzung):

Aussagen der DIN 33430 (in der Fassung der DIN Screen; Kersting, 2008a) und ihre Zuordnung zu den Standards (ein doppeltes oder einfaches Pluszeichen weist auf eine hohe bzw. mittlere Korrespondenz hin, ein fehlender Eintrag auf fehlende Übereinstimmung)

DIN 33430, Aussagen zu Anforderungen an die Qualifikation der an der Eignungsbeurteilung beteiligten Personen (Checkliste 6)			
	<i>DIN 33430, Aussagen zu Qualitätsanforderungen zur Durchführung und Auswertung von Eignungsinterviews</i>	<i>Korrespondenz</i>	<i>Standards</i>
V68	<i>Verzweigungsfrage: Sind der Auftragnehmer (und ggf. Mitwirkende) an der Durchführung und Auswertung von Eignungsinterviews beteiligt? Falls „nein“ → Bitte weiter bei Verzweigungsfrage V69.</i>		
	Der Auftragnehmer <u>und</u> (sofern beteiligt) die Mitwirkenden verfügt/verfügen über ...		
310	– Kenntnisse der Rahmenbedingungen von Verfahren zur mündlichen Informationsgewinnung (DIN, S. 11)		
311	– Kenntnisse über einschlägige Evaluationen von Eignungsinterviews (DIN, S. 11)		
312	– Kenntnisse über Interviewklassifikationen (DIN, S. 11)		
313	– Kenntnisse über die Handhabung von Interviewleitfäden (DIN, S. 11)		
314	– Kenntnisse über Fragetechniken <u>und</u> Formulierungstechniken (DIN, S. 11)		
315	– Kenntnisse über Kriterien zur Beurteilung der Aussagen des Interviewten (DIN, S. 11)		
316	– Kenntnisse über Fragebereiche		
317	– Kenntnisse über die rechtliche Zulässigkeit von Interviewfragen. (DIN, S. 11)		
V69	<i>Verzweigungsfrage: Werden Sachverständige zur Durchführung der Verfahren der mündlichen Informationsgewinnung hinzugezogen, um fachliche Kenntnisse und/oder Fertigkeiten der Kandidaten zu erkunden? Falls „nein“ → Ende der Checkliste 6.</i>		
318	Die zur Durchführung der Verfahren der mündlichen Informationsgewinnung hinzugezogenen Personen, die einen Beitrag zur Erkundung der fachlichen Kenntnisse und/oder Fertigkeiten der Kandidaten leisten, arbeiten (sofern sie nicht selbst im Sinne der DIN 33430 qualifiziert sind) nicht allein, sondern mit Personen zusammen, die im Sinne der DIN 33430 qualifiziert sind. (DIN, S. 12)		

4 Vergleich der Standards mit der DIN 33430

Im vorherigen Kapitel wurden die normativen sowie einige nicht normative Aussagen der DIN 33430 (in der Ordnung der DIN Screen, Kersting, 2008a) den Standards gegenübergestellt. Im Folgenden werden die Ergebnisse der Gegenüberstellung ausgewertet, dabei soll folgenden Fragen nachgegangen werden: Welche Gemeinsamkeiten und Unterschiede finden sich in den beiden Regelwerken? Allein von der Quantität her muss die Teilmenge der Gemeinsamkeiten gering ausfallen, die weiter oben in Abschnitt 3 wiedergegebenen Aussagen der DIN 33430 entsprechen nur knapp 40 % der Textmenge der Standards. Bedenkt man außerdem den deutlich weiter gefassten Geltungsbereich der Standards, muss es überraschen, dass sich für 29 % der Standards eine zumindest mittlere Korrespondenz zu den Aussagen der DIN 33430 ergibt, wobei die Prozentwerte vor dem Hintergrund zu interpretieren sind, dass zahlreiche Aussagen der DIN 33430 mehreren Standards zugeordnet wurden. Dies wiederum ist der Gliederung der Standards in Anwendungsbereiche geschuldet, die einige inhaltliche Wiederholungen und Redundanzen mit sich bringt. Dies lässt sich an einem Beispiel verdeutlichen: Welche Angaben zu empirischen Untersuchungen notwendig sind, wird in der DIN 33430 einmal (und nur einmal, DIN-Screen-Aussagen 45 bis 49), in den Standards aber in verschiedenen Abschnitten erläutert (z. B. Standards 1.5., 1.13, 2.4, 3.8 und 4.6). Dies erklärt, warum etliche Aussagen der DIN 33430 mehreren Standards zugeordnet werden konnten und mussten.

Überraschend ist, dass es nicht nur Standards gibt, für die es keine Entsprechung in der DIN 33430 gibt, sondern dass umgekehrt für 188 Aussagen der DIN 33430 keine Entsprechung in den Standards gefunden werden konnte (vgl. Tab. 5), dies entspricht 60 % der Aussagen der DIN 33430. Das heißt, dass neben Gemeinsamkeiten beide Qualitätsstandards auch Eigenständigkeiten aufweisen, deren Analyse zum Verständnis der jeweils unterschiedlichen diagnostischen Qualitätsauffassungen beiträgt und Entwicklungspotenziale für beide Regelwerke aufzeigt.

4.1 Standards und DIN 33430: Gemeinsamkeiten und Unterschiede aus der Perspektive der Standards

Tabelle 6 listet für jedes Kapitel der Standards den Anteil der Aussagen auf, die eine zumindest mittlere Korrespondenz zu den Forderungen der DIN 33430 aufweisen. Da die Standards nach Einschätzung von Jeanneret und Zedeck (2010), die einen Vergleich der drei Richtlinien (1) Standards, (2) Principles for the validation and use of personnel selection procedures (vgl. Abschnitt 2.1.3) und (3) Uniform Guidelines on Employee Selection Procedures (vgl. Abschnitt

2.1.4) durchgeführt haben, eine hohe Übereinstimmung mit den Principles aufweisen, können die nachfolgend dargestellten Ergebnisse teilweise auch auf die Principles der SIOP generalisiert werden. Der Vergleich der Standards mit der DIN 33430 verdeutlicht die unterschiedliche Testpraxis in Nordamerika und Deutschland. In den USA sind Testprogramme seit langem weit verbreitet, wie sie in Deutschland erst durch die PISA-Studie in der breiten Öffentlichkeit bekannt wurden. Entsprechend finden sich zahlreiche Richtlinien in den Standards (nicht aber in der DIN), die sich auf Testprogramme, die Verwendung von Tests zur Programmevaluation und für Lizenzierungen bzw. Zertifizierungen (z. B. Standards 14.14 bis 14.17) sowie auf Large Scale Testing beziehen. Häufiger als in Deutschland stützt sich in Nordamerika die Politik auf Testergebnisse, dies wird entsprechend in den Standards thematisiert (z. B. Standards 15.4, 15.5, 15.11, 15.12 und 15.13). Die Standards gehen explizit auf die aus der Testung abgeleiteten Entscheidungen und deren rechtlichen Anfechtbarkeit ein (z. B. Standards 1.23, 1.24, 2.19, 6.6, 7.9, 8.10 bis 8.13; 11.11, 11.17, 11.18 und 11.24). Das gesamte Kapitel 13 widmet sich dem in der DIN 33430 wegen des eignungsdiagnostischen Fokus kaum thematisierten pädagogischen Testen, wobei auch hier stets die Auswirkungen betrachtet werden (z. B. Standard 15.7). Die unterschiedliche Testpraxis zeigt sich u. a. auch darin, dass die Standards sich zu speziellen Dienstleistungen äußern, z. B. zu Testauswertediensten (z. B. Standards 5.8 und 6.12), die in Deutschland (noch) kaum nachgefragt werden und entsprechend auch nicht geregelt sind. Auch der Umgang mit den Ergebnissen lokaler Studien (z. B. Standards 2.13, 13.4, 14.2 und 14.6) ist das Thema von Richtlinien für ein großes, nicht aber für ein relativ kleines Land wie Deutschland.

In den USA spielt naturgemäß das Thema Diversity (z. B. Standards 6.10, 7.3, 7.4, 7.7, 7.8, 7.10., 7.11 und 12.2) eine viel größere Rolle als in Deutschland, wobei insbesondere auch sprachliche Unterschiede und ihre potenziellen Auswirkungen auf die Diagnostik betrachtet werden (z. B. das gesamte Kapitel 9 der Standards). Auch der diagnostische Umgang mit behinderten Menschen, der in der DIN 33430 nach Ansicht von Wegener (2003) ungebührlich vernachlässigt wird, nimmt in den Standards (schon allein aufgrund einer ganz eigenen Rechtsgeschichte in Nordamerika) einen großen Raum ein (z. B. das gesamte Kapitel 10, sowie Standard 11.23). Zu den insgesamt 23 Richtlinien der Kapitel 9 und 10 der Standards, finden sich keine korrespondierenden Aussagen der DIN 33430.

Die Unterschiedlichkeit der nordamerikanischen und deutschen Qualitätsstandards spiegelt auch die unterschiedlichen Rechtssysteme der beiden Länder wieder. Über einen Ländervergleich des rechtlichen Umfelds der Diagnostik, in den 22 Länder einbezogen wurden, berichten Myors et al. (2008) sowie Sackett et al. (2010). In den USA als Einwanderungsland gibt es deutlich mehr rechtliche Regelungen in Bezug auf unterschiedliche Bevölkerungsgruppen als in

Tabelle 6:
Korrespondenz zwischen den Standards und der DIN 33430 auf der Basis der Gliederung der Standards

Ab-schnitt	Thema	Anzahl der Aussagen	Anteil der Aussagen, die eine zumindest mittlere Korrespondenz zu den Forderungen der DIN 33430 aufweisen
06	Unterstützende Dokumentation für Tests	15	60 %
01	Validität	24	58 %
03	Testentwicklung und Testrevision	27	48 %
05	Testdurchführung, Testbewertung und Ergebnisdarstellung	16	38 %
12	Psychologisches Testen und psychologische Diagnostik	20	35 %
08	Rechte und Pflichten der Testteilnehmer	13	31 %
07	Fairness bei der Testanwendung und bei der Testung	12	25 %
11	Pflichten der Testanwender	24	25 %
04	Skalierung, Normen und Testwertvergleichbarkeit	21	24 %
14	Testen im Kontext der Eignungsdiagnostik und der beruflichen Zulassung und Zertifizierung	17	24 %
02	Reliabilität und Messfehler	20	20 %
15	Testen im Kontext von (Lehr-)Programmevaluationen sowie zur Unterstützung (bildungs-)politischer Entscheidungen	13	8 %
13	Pädagogisches Testen und pädagogische Diagnostik	19	5 %
09	Das Testen von Personen mit unterschiedlichem sprachlichen Hintergrund	11	0 %
10	Das Testen von Personen mit Behinderungen	12	0 %
Gesamt		264	29 %

Deutschland. Eine weitere Besonderheit, die sich indirekt auf die Standards auswirkt, ist das extensive Fallrecht der USA. Deutsches Recht ist demgegenüber kodifiziertes Recht, der Umgang mit personenbezogenen Daten ist beispielsweise klar geregelt, die DIN 33430 fordert entsprechend lediglich (z. B. DIN-Screen-Aussage 230) alle gesetzlichen Vorgaben einzuhalten (DIN 2002, S. 12). Demgegenüber finden sich dazu mehrere Richtlinien in den Standards, so muss etwa extra geregelt werden, dass bei der Übermittlung von Testwerten die Vertraulichkeit der Testwerte nicht gefährdet werden darf (Standard 5.13), dass Organisationen, die Testergebnisse aufbewahren, bezüglich dieser Aufbewahrung klare Richtlinien benötigen (z. B. Aufbewahrungsdauer, Verwendung etc., vgl. Standards 5.15 und 5.16).

Zahlreiche Korrespondenzen zwischen den Standards und der DIN 33430 finden sich bei den Themen Durchführung, Auswertung und Interpretation diagnostischer Verfahren und bei den Testgütekriterien sowie vor allem beim Thema der transparenten Information über den Test (Kapitel 6 der Standards einerseits und Checkliste 1 der DIN Screen andererseits). Bezüglich der Themen Testentwicklungen und Testgütekriterien sind die Standards aber wesentlich detaillierter als die DIN 33430. So werden beispielsweise Besonderheiten der Testentwicklung angesprochen wie Testverkürzungen und variable Testinstruktionen (z. B. Standards 3.2, 3.3, 3.10, 3.13, 3.16, 3.18 und 3.21), die Kombination von Tests (z. B. Standards 12.4 und 12.5), das kriteriumsorientierte Testen (z. B. Standard 3.4) und die Nutzung von Geräten (auch unter Anwenderperspektive, z. B. Standard 5.5). Die Standards gehen wesentlich ausführlicher als die DIN 33430 auf die Themen Reliabilität im Allgemeinen und Item-Response-Theorie (z. B. Standard 3.9) im Besonderen ein, außerdem widmen sie sich wesentlich ausführlicher dem Thema Skalierung. Bezüglich dieser Aspekte könnte eine Fortschreibung der DIN 33430 von den Standards lernen. Andere Alleinstellungsmerkmale sind die Standards zum adaptiven Testen (z. B. Standards 2.16 und 3.12) sowie die zahlreichen Hinweise auf mögliche Fehler beim Testen bzw. bei der Testinterpretation (z. B. Standards 12.14 und 12.19). Anders als die DIN 33430 formulieren die Standards auch methodische Vorgaben, wie eine bestimmte Analyse vorzunehmen ist (z. B. Standard 7.6). Während die DIN sich lediglich allgemein zum Thema „Normen“ äußert, thematisieren die Standards im Kapitel 4 den Prozess der Normierung auf der Detailebene und beleuchten etwa das Thema „Equating“ (z. B. Standards 4.11 bis 4.14) und spezifische Normierungstechniken (z. B. Standard 4.15).

Bezüglich der Gütekriterien konnten zwar etliche Anforderungen der DIN den Standards zugeordnet werden (z. B. die DIN-Screen-Aussage 68 den Standards 2.1, 2.17 und 2.18), man muss aber eingestehen, dass die Standards beispielsweise das Thema Reliabilität wesentlich gründlicher behandeln und z. B. die Bedeutung von Standardfehlern diskutieren. Die Inhaltsvalidität wird stärker

als in Deutschland auch als Nachweis für eignungsdiagnostische Tests diskutiert und akzeptiert (z. B. die Standards 14.9 ff.), ggf. liegt das an der stärkeren Verbreitung „echter“ Arbeitsproben in Nordamerika (Roth, Bobko & McFarland, 2005).

4.2 Standards und DIN 33430: Gemeinsamkeiten und Unterschiede aus der Perspektive der DIN

In den Tabellen 5 und 7 wird die Korrespondenz der Standards mit der DIN 33430 aus der Perspektive der DIN 33430 dokumentiert, Tabelle 5 leistet dies auf der Ebene der 318 Einzelanforderungen der DIN 33430, Tabelle 7 gibt einen Überblick auf der Ebene der Gliederung der DIN Screen (Kersting, 2008a).

Die Anforderungen an die Qualifikation der an der Eignungsbeurteilung beteiligten Personen werden in der DIN 33430 deutlich ausführlicher und spezifischer thematisiert als in den Standards. Wie bereits ausgeführt hat dieses

Tabelle 7:
Korrespondenz zwischen der DIN 33430 und den Standards auf der Basis der Gliederung der DIN Screen (Kersting, 2008a)

Thema der DIN 33430	Anzahl der Aussagen	Anteil der Aussagen, die eine zumindest mittlere Korrespondenz zu den Standards aufweisen
Anforderungen an Verfahrenshinweise	140	61 %
Durchführung, Auswertung und Interpretation der Verfahrensergebnisse sowie Urteilsbildung	55	51 %
Auswahl und Zusammenstellung von Verfahren	20	45 %
Planung von berufsbezogenen Eignungsbeurteilungen	33	18 %
Dokumentation	22	9 %
Anforderungen an die Qualifikation der an der Eignungsbeurteilung beteiligten Personen	48	0 %
Gesamt	318	41 %

Thema in Nordamerika u. a. die „Task Force on Test User Qualifications“ aufgegriffen (vgl. z. B. Fremmer, 1996; Turner et al., 2001).

Die DIN 33430 sieht, wie alle Richtlinien die von Normungsinstitutionen wie DIN, CEN oder ISO erstellt werden, sehr umfassende Dokumentationen vor. Diese Dokumentationen sind die Grundlage für Soll-Ist-Vergleiche. In dieser Hinsicht ist die DIN 33430 den Standards deutlich überlegen (vgl. Checkliste 5 des DIN Screen).

Eigenständig ist die DIN 33430 im Vergleich zu den Standards auch in der Frage der Auftragsklärung und der eindeutigen Regelung von Verantwortlichkeiten, beides Aspekte der Planung des diagnostischen Prozesses (Checkliste 2 des DIN Screen). Die Themen Vorauswahl (Aussagen 225 bis 228) und Anforderungsanalyse (Aussagen 151 bis 153 sowie 166 bis 172) werden in der DIN 33430 wesentlich ausführlicher reflektiert als in den Standards.

Andere Unterschiede beziehen sich auf Einzelaspekte. Wenngleich das Thema Reliabilität in den Standards insgesamt ausführlicher behandelt wird (vgl. Abschnitt 4.1), findet sich nur in der DIN 33430 die ausdrückliche Forderung des Nachweises der Retest-Reliabilität für solche Verfahren, die Merkmale erfassen, für die eine zumindest relative Zeit- und Situationsstabilität angenommen wird (Aussage 74). Einzig in der DIN 33430 wird das Problem der „Veralterung“ von Analyseergebnissen dadurch geregelt, dass konkrete Prüfintervalle angegeben werden. So müssen z. B. die Reliabilitäts- und Validitätsangaben zu einem Verfahren mindestens alle 8 Jahre überprüft werden (Aussagen 76 und 90). Ebenso ist die Angemessenheit der Normwerte spätestens alle 8 Jahre zu prüfen (Aussage 58). Die Aktualität von Normwerten wird in den Standards zwar auch angesprochen (z. B. Standard 4.18), hier wird aber kein konkretes Zeitintervall genannt.

Anders als die Standards gehen die Richtlinien der DIN 33430 explizit auf Aspekte der Objektivität und Reliabilität bei Verfahren ein, bei der Personen aufgrund von mündlich gewonnenen Informationen und/oder aufgrund von Verhaltensbeobachtungen beurteilt werden (z. B. Interviews oder Assessment-Center). Für den Fall, dass mehrere Beurteiler eine Person beurteilen, fordert die DIN 33430 eine Analyse der Beurteilerübereinstimmung (Aussagen 210 und 263).

5 Fazit und Ausblick

Diagnostisches Handeln bedarf systematisch entwickelter „Leitlinien“, die auf aktuellen wissenschaftlichen Erkenntnissen und Praxiserfahrungen beruhen und in einem Konsensprozess von Experten gewonnen wurden. Es gibt zahlreiche

verschiedene diagnostische Qualitätsstandards, die sich deutlich voneinander unterscheiden. Ausgewählte Qualitätsstandards wurden im vorliegenden Beitrag besprochen, wobei die nordamerikanischen Standards und die DIN 33430 im Fokus des Interesses standen, da diese beiden Ansätze in mehrfacher Hinsicht unterschiedliche Perspektiven auf das Thema Qualität der Diagnostik repräsentieren. Die Unterschiede zwischen den Standards einerseits und der DIN 33430 andererseits sind zum einen darauf zurückzuführen, dass die thematischen Schwerpunkte der Standards von dem rechtlichen Rahmenbedingungen und der kulturell-soziologischen Situation in den USA und somit national geprägt sind. Zum anderen repräsentieren die Standards stärker die Perspektive der Wissenschaft, während die DIN die Perspektive der Wirtschaft stärker berücksichtigt. Diese Unterscheidung von Wissenschaft und Wirtschaft bezieht sich hier weder auf die Inhalte, noch auf die Autoren (sowohl die nordamerikanischen Standards als auch die DIN 33430 wurde von namhaften Wissenschaftlern verfasst), sondern auf die gewählte Methode der Qualitätssicherung.

Die Standards repräsentieren das qualitätssichernde Vorgehen einer von Wissenschaftlern dominierten Fachwelt, die vor allem auf die bindende Kraft der von Autoritäten erstellten, fachlich einwandfreien und möglichst erschöpfenden Qualitätsstandards setzt. Tatsächlich wird die Breite und Tiefe der in den Standards verarbeiteten Kenntnisse von kaum einer anderen Richtlinie erreicht. Dies liegt u. a. daran, dass der Mitarbeit an Qualitätsstandards in Nordamerika interkollegial eine hohe Bedeutung eingeräumt wird. So konnten die nordamerikanischen Standards von fast 8.000 Seiten Feedback von namhaften Personen und Organisationen zu den ersten Entwürfen profitieren (vgl. das Vorwort zu den Standards, APA, 1999). Im Vergleich dazu wurden zur Vorveröffentlichung der DIN 33430 (dem sogenannten Gelbdruck) gerade einmal 55 Einsprüche formuliert, die 201 Seiten füllten (vgl. Kersting, 2008a).

Das grundsätzliche Problem berufsständisch formulierter Richtlinien gilt aber auch für die Standards: deren Unverbindlichkeit. Mit den Richtlinien werden Ziele formuliert, ohne dass die Zielerreichung und ihre Nachprüfung thematisiert werden. Entsprechend werden die Texte häufig als berufsständische Tugendlehre rezipiert, viel gelesen und von einer moralischen Komponente durchtränkt zitiert, aber in der Praxis wenig beachtet. Über die mit den Qualitätsstandards erreichte Zielformulierung hinaus kommt es darauf an, den Verantwortlichen ein Feedback über ihre Zielerreichung zur Verfügung zu stellen und somit die Grundlage für eine lernende Diagnostik zu schaffen. Die DIN 33430 ist der erste diagnostische Standard, der sich den in der Wirtschaft üblichen Techniken des Qualitätsmanagements bedient und außerhalb einer Berufsstandesorganisation verfasst wurde. Mit der DIN 33430 sowie deren Folgeinitiativen werden diagnostische Qualitätsanforderungen erstmalig nicht nur beschrieben, sondern auch (z. B. durch Checklisten, Testbeurteilungen, Lizenzprüfungen, Zertifizie-

rungen und/oder Auditierungen) einer gesellschaftlichen Bewertung zugeführt, sodass Qualität (im Sinne von DIN-Normen, vgl. Abschnitt 4.2) nach außen sichtbar wird. Die DIN 33430 und die darauf aufbauenden Checklisten, Lizenzen, Zertifikate und Testbeurteilungen stellen einen Gütemaßstab dar, der bei leistungsmotivierten Personen Erwartungen weckt. Erst durch die Verbindung des diagnostischen Handelns mit einem klaren Ziel einerseits und einem Feedback andererseits wird psychologisch gesehen aus der diagnostischen Handlung eine Leistung (vgl. Kersting, 2008a).

Es liegt nahe, die beiden Qualitätssicherungsansätze der Wissenschaft und der Wirtschaft in Zukunft noch weiter zu verzahnen. Die in 2006 gestartete Entwicklung einer ISO-Norm „Assessment Service Delivery: Procedures and methods to assess people in work and organizational settings“ bietet hierzu eine gute Gelegenheit, da sowohl die Standards als auch die DIN 33430 (neben weiteren nationalen und internationalen Texten) als Grundlage herangezogen werden.

Erstaunlicherweise überlappen sich die Aussagen der auf die Eignungsdiagnostik beschränkten DIN 33430 und der für alle Bereiche der Diagnostik formulierten Standards in vielen Bereichen erheblich. Dabei ist beachtenswert, dass die meisten Korrespondenzen zwischen den Standards und der DIN 33430 keinesfalls im Abschnitt 14 zur Eignungsdiagnostik erzielt werden, sondern in anderen Bereichen wie den Informationsanforderungen an Verfahrenshinweise (Abschnitt 6 der Standards), Validität (Abschnitt 1), Testentwicklung und -revision sowie Testdurchführung, Testauswertung und -interpretation (Abschnitte 3 und 5). Daraus lässt sich ableiten, dass die Einschränkung des Geltungsbereiches der DIN 33430 auf die Eignungsdiagnostik nicht notwendig ist, sondern die DIN 33430, wie die Standards, in allen Bereichen der Diagnostik genutzt werden kann. Diese Erkenntnis wurde mit dem Testbeurteilungssystem des Testkuratoriums (vgl. Abschnitt 2.3.3.3) bereits umgesetzt, welches 140 Aussagen der DIN 33430 inkorporiert und offiziell zum „Standard zur Information und Dokumentation von Instrumenten zur Erfassung menschlichen Erlebens und Verhaltens des Testkuratoriums der Föderation Deutscher Psychologenvereinigungen“ ausruft, und zwar für Tests aus allen Anwendungsbereichen der Psychologie. Die hier geleistete Auswertung zeigt, dass noch weitere Bestandteile der DIN 33430 für alle Anwendungsgebiete der Diagnostik genutzt werden können.

Als weiteres Element der Qualitätssicherung und -optimierung sollten Wettbewerbe von diagnostisch arbeitenden Personen und/oder Organisationen um Qualitätspreise hinzukommen. Sofern die Preise ein hohes fachliches wie gesellschaftliches Ansehen haben, kann die Preisverleihung sich positiv auf die Preisträger auswirken. Dies sollte dazu beitragen, dass Qualität kein Selbst-

zweck ist, sondern in einem härter werdenden Wettbewerb die Marktchancen der Anbieter mit gesicherter diagnostischer Qualität erhöht. Die wirtschaftliche Seite wurde bei den bisherigen Qualitätsstandards vernachlässigt. Grundsätzlich kann soziale Kohäsion durch Zwangsmittel (Sanktionen), normative Mittel (Stimme der Fachwelt/Gemeinschaft) und utilitäre Mittel (ökonomische Anreize) herbeigeführt werden. Mit wenigen Ausnahmen findet psychologische Diagnostik in einem gesetzlich nicht geregelten Bereich statt, sodass Zwangsmittel entfallen. Zumindest die beiden verbleibenden Mittel sollten ausgeschöpft werden. Somit sollten zusätzlich zu den normativen Mitteln, dem bislang einzigen Mittel, das die Wissenschaftler nutzen, auch utilitäre Mittel zum Einsatz kommen. Darüber hinaus spricht das folgende Argument für Wettbewerbe: Allen Richtlinien, also auch den Standards und der DIN Norm, liegt ein potenziell statisches und somit innovationshemmendes Erfüllungsparadigma zugrunde, welches dem dynamischen Erkenntnisfortschritt nicht gerecht wird. Richtlinien sollten daher durch Wettbewerbe um ein dynamisches Optimierungsparadigma ergänzt werden.

Schließlich ist es notwendig, nicht nur Anforderungen an die Qualifikation von diagnostisch tätigen Personen zu formulieren, sondern in Form von systematischen Aus- und Fortbildungen auch Hilfestellungen zum Qualifikations- und Kompetenzerwerb und -ausbau anzubieten (vgl. Abschnitt 2.2 sowie Kersting, 2010). Entsprechende Angebote sollte es sowohl für Personen geben, die ein Psychologiestudium erfolgreich absolviert haben, aber noch einen spezifischen Nachholbedarf in Diagnostik sehen, als auch für Personen, die kein einschlägiges Studium/keine einschlägige Ausbildung absolviert haben und aufgrund ihrer betrieblichen Position wie Aufgaben dessen bedürfen.

Literatur

- Abeln, C. & Reimann, G. (2004). DIN 33430 und die Folgen. Personalauswahl und -entwicklung im Umbruch. *Arbeit und Arbeitsrecht*, 11, 8–15.
- American Educational Research Association (AERA), American Psychological Association (APA) & National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- American Psychological Association (1950). Ethical standards for the distribution of psychological tests and diagnostic aids. *American Psychologist*, 5, 620–626.
- American Psychological Association (1981). Specialty guidelines for the delivery of services by industrial-organizational psychologists. *American Psychologist*, 36, 664–669.
- American Psychological Association (2002). Ethical principles of psychologists and code of conduct. *American Psychologist*, 57, 1060–1073.

- Arbeitskreis Assessment Center (2004). *Standards der Assessment Center Technik*. Hamburg: Autor.
- Arbeitskreis Assessment Center (2008). *Interview-Standards*. Hamburg: Autor.
- Association of Test Publishers (2002). *Guidelines for computer-based testing*. Washington, DC: Author.
- Bartram, D. (1996). Test qualifications and test use in the UK: The competence approach. *European Journal of Psychological Assessment, 12*, 62–71.
- Bartram, D. (2001). Guidelines for test users: A review of national and international initiatives. *European Journal of Psychological Assessment, 17*, 173–186.
- Bartram, D. & Hambleton, R. (Eds.). (2006). *Computer-based testing and the internet: Issues and advances*. New York: Wiley.
- Bartram, D., Lindley, P. A. & Foster, J. M. (1990). *A review of psychometric tests for assessment in vocational training*. Sheffield: The Training Agency.
- Berufsverband Deutscher Psychologinnen und Psychologen (n. d.). *Psychologische Testverfahren bei Einstellungsuntersuchungen, 7 Tipps für Bewerber*. Bonn: Autor.
- Berufsverband Deutscher Psychologinnen und Psychologen (1998). *Berufsordnung für Psychologen*. Bonn: Deutscher Psychologenverlag.
- Brähler, E., Holling, H., Leutner, D. & Petermann, F. (2002). *Brickenkamp Handbuch psychologischer und pädagogischer Tests* (3., vollständig überarbeitete und erweiterte Aufl., Band 1 und 2). Göttingen: Hogrefe.
- Chartered Institute of Personnel and Development, SHL, The British Psychological Society, The Department of Trade and Industry, University of Surrey Roehampton (2004). *360 degree feedback: Best practice guidelines*. London: Author.
- Conway, J. M., Jako, R. A. & Goodman, D. F. (1995). A meta-analysis of interrater and internal consistency reliability of selection interviews. *Journal of Applied Psychology, 80*, 565–579.
- Cronbach, L. J. & Gleser, G. C. (1965). *Psychological tests and personal decisions*. Chicago, IL: University of Illinois Press.
- Davidson, G. (1997). The ethical use of psychological tests: Australia. *European Journal of Psychological Assessment, 13*, 132–139.
- DIN (2002). *DIN 33430: Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen*. Berlin: Beuth.
- Dormann, C., Moosbrugger, H., Stemmler, G. & Maier, G. A. (2009). Erwerb von Personenlizenzen zur DIN 33430 im Rahmen des Psychologiestudiums. *Psychologische Rundschau, 60*, 23–26.
- Eignor, D. R. (2001). Standards for the development and use of tests: The standards for educational and psychological testing. *European Journal of Psychological Assessment, 17*, 157–163.
- Equal Employment Opportunity Commission, Department of Labor, Department of Justice & Civil Service Commission. (1978). *Uniform Guidelines on Employee Selection Procedures*. 29 CFR, 1607.
- Evers, A. (1996). Regulations concerning test qualifications and test use in The Netherlands. *European Journal of Psychological Assessment, 12*, 153–159.

- Evers, A. (2001a). Improving test quality in The Netherlands: Results of 18 years of test ratings. *International Journal of Testing, 1*, 137–153.
- Evers, A. (2001b). The Revised Dutch Rating System for Test Quality. *International Journal of Testing, 1*, 155–182.
- Evers, A., van Vliet-Mulder, J. C. & Groot, C. (2000). *Documentatie van tests en testresearch in Nederland, dl. 1 en 2*. Amsterdam: NIP/Van Gorcum.
- Fernández-Ballesteros, R., De Bruyn, E. E. J., Godoy, A., Hornke, L. F., Ter Laak, J., Vizcarro, C. et al. (2001). Guidelines for the assessment process (GAP): A proposal for discussion. *European Journal of Psychological Assessment, 17*, 187–200.
- Fischer, G. H. (1968). *Einführung in die Theorie psychologischer Tests: Grundlagen und Anwendungen*. Bern: Huber.
- Fisseni, H. & Fennekels, G. P. (1995). *Das Assessment-Center*. Göttingen: Verlag für Angewandte Psychologie.
- Föderation Deutscher Psychologenverbände (1986). Beschreibung der einzelnen Kriterien für die Testbewertung. *Diagnostica, 32*, 358–360.
- Franke, J. (1969). Eine Konzeption zum systematischen Aufbau von Eignungsuntersuchungen. *Psychologische Beiträge, 11*, 390–405.
- Fremer, J. (1996). Promoting high standards for test use: Developments in the United States. *European Journal of Psychological Assessment, 12*, 160–168.
- Häcker, H., Leutner, D. & Amelang, M. (1998). *Standards für pädagogisches und psychologisches Testen*. Göttingen: Hogrefe.
- Hall, J. D., Howerton, D. L. & Bolin, A. U. (2005). The use of testing technicians: Critical issues for professional psychology. *International Journal of Testing, 5*, 357–375.
- Hambleton, R. K. (2001). The next generation of the ITC test translation and adaptation guidelines. *European Journal of Psychological Assessment, 17*, 164–172.
- Hambleton, R. K. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In R. K. Hambleton, P. Merenda & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 3–38). Hillsdale, NJ: Erlbaum.
- Herrmann, T. (1966). Sammelreferat: Zur Geschichte der Berufseignungsdiagnostik. *Archiv für die gesamte Psychologie, 118*, 253–278.
- Hilke, R. (2004). Qualitätsmanagement im psychologischen Dienst der Bundesanstalt für Arbeit. In L. F. Hornke & U. Winterfeld (Hrsg.), *Eignungsbeurteilungen auf dem Prüfstand: DIN 33430 zur Qualitätssicherung* (S. 81–102). Heidelberg: Spektrum.
- Höft, S. & Funke, U. (2006). Simulationsorientierte Verfahren der Personalauswahl. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (2. Aufl., S. 145–187). Göttingen: Hogrefe.
- Hornke, L. F. (2005). Die englische Fassung der DIN 33430. In K. Westhoff et al. (Hrsg.), *Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33430* (2. überarb. Aufl., S. 255–283). Lengerich: Pabst.
- International Test Commission (2001). International guidelines on test use. *International Journal of Testing, 1*, 95–114.

- International Test Commission (2006). International guidelines on computer-based and internet-delivered testing. *International Journal of Testing*, 6, 143–171.
- Jeanneret, P. R. & Zedeck, S. (2010). Professional guidelines/standards. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection* (pp. 593–625). New York: Routledge.
- Joint Committee on Testing Practices (1988). *The code of fair testing practices in education*. Washington, DC: American Psychological Association.
- Joint Committee on Testing Practices (1993). *Responsible test use: Case studies for assessing human behaviour*. Washington, DC: American Psychological Association.
- Joint Committee on Testing Practices (2000). *Rights and responsibilities of test takers: Guidelines and expectations*. Washington DC: Joint Committee on Testing Practices.
- Kanning, U. P. (2007). *Förderung sozialer Kompetenzen in der Personalentwicklung*. Göttingen: Hogrefe.
- Kersting, M. (2006). Zur Beurteilung der Qualität von Tests: Resümee und Neubeginn. *Psychologische Rundschau*, 57, 243–253.
- Kersting, M. (2008a). *Qualität in der Diagnostik und Personalauswahl: Der DIN Ansatz*. Göttingen: Hogrefe.
- Kersting, M. (2008b). Zur Akzeptanz von Intelligenz- und Leistungstests. *Report Psychologie*, 33, 420–433.
- Kersting, M. (2009). Qualität im Assessment Center. *Zeitschrift für Arbeits- und Organisationspsychologie*, 53, 70–74.
- Kersting, M. (2010). Diagnostische Fortbildung am Beispiel des Trainings zur Eignungsbeurteilung nach DIN 33430. In U. P. Kanning, L. v. Rosenstiel & H. Schuler (Hrsg.), *Jenseits des Elfenbeinturms. Psychologie als nützliche Wissenschaft* (S. 223–240). Göttingen: Vandenhoeck & Ruprecht.
- Kersting, M. & Püttner, I. (2006). Personalauswahl: Qualitätsstandards und rechtliche Aspekte. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (2. Aufl., S. 841–861). Göttingen: Hogrefe.
- Keyser, J. D. & Sweetland, R. C. (1984). *Test critiques*. Austin, TX: Pro-Ed.
- Leach, M. M. & Oakland, T. (2007). Ethics standards impacting test development and use: A review of 31 ethics codes impacting practices in 35 countries. *International Journal of Testing*, 7, 71–88.
- Liepmann, D., Beauducel, A., Brocke, B. & Amthauer, R. (2007). *Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R)* (2., überarbeitete und erweiterte Aufl.). Göttingen: Hogrefe.
- Lindsay, G., Koene, C., Øvreide, H. & Lang, F. (2008). *Ethics for European psychologists*. Göttingen: Hogrefe.
- Moosbrugger, H., Stemmler, G. & Kersting, M. (2008). Qualitätssicherung und -optimierung im Aufbruch: Die ersten Testrezensionen nach dem neuen TBS-TK-System. *Psychologische Rundschau*, 59, 182–184.
- Myers, B., Lievens, F., Schollaert, E., Van Hove, G., Cronshaw, S. F., Mladinic, A. et al. (2008). International perspectives on the legal environment for selection. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 1, 200–256.
- Naglieri, J. A., Drasgow, F., Schmit, M., Handler, L., Prifitera, A., Margolis, A. & Velasquez, R. (2004). Psychological testing on the Internet: New problems, old issues. *American Psychologist*, 59, 150–162.
- Neubauer, R. & Höft, S. (2006). Standards der Assessment Center-Technik – Version 2004. Überblick und Hintergrundinformationen. *Wirtschaftspsychologie*, 8, 77–82.
- Österreichisches Normungsinstitut (2005). *ÖNORM D 4000: Anforderungen an Prozesse und Methoden in der Personalauswahl und -entwicklung*. Wien: Autor.
- Plake, B. S. & Impara, J. C. (2001). *The fourteenth mental measurements yearbook*. Lincoln, NB: The Buros Institute of Mental Measurement.
- Puzicha, K. (1997). *Neue Wege in der Personalpsychologie* (Arbeitsbericht Nr. 1/1997 des Psychologischen Dienstes der Bundeswehr). Bonn: Bundesministerium der Verteidigung P II 4.
- Roth, P. L., Bobko, P. & McFarland, L. A. (2005). A meta-analysis of work sample test validity: Updating and integrating some classic literature. *Personnel Psychology*, 58, 1009–1037.
- Sackett, P. R., Shen, W., Myers, B., Lievens, F., Schollaert, E., Van Hove, G. et al. (2010). Perspectives from twenty-two countries on the legal environment for selection. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection* (pp. 651–676). New York: Routledge.
- Sarges, W. & Wottawa, H. (Hrsg.) (2004). *Handbuch wirtschaftspsychologischer Testverfahren* (2. überarb. und erw. Aufl.). Lengerich: Pabst.
- Schmidt, F. L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist*, 47, 1173–1181.
- Schmidt-Atzert, L. & Rauch, W. (2008) TBS-TK Rezension: „Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R) 2., erweiterte und überarbeitete Auflage“. *Report Psychologie*, 33, 303–304.
- Schuler, H. (1990). Personalauswahl aus der Sicht der Bewerber: Zum Erleben eignungsdiagnostischer Situationen. *Zeitschrift für Arbeits- und Organisationspsychologie*, 34, 184–191.
- Schuler, H. (1991). Ethische Probleme der (sozial-)psychologischen Forschung. In H. Lenk (Hrsg.), *Wissenschaft und Ethik* (S. 331–355). Stuttgart: Reclam.
- Schuler, H. (1993). Social validity of selection situations: A concept and some empirical results. In H. Schuler, J. L. Farr & M. Smith (Eds.), *Personnel selection and assessment. Individual and organizational perspectives* (pp. 11–26). Hillsdale, NJ: Erlbaum.
- Schuler, H. (1996). *Psychologische Personalauswahl. Einführung in die Berufsleistungsdiagnostik*. Göttingen: Hogrefe.
- Schuler, H. (2002). *Das Einstellungsinterview*. Göttingen: Hogrefe.
- Schuler, H. (2003). Kommentar zum Diskussionsbeitrag von Kersting und Hornke zur Qualitätssicherung und -optimierung in der Diagnostik. *Psychologische Rundschau*, 54, 180–181.
- Schuler, H. & Stehle, W. (1983). Neuere Entwicklungen des Assessment-Center-Ansatzes – beurteilt unter dem Aspekt der sozialen Validität. *Zeitschrift für Arbeits- und Organisationspsychologie*, 27, 33–44.
- Sektion Arbeits-, Betriebs- und Organisationspsychologie im Berufsverband Deutscher Psychologen. (1980). *Grundsätze für die Anwendung psychologischer Eignungsuntersuchungen in Wirtschaft und Verwaltung*. Bonn: Berufsverband Deutscher Psychologen.

- Simner, M. L. (1996). Recommendations by the Canadian Psychological Association for improving the North American safeguards that help protect the public against test misuse. *European Journal of Psychological Assessment, 12*, 72–82.
- Sireci, S. G. (1998). The construct of content validity. *Social Indicators Research, 45*, 83–117.
- Spies, R. A., Carlson, J. F. & Geisinger, K. F. (2010). *The eighteenth mental measurements yearbook*. Lincoln, NE: University of Nebraska Press.
- Task force on Assessment Center (2009). Guidelines and ethical considerations for assessment center operations. *International Journal of Selection and Assessment, 17*, 243–254.
- Testkuratorium der Föderation Deutscher Psychologenvereinigungen (2010). TBS-TK. Testbeurteilungssystem des Testkuratoriums der Föderation Deutscher Psychologenvereinigungen. Revidierte Fassung vom 09. September 2009. *Psychologische Rundschau, 61*, 52–56.
- Turner, S. M., DeMers, S. T., Fox, H. R. & Reed, G. M. (2001). APA's guidelines for test user qualifications. An executive summary. *American Psychologist, 56*, 1099–1113.
- Vijver, F. J. R. van de (1997). The ethical use of psychological tests: Australia and the United States. *European Journal of Psychological Assessment, 13*, 131–131.
- Wegener, M. (2003). Rechtliche Verbindlichkeit der DIN 33430 für Behörden und Gerichte. *DGP Informationen, 48*, 7–11.
- Westhoff, K., Hagemeyer, C., Kersting, M., Lang, F., Moosbrugger, H., Reimann, G. & Stemmler, G. (Hrsg.). (2010). *Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33430* (3., überarbeitete Aufl.). Lengerich: Pabst.
- Westhoff, K., Hornke, L. F. & Westmeyer, H. (2003). Richtlinien für den diagnostischen Prozess – Zur Diskussion gestellt. *Report Psychologie, 28*, 504–517.
- Whitfield, E. A., Feller, R. & Wood, C. (Eds.). (2009). *A counselor's guide to career assessment instruments* (5th ed.). Broken Arrow, OK: National Career Development Association.
- Woehr, D. J. & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology, 67*, 189–205.
- Zollondz, H.-D. (2006). *Grundlagen Qualitätsmanagement. Einführung in die Geschichte, Begriffe, Systeme und Konzepte* (2. Aufl.). München: Oldenbourg.