

# Kapitel 9

## Qualitätsstandards der Personalauswahl

Martin Kersting

*in: H. Schuler & U. P. Kanning (Hrsg.) (2014)  
Lehrbuch der Personalpsychologie  
3. Auflage, S. 325-356. Göttingen: Hogrefe.*

### Inhaltsübersicht

Überblick .....	325	9.3.2 Die ISO-Normen 10667-1 und 10667-2 .....	345
9.1 Definition des Begriffs „Qualität“ .....	326	9.3.3 Das Binnenverhältnis der DIN- und ISO-Normen zur Personalauswahl .....	347
9.2 Qualitätsstandards .....	329	9.4 Fazit .....	347
9.2.1 Allgemeine und verfahrensspezifische Richtlinien .....	330	Tipps für die Praxis .....	349
9.2.2 Kompetenzorientierte Qualitätsansätze .....	333	Lernquiz .....	350
9.2.3 Beurteilungssysteme für Tests und Fragebogen .....	334	Übung .....	352
9.3 Normen zur Personalauswahl .....	338	Weiterführende Literatur .....	352
9.3.1 Die DIN 33430 .....	340	Literatur .....	353

### Überblick

In Deutschland werden pro Jahr ca. 30 bis 50 Millionen Personalentscheidungen gefällt (Wottawa & Oenning, 2002). Dabei wird der Rechtmäßigkeit, der Qualitätssicherung und der Wirtschaftlichkeit eine zunehmend große Bedeutung beigegeben. Die interne und externe Personalauswahl muss rechtliche Vorgaben beachten (siehe Kapitel 28), sie sollte effizient (d. h. mit minimalen Kosten pro Leistung, siehe Kapitel 26 und 27) funktionieren und sollte einem definierten Qualitätsstandard gerecht werden. Qualitätsstandards stellen operatives Hintergrundwissen und standardisierte Techniken (im Sinne normierter Handlungsanweisungen) bereit, ihre Beachtung trägt zu einer qualitativ hochwertigen und rechtlich einwandfreien Personalauswahl bei. Der Beitrag thematisiert zunächst die Frage, wie Qualität in der Personalauswahl definiert werden kann, und gibt anschließend einen Überblick über einschlägige nationale und internationale Qualitätsstandards. Ausführlicher vorgestellt werden die DIN 33430 mit ihren „Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen“ sowie die ISO-Norm zur Eignungsdiagnostik.

#### Lernzielfragen:

- Was bedeutet Qualität in der Personalauswahl?
- An welchen konkreten Standards kann man die Gestaltung der Personalauswahl orientieren?

- Welche postgradualen Qualifizierungsangebote gibt es für Menschen, die im Rahmen der Personalauswahl (Teil-)Aufgaben übernehmen?
- Wie lässt sich die Qualität von Tests und Fragebogen beurteilen?
- Wo findet man Testrezensionen?

### Fallbeispiel: Problemstellung

In der deutschen Niederlassung eines weltweit agierenden Pharmakonzerns ist die Personalauswahl bis zur Ebene der Gruppenleiter dezentral organisiert. Während die Zentrale in München bei der Personalentscheidung viel Wert auf die Ergebnisse im mehrtägigen Assessment Center (AC) legt, trifft man die Entscheidung in Berlin aufgrund einer Kombination aus den Ergebnissen in einem Persönlichkeitsfragebogen (eingesetzt wird der Golden Profiler of Personality) sowie einem Interview. Schon häufig gab es Kritik an der ungleichen Vorgehensweise, zumal sich sowohl Gruppenleiter aus Berlin als auch aus München um Positionen als Abteilungsleiter bewerben können. Zwar wird die Auswahl der Abteilungsleiter aufgrund eines zentralen Verfahrens in Frankreich entschieden, durch die unterschiedlichen Verfahren in Berlin und München würden aber, so die Kritik, unterschiedliche hohe Hürden für die Zulassung zum Abteilungsleiterverfahren aufgebaut. In letzter Zeit wurde aufgrund von Einzelfällen die Kritik besonders laut, im AC würde lediglich nach Nasenfaktor entschieden. Der in Berlin zu absolvierende Persönlichkeitsfragebogen komme, so die Kritik, hingegen einem Seelenstriptease gleich. Außerdem würde – je nach Sympathie – mal das Ergebnis im Persönlichkeitsfragebogen und mal das Interview den Ausschlag für die Personalauswahl geben. Schließlich hat die Personalvertretung um eine Überprüfung der Auswahlverfahren gebeten. Das Beispiel veranschaulicht mögliche negative Auswirkungen einer Vernachlässigung des Themas der Qualität der Personalauswahl und dient im Folgenden zur Illustration des Nutzens von Qualitätsstandards.

## 9.1 Definition des Begriffs „Qualität“

*Everybody has won, and all must have prizes.*  
(Alice's Adventures in Wonderland)

Qualität in der Personalauswahl – wer würde diese nicht fordern und für seine Arbeit in Anspruch nehmen. Der Begriff „Qualität“ ist derart positiv besetzt, dass die meisten Anbieter und Anwender in Folge ihres verständlichen Strebens, sich selbst in positivem Licht zu sehen und darzustellen, überzeugt sind, ihre Prozesse und Verfahren seien qualitativ hochwertig. Umgekehrt wirft jeder dem anderen mangelhafte Qualität vor. Und auf den ersten Blick hat jeder Recht mit seiner Überzeugung, denn Qualität ist ein unscharfer Begriff, den man auf unterschiedliche Art und Weise deuten kann. Dies zeigt sich in unserem Fallbeispiel an der Einschätzung des genutzten Persönlichkeitsfragebogens. Während die einen nur einem

Persönlichkeitsfragebogen mit einer hohen Reliabilität Qualität zusprechen, ziehen die anderen ihre Qualitätsüberzeugung aus dem Markt: Qualität ist demzufolge, was am Markt Erfolg hat – und das sind häufig Verfahren, über deren Reliabilität keine Informationen vorliegen. Argumentieren die einen mit der Validität des Fragebogens, führen die anderen ihr subjektives Evidenzgefühl an: Das aus dem Persönlichkeitsfragebogen abgeleitete Gutachten empfinden sie als „absolut stimmig“ (was sich auch bei wertlosen Gutachten durch den Forer- oder Barnum-Effekt gut erklären lässt, siehe Kanning, 2010, Kersting, 2010a).

**Beachte:**

Die Frage, was unter „Qualität“ zu verstehen ist, kann nicht in allgemeiner Form beantwortet werden. Daher sind Diskussionen über die Qualität der Personalauswahl fruchtlos, solange man nicht klärt, welchen Qualitätsbegriff man zugrunde legt. Im Alltagssprachgebrauch wird das Wort Qualität häufig unreflektiert im Sinne von „Güte“ genutzt und mit kategorialen Einstufungen („sehr gut“, „mangelhaft“ usw.) verbunden. Im Sinne des professionellen Qualitätsmanagements, im Sinne von DIN und ISO-Normen (siehe unten), ist Qualität hingegen ein relationaler Begriff und wird als der Grad definiert, in dem ein Satz inhärenter Merkmale Anforderungen erfüllt (EN ISO 9000:2005).

Die Beurteilung der Qualität von Prozessen und/oder Instrumenten der Personalauswahl orientiert sich an bestimmten bereichsspezifischen Kriterien und erfolgt im Zuge eines Abgleichs von Soll und (messbarem) Ist. Qualität als Fachbegriff bezeichnet keine absolute „Güte“, sondern eine Beschaffenheit und bezieht sich auf Qualitätsforderungen, ist also relativ. Die Qualitätsforderungen wiederum variieren in Abhängigkeit von der jeweiligen diagnostischen Zielstellung. Versteht man Qualität als absolute Güte, müsste ein Fragebogen oder Test beispielsweise stets die höchste Reliabilität oder eine definierte Mindestreliabilität (z. B. mindestens .85) aufweisen. Berücksichtigt man hingegen die diagnostische Fragestellung, so leitet sich die Qualitätsforderung aus der Frage ab, wie reliabel der 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 festgesetzte 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. Für ein erstes Screening erfüllen Tests und Fragebogen mit einer Reliabilität von .60 unter bestimmten Rahmenbedingungen durchaus ihren Zweck. 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. 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.

Die logische Struktur der Qualitätsbeurteilung ist nach Kersting und Westmeyer (2013) die Folgende:  $x$  (das Beurteilungssubjekt) beurteilt die Qualität von

y (des Beurteilungsobjekts) als z (das Beurteilungsergebnis) relativ zu k (den Beurteilungskriterien) zur Zeit t. Die Beurteilungskriterien stellen die zwischen verschiedenen, miteinander kooperierenden Einheiten (Menschen, Organisationen) getroffenen Konventionen über Gegenstände und Prozesse dar. Die Konventionen werden durch soziale Transaktionen zwischen den Beteiligten ausgehandelt. Die weiter unten besprochenen DIN und ISO-Normen sind Beispiele für solche Konventionen. Konventionen variieren in Abhängigkeit vom jeweiligen Zeitgeist, daher ist es wichtig, den Zeitpunkt der jeweiligen Qualitätskonstruktion zu berücksichtigen. Anhand der verschiedenen Ausgaben der weiter unten näher erläuterten „Standards für pädagogisches und psychologisches Testen“ (erste Fassung 1954, aktuelle Fassung 1999, AERA, APA & NCME, 1999) hat Sireci (1998) beispielsweise aufgezeigt, wie sich die wissenschaftlichen Vorstellungen über Qualitätsmerkmale der Diagnostik über die Zeit hinweg wandeln.

Häufig werden, insbesondere in der Wissenschaft, empirische Belege genutzt, um bestimmte Beurteilungskriterien zu etablieren. So lassen sich beispielsweise aus den Ergebnissen einer Metaanalyse über die Kriteriumsvalidität von Assessment Centern (AC) Schlussfolgerungen für eine, in einem bestimmten Sinne qualitativ hochwertige Assessment Center Praxis ziehen (siehe Studie im Kasten). Kanning (2004) hat zahlreiche Forschungsergebnisse ausgewertet, um daraus Standards für die Personaldiagnostik abzuleiten. Die Empirie ist ein gutes und wichtiges Argument, aber eben nur eines neben anderen, wie z. B. dem Argument der Nützlichkeit, Schnelligkeit und Praktikabilität von Personalauswahlverfahren (Anderson, Herriot & Hodgkinson, 2001; Muchinsky, 2004). Das Ausmaß, in dem sich bestimmte Auffassungen durchsetzen und über die Zeit hinweg aufrechterhalten werden, hängt nach Gergen (Gergen, 1985, zitiert nach Westmeyer, 2011) nicht so sehr von der empirischen Validität der Auffassungen ab, sondern von den Eventualitäten sozialer Prozesse (z. B. Kommunikation, Verhandlung, Konflikt, Rhetorik). So konnte sich z. B. das Konzept der „emotionalen Intelligenz“ etablieren, obwohl die Argumente zum damaligen Zeitpunkt empirisch nicht fundiert waren (siehe Schuler 2002a, Weber und Westmeyer, 1999).

#### **Erfolgsfaktoren der AC Gestaltung:**

Becker, Höft, Holzenkamp und Spinath (2011) reanalysierten Einzelstudien aus den Jahren 1987 bis 2007 zur Kriteriumsvalidität von Assessment Centern (AC) im deutschsprachigen Raum. Insgesamt konnten 24 Validitätskoeffizienten aus 19 Studien berücksichtigt werden. Die Daten wurden an insgesamt 3.556 Personen erhoben. Zur Analyse wurde die Methode der Validitätsgeneralisierung nach Hunter und Schmidt (2004) genutzt. Die resultierenden Validitätsschätzungen wurden gegen (1) den Stichprobenfehler, (2) gegen die mangelnde Reliabilität im Kriterium sowie (3) gegen direkte Streuungseinschränkungen korrigiert. Der entsprechend bereinigte Validitätskoeffizient für Assessment Center im deutschsprachigen Raum betrug .40. Da sich nur ein unzureichender Anteil der Gesamtvarianz auf die berücksichtigten Artefakte zurückführen ließ, wurden Moderatoranalysen gerechnet. Für die nachfolgenden acht Moderatoren zeigten sich Effekte.

- Art der Kriterienoperationalisierung: Vorgesetztenurteile ließen sich besser vorhersagen (.56) als andere Kriterien, insbesondere das „harte“ Kriterium Verkaufszahlen ließ sich ungleich schlechter (.04) prognostizieren.
- Die Studien erwiesen sich als umso treffsicherer, je mehr Zeit zwischen der Durchführung des AC und der Erhebung der Kriteriendaten lag (.48 zu .22).
- AC, die mit dem Ziel der Potentialanalyse eingesetzt wurden, erzielten höhere Validitäten (.51) als AC, die der Personalentwicklung (.33) oder der Personalauswahl (.26) dienen.
- Bezüglich der Zielgruppen erwiesen sich AC für interne Kandidaten als aussagekräftiger (.44) als für externe Kandidaten (.22).
- Die späteren Leistungen von jüngeren AC-Teilnehmern konnten besser vorhergesagt werden (.50) als die von Lebensälteren.
- Hinsichtlich der inhaltlichen Gestaltung erwiesen sich AC insbesondere dann als aussagekräftig, wenn im Rahmen des AC ein Intelligenztest eingesetzt wurde (.56 zu .25).
- Außerdem konnten hinsichtlich der Kriteriumsvalidität solche AC überzeugen, in denen eine hohe Anzahl von Instrumenten zum Einsatz kam (.53 zu .27).
- Die hohe Anzahl an Instrumenten darf aber nicht mit der Dauer des AC gleichgesetzt werden, im Gegenteil war es so, dass kürzere AC (max. ein Tag, .49) unter Validitätsgesichtspunkten besser abgeschnitten haben als längere AC (.26).

Die Interpretation dieser Befunde wird allerdings dadurch erheblich erschwert, dass mit Ausnahme der Variablen „Art des Kriteriums“ und „zeitlicher Abstand zwischen Prädiktor- und Kriteriumserhebung“ alle Moderatoren interkorreliert waren. Dennoch lässt sich aus der Studie ableiten, dass AC im deutschsprachigen Raum aussagekräftig sind und dass ein kurzes (eintägiges) AC, in dem eine große Anzahl unabhängiger Instrumente zum Einsatz kommen, besonders gute Chancen auf aussagekräftige Ergebnisse bieten. Holzenkamp, Spinath und Höft (2010) empfehlen aufgrund der Studienergebnisse im Rahmen von AC Intelligenztests einzusetzen, wobei die bei Intelligenztest zu befürchtenden Akzeptanzprobleme durch eine berufsbezogene Gestaltung dieser Tests überwunden werden könnten.

## 9.2 Qualitätsstandards

Qualitätsstandards formulieren Regeln und Prinzipien (Soll-Werte), deren Beachtung zur Qualitätssicherung und -optimierung beiträgt. Darüber hinaus erfüllen sie auch (standes-)ethische Funktionen (siehe Schuler, 1991, S. 343; Schuler, 1996, S. 188 ff.). Qualitätsstandards werden in der Regel von Gruppen konstruiert, die über Definitionsmacht verfügen. Die für die Diagnostik und Personalauswahl relevanten Qualitätsstandards wurden von Gruppen konstruiert, die von nationalen bzw. internationalen und transnationalen berufsständischen und wissenschaftlichen Vereinigungen oder von Normungsinstituten mit Definitionsmacht ausgestattet sind. Der folgende Abschnitt geht auf ausgewählte Qualitätsstandards der Perso-

nalauswahl ein. Innerhalb der Standards kann zwischen Richtlinien (vgl. Kapitel 9.2.1), kompetenzorientierten Qualitätsansätzen (vgl. Kapitel 9.2.2) und Beurteilungssystemen für Tests und Fragebogen (vgl. Kapitel 9.2.3) unterschieden werden, obgleich der Übergang zwischen den Kategorien teilweise fließend ist bzw. einige Qualitätsstandards mehreren Kategorien zugeordnet werden könnten. Im Folgenden werden ausgewählte Qualitätsstandards der Personalauswahl überblicksartig erwähnt, bevor die für die Personalauswahl besonders einschlägigen Standards des DIN- und des ISO-Instituts ausführlicher geschildert werden. Eine wesentlich umfangreichere Darstellung von Qualitätsstandards der Diagnostik für alle Anwendungsbereiche der Psychologie (nicht nur Personalauswahl) findet sich bei Kersting, Häcker und Hornke (2011).

### 9.2.1 Allgemeine und verfahrensspezifische Richtlinien

Von zentraler Bedeutung für die Qualität in der Diagnostik im Allgemeinen und somit auch für die Qualität der Diagnostik in der Personalauswahl im Besonderen, sind die von verschiedenen US-amerikanischen Organisationen (American Educational Research Association [AERA], American Psychological Association [APA], National Council on Measurement in Education [NCME]) in Gemeinschaftsarbeit herausgegebenen „Standards für pädagogisches und psychologisches Testen“ (nachfolgend mit „Standards“ bezeichnet), obwohl diese sich – mit Ausnahme des vierzehnten Kapitels – nicht spezifisch auf die Personalauswahl beziehen. Die Standards waren und sind Vorbild verschiedener nationaler und nationenübergreifender sowie fachspezifischer Richtlinienwerke.

Die erste Ausgabe der „Standards“ erschien 1954; seit 1999 liegt die aktuellste, fünfte, überarbeitete Fassung vor (American Educational Research Association [AERA], 1999; Eignor, 2001). Die Standards werden regelmäßig überarbeitet, zum Zeitpunkt der Publikation des vorliegenden Lehrbuchkapitels wurde bereits eine Vorversion der sechsten Fassung veröffentlicht. Von der 1985 publizierten vierten Fassung wurde 1998 eine deutschsprachige Ausgabe publiziert (Häcker, Leutner und Amelang, 1998). Die 264 „Standards“ der aktuellen, fünften Auflage definieren Kriterien für die Anwendung und Evaluation von pädagogischen und psychologischen Tests und Testanwendungen, wobei der Testbegriff sehr weit gefasst ist und teilweise auch Interviews und Simulationsübungen inkludiert. Dabei wird u. a. auf die Gütekriterien und die Fairness diagnostischer Verfahren und Instrumente eingegangen. Um die Personalauswahl geht es insbesondere im vierzehnten Abschnitt der Standards. So findet sich bezüglich der Kriteriumsvalidität beispielsweise ein Hinweis, unter welchen Bedingungen die Ergebnisse von Studien zur Kriteriumsvalidität auf die aktuelle Situation generalisiert werden können (Standard 14.6, vgl. Kasten). Bei 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 Vorgehensweisen und Tests zur Lizenzierung im Kontext beruflicher Qualifikationen, die in verschiedenen Aussagen thematisiert werden. Die Thematik der beruflichen Lizenzierung wird in Europa aufgrund der Freizügigkeit der Berufsausübung in Zukunft eine größere Rolle spielen.

**Beispiel aus den Standards for educational and psychological testing (AERA, 1999, S. 159):**

**Standard 14.6**

Evidence of predictor-criterion relationships in a current local situation should not be inferred from a single previous validation study unless the previous study of the predictor-criterion relationship was done under favourable conditions (i. e., with a large sample size and a relevant criterion) and if the current situation corresponds closely to the previous situation.

Die „*Principles for the validation and use of personnel selection procedures*“ der SIOP (2003) ([http://siop.org/\\_Principles/principles.pdf](http://siop.org/_Principles/principles.pdf)) beziehen sich explizit auf die „Standards“, konzentrieren sich aber ausschließlich auf die Personalauswahl und auf die in diesem Kontext relevanten Themen, wie z. B. Anforderungsorientierung sowie Selektions- und Entscheidungsstrategien.

Im Jahre 1978 wurden die von der Equal Employment Opportunity Commission, den Departments of Labor sowie of Justice und der Civil Service Commission herausgegebenen „*Uniform Guidelines on Employee Selection Procedures*“ (<http://www.uniformguidelines.com>) publiziert. Die Beachtung dieser 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. Jeanneret und Zedeck (2010) weisen darauf hin, dass die Uniform Guidelines in einigen Punkten den „Standards“ sowie den „Principles for the validation and use of personnel selection procedures“ widersprechen und insgesamt fachlich umstritten sind. Dieses Beispiel zeigt, dass es keine Einheitsdefinition von Qualität in der Personalauswahl gibt.

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

Die Internationalisierung des Testmarktes erfordert nach Ansicht der ITC nationalübergreifende Richtlinien für die Testanwendung. Entsprechend hat die ITC die „*International Guidelines for Test Use*“ erarbeitet (ITC, 2000), die auch bereits ins Deutsche übertragen wurden (siehe [www.intestcom.org](http://www.intestcom.org)). Durch die explizite Berücksichtigung der Anwendungskontexte soll es möglich werden, gleichzeitig lokal heterogene Anforderungen und international homogene Standards zu erfüllen. Die Guidelines stellen hohe Anforderungen an den Testanwender und formulieren Best-practice-Beispiele für alle diagnostischen Phasen.

Ebenfalls durch die Internationalisierung bedingt wird es immer häufiger notwendig, bereits vorhandene monolinguale bzw. monokulturelle Testformate an andere Sprach- und Kulturräume zu adaptieren. Dabei sind die – gleichfalls vom ITC erarbeiteten – „*Guidelines for Test Translation and Adaptation*“ zu beachten (siehe z. B. Hambleton, 2001). Die Richtlinien empfehlen die Einbindung erfahrener Übersetzer und die Berücksichtigung der eventuell bestehenden „kulturellen Distanz“, die sich in Bereichen wie Sprache, Familienstruktur, Religion, Lebensstil und Wertekanon manifestieren kann.

Eine weitere Richtlinie der ITC widmet sich der zunehmend an Bedeutung gewinnenden computer- und internetbasierten Testung (siehe auch Bartram & Hambleton, 2006; Naglieri, Drasgow, Schmit, Handler, Prifitera, Margolis & Velasquez, 2004). Die ITC „*International Guidelines on Computer-Based and Internet-Delivered Testing*“ (ITC, 2006) thematisieren u. a. Anforderungen an den Datenschutz, die technischen Voraussetzungen und die Frage der Verantwortlichkeiten für diese Art von Testungen. In diesem Zusammenhang sei auch auf die „*Guidelines for computer-based testing*“ verwiesen, die von der Association of Test Publishers (ATP, 2002) herausgegeben wurden.

Speziell auf die Assessment Center-Methode zugeschnittenen Richtlinien wurden im englischsprachigen Raum von der Task force on „AC Guidelines“ (2009) entwickelt. In der aktuellen fünften Fassung wird auch die angesichts der Globalisierung in der Personalauswahl immer bedeutsamere Entwicklung und der Einsatz von Assessment Centern für/in unterschiedlichen Sprach- und Kulturräumen thematisiert. Für den deutschsprachigen Raum sind die „*Standards der Assessment Center Technik*“ des Arbeitskreis Assessment Center (AKAC 2004, siehe <http://www.arbeitskreis-ac.de>) zu nennen. Die 1992 erstmals veröffentlichten deutschen Standards liegen seit 2004 in der zweiten, überarbeiteten Auflage vor, die sowohl in deutscher als auch in englischer Sprache verfügbar ist. Insgesamt werden neun Standards formuliert, die sich an dem Prozess der Assessment Center-Konstruktion und -Durchführung orientieren. Jeder Standard zielt auf einen Prozessabschnitt, beispielsweise auf die Arbeits- und Anforderungsanalyse, die Übungs-konstruktion oder die Beobachterausswahl und -vorbereitung. Der Inhalt und die Nützlichkeit jedes einzelnen Standards werden erläutert. Dieser Erläuterung folgen Hinweise zur konkreten Umsetzung des Standards. Während die übrigen Richtlinien fast ausschließlich positiv formuliert sind, finden sich in den AKAC-Standards unter der Überschrift „Verstöße“ auch Beispiele für Vorgehensweisen, die dem jeweiligen Standard zuwiderhandeln. Diese Ergänzung um Negativbeispiele kann durchaus sinnvoll sein, da Verhalten durch die Darbietung positiver und negativer Modelle reguliert werden kann.

Die Qualitätsstandards des Arbeitskreis Assessment Center Schweiz entsprechen weitgehend den deutschen Standards ([http://www.akac.ch/files/ac\\_standards\\_akac.pdf](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.

Auch der Vergleich der Assessment Center Qualitätsstandards aus Nordamerika mit den entsprechenden deutschen Qualitätsstandards zeigt, dass es keine Qualität

„an sich“ in der Personalauswahl gibt. Während in den deutschen AC-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, 2009, S. 246). In den deutschen AKAC-Qualitätsstandards werden psychometrische Tests und Interviews als „Nicht-Simulationen“ bezeichnet und tendenziell als zweite Wahl klassifiziert. Dem steht die folgende Aussage der amerikanischen guidelines gegenüber: „Multiple assessment techniques must be used“ (ebd., 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. Teilnehmermerkmale wie Alter, ethnischer und kultureller Hintergrund und Geschlecht werden in den nordamerikanischen guidelines viel deutlicher thematisiert als in den deutschen AC-Qualitätsstandards. Dies lässt sich darauf zurückführen, dass der Integration unterschiedlicher Kulturen und dem Schutz von Minderheiten in Nordamerika eine höhere Bedeutung zukommt als in Deutschland. Das Beispiel zeigt, wie die jeweilige Historie und Kultur die Vorstellungen über Qualität in der Personalauswahl bedingen.

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

Die vom DIN sowie vom ISO Institut herausgegebenen Normen zur Personalauswahl zählen ebenfalls zum Richtlinienansatz, werden aber aufgrund ihrer besonderen Bedeutung im Kapitel 9.3 separat besprochen.

### 9.2.2 Kompetenzorientierte Qualitätsansätze

Qualität in der Personalauswahl erfordert nicht nur hochwertige Verfahren und Vorgehensweisen, sondern auch qualifiziertes Personal. Viele der oben genannten Richtlinien formulieren Anforderungen an die Qualifikation der an der Eignungsbeurteilung beteiligten Personen. Die 1996 gegründete „*Task Force on Test User Qualifications*“ (siehe z. B. Turner, DeMers, Fox & Reed, 2001) kümmert sich vorrangig um die notwendigen Kompetenzen der Anwender. Während die Richtlinien die notwendigen Kompetenzen der beteiligten Personen aber zumeist lediglich beschreiben, thematisiert der kompetenzorientierte Qualitätsansatz auch den Erwerb, die Vermittlung sowie die Überprüfung/Zertifizierung der Kompetenzen und Fertigkeiten. Dabei kann zwischen der universitären und der postgradualen Qualifizierung unterschieden werden. Zur universitären Qualifizierung zählt die im Rahmen des Psychologiestudiums geleistete Ausbildung in psychologischer Diagnostik. In einigen Ländern (z. B. Australien, Kanada) wurde ein „Testschutz“ etabliert, der auf die universitäre Ausbildung vertraut, indem bestimmte Verfahren eigentlich nur von universitär vorgebildeten Personen bezogen werden dürfen. In Deutschland ist die Anwendung von Testverfahren entgegen der weitverbreiteten anderslautenden Meinung hingegen nicht gesetzlich geregelt. Falls 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. Die Frage, ob und in welchem Ausmaß zusätzlich zu den Psychologen Angehörige anderer Berufsgruppen an berufsbezogenen Eignungsbeurteilungen beteiligt werden sollten, ist in Deutschland umstritten. Der Streit darüber 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 (siehe Kersting, 2008, S. 16 f., Schuler, 1996, S. 195; 2003, S. 180).

Weit verbreitet sind *Interviewtrainings* sowie *Beobachtertrainings* für Assessoren im Assessment Center. (Für entsprechende Trainingskonzepte siehe u. a. Kanning (2007) und Schuler (2002b) (beide zu Interviewtrainings) sowie Kapitel 6 für Beobachtertrainings. Zur Effektivität siehe u. a. Conway, Jako und Goodman (1995) für Interviewtrainings und Woehr und Huffcutt (1994) für Beobachtertrainings. Die Trainings sind schwer einzuordnen. Sie sind in der Regel wenig formalisiert und vernachlässigen die Überprüfung und Zertifizierung der vermittelten Kompetenzen und Fertigkeiten. Ähnliches gilt für die zahlreichen *testbezogenen Qualifizierungen*, die teilweise sogar verpflichtende Voraussetzung für den Testwerb und/oder die Testinterpretation sind und bei denen die Vermittlungsmethode in der Regel ebenfalls nicht curricular geregelt ist und häufig keine Trainingserfolgsüberprüfung stattfindet.

Ein Beispiel für formalisierte postgraduale Qualifizierungsangebote geben die Niederlande, indem sich akademisch ausgebildete Psychologen mit psychodiagnostischem Tätigkeitsschwerpunkt von der Psychologenvereinigung NIP (Niederlands Instituut van Psychologen, <http://www.psynip.nl>) registrieren lassen können. Nachzuweisen ist eine universitäre Ausbildung, ein Training in psychologischer Diagnostik sowie drei Fallbearbeitungen unter Supervision eines akkreditierten Psychologen (Evers, 1996). Ein Zertifizierungssystem für Testanwender existiert auch in Großbritannien. Im Gegensatz zu den Niederlanden wird die Zertifizierung hier sowohl für Psychologen, als auch für Angehörige anderer Berufsgruppen angeboten. Der Erwerb des „*Certificate of Competence in Occupational Testing*“, setzt den Besuch spezifischer Qualifizierungskurse voraus, die mit einer Prüfung enden (siehe Bartram, 1996). Die in der Regel fünftägigen Kurse und die Prüfung werden von Personen durchgeführt, die von der British Psychological Society (BPS, <http://www.bps.org.uk>) akkreditiert sind.

Auch die weiter unten beschriebenen Fortbildungen und Lizenzprüfungen zur DIN 33430 zählen zum kompetenzorientierten Qualitätsansatz.

### 9.2.3 Beurteilungssysteme für Tests und Fragebogen

In der Personalauswahl werden Tests und Fragebogen eingesetzt. Die Qualitätsbeurteilung dieser Instrumente wurde früher häufig aus Richtlinien zur Gestaltung von Informationen über Verfahren sowie zur Beurteilung von Verfahren abgeleitet. Ein Beispiel ist der von der Föderation Deutscher Psychologenverbände (1986)

herausgegebene *Kriterienkatalog für die Testbeurteilung*. In Nordamerika wurde eine Vielzahl von Tests nach dem nordamerikanischen BUROS-System besprochen. Das Bueros-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). Wie auch anderen Testbeschreibungsansätzen liegen dem BUROS-System Richtlinien zugrunde, die die zur Information und Beurteilung kritischen Bereiche kennzeichnen und Beurteilungshinweise geben. Im Gegensatz zu den eigentlichen Beurteilungssystemen für Tests und Fragebogen, die im vorliegenden Abschnitt beschrieben werden sollen, fehlen in solchen Richtlinien aber eindeutige Bewertungsvorschriften. Reviews, die den oben genannten Kriterien für die Testbeurteilung folgen, lassen beispielsweise in der Regel eine klare abschließende Beurteilung des besprochenen Tests sowie eine eindeutige Anwendungsempfehlung vermissen. Auch die Vorgabe eines Test-Beschreibungsrasters, wie sie z. B. von dem ZPID (<http://www.zpid.de/index.php?wahl=products&uwahl=printed&uuwahl=testverzeichnis>) oder der ZUMA (ZUMA-Informationssystem (ZIS) (siehe <http://www.gesis.org/methodenberatung/zis/>)) geleistet wird, kann – ungeachtet der damit verbundenen wertvollen Verbesserung der Dokumentation und der Erhöhung der Zugänglichkeit – nicht als Beurteilungssystem im engeren Sinne gewertet werden. Der Übergang von einer Richtlinie zur Gestaltung von Informationen über Verfahren sowie 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.

Einen Grenzbereich stellen beispielsweise die sogenannten „Labels“ (Prüfsiegel, Prüfzeichen, Gütesiegel) der Fachgruppe Diagnostik des Schweizerischen Dienstleistungszentrums Berufsbildung/Berufs-, Studien- und Laufbahnberatung (vormals Diagnostikkommission des Schweizerischen Verbandes für Berufsberatung) dar. Mittels der Labels werden die wesentlichen Aspekte von diagnostischen Instrumenten einem standardisierten Aufbau folgend beschrieben und beurteilt. Die aktuell über 50 vorhandenen Labels können auf der Homepage der Diagnostikkommission eingesehen werden (siehe <http://www.diagnostik.sdbb.ch/labels>).

Ein Beispiel für ein ausgearbeitetes Beurteilungssystem ist das sogenannte *Cotan-System der Niederlande*. Von 1980 bis 2012 wurden 964 Tests einheitlich nach diesem System analysiert und bewertet (Evers, 2012, siehe auch [www.cotandocumentatie.nl](http://www.cotandocumentatie.nl)).

Das Testbeurteilungssystem der European Federation of Psychologists Associations (EFPA, siehe Lindley & Bartram, 2012) weist deutliche Ähnlichkeiten mit dem niederländischen System auf. Das „*EFPA Review Model for the Description and Evaluation of Psychological Tests*“ wird u. a. im Internet dargestellt (<http://www.efpa.be>). Das System sieht einerseits ein Formblatt zur Beschreibung und Bewertung der Verfahren vor und bietet andererseits den Reviewern ausführliche

ippen verkauft wer-  
Regelung der priva-  
zu den Psychologen  
nungsbeurteilungen  
it darüber lässt sich  
ann [1966, S. 263]  
is dem Jahre 1927)  
2008, S. 16f., Schu-

inings für Assesso-  
ypte siehe u. a. Kan-  
sowie Kapitel 6 für  
ako und Goodman  
für Beobachtertrai-  
Regel wenig forma-  
ng der vermittelten  
chen *testbezogenen*  
nung für den Tester-  
mittlungsmethode  
keine Trainingser-

angebote geben die  
en mit psychodiag-  
igung NIP (Neder-  
eren lassen können.  
in psychologischer  
ines akkreditierten  
tanwender existiert  
wird die Zertifizie-  
nderer Berufsgrup-  
*Occupational Tes-*  
raus, die mit einer  
igen Kurse und die  
Psychological So-

zenzprüfungen zur  
z.

etzt. Die Qualitäts-  
rien zur Gestaltung  
Verfahren abgelei-  
genverbände (1986)

Beurteilungshinweise, die eine einheitliche Bewertung sichern soll. Für alle Verfahrensmerkmale, die eine Bewertung erfahren, ist eine vierstufige Skala (von „inadequate“ to „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 allerdings dazu, dass von einer Anwendung des Verfahrens abgeraten wird. Für die einzelnen Bewertungsstufen gibt es eindeutige Vorgaben, beispielsweise werden die Kennwerte für die interne Konsistenz entsprechend ihrer Höhe klassifiziert, wobei die Einordnung aber je nach Anwendungszweck (Entscheidungen über Individuen oder Gruppen) strenger oder milder ausfällt. Das Review endet mit einer eindeutigen Bewertung des Verfahrens und einer Empfehlung zu seiner Anwendung. Obwohl das Gesamturteil sich nicht arithmetisch aus den Einzelurteilen herleitet, sollen definierte Mindestanforderungen sicherstellen, dass problematischen Verfahren keine Empfehlung ausgesprochen wird.

Das System der Testdokumentation und -bewertung der British Psychological Society baut auf dem EFPA Modell auf, unter <http://www.psychtesting.org.uk> können entsprechende Reviews (in der Kurzform) eingesehen werden.

Kersting (2006) hat eine Bestandsaufnahme zur Situation der Information über und der Beurteilung von Tests sowie Vorschläge für eine verbesserte Praxis vorgelegt. Das Diagnostik- und Testkuratorium (DTK) der Föderation der Deutschen Psychologinnenvereinigungen hat daraufhin ein System zur Information über und Beurteilung von Tests entwickelt (Testbeurteilungssystem des Testkuratoriums (TBS-TK, Testkuratorium 2010), siehe <http://www.zpid.de/Testkuratorium>), welches die DIN 33430 (siehe unten) berücksichtigt. Das TBS-TK ersetzt die oben genannten Kriterien für die Testbeurteilung der Föderation Deutscher Psychologinnenverbände (1986).

Der Beurteilungsprozess gemäß TBS-TK verläuft vereinfacht dargestellt in zwei Schritten: Der erste Schritt umfasst eine Prüfung der Verfahrenshinweise (i. d. R. des Testmanuals) auf die Erfüllung der in der DIN 33430 (siehe unten) formulierten Anforderungen an die Vollständigkeit der in den Verfahrenshinweisen (Testmanuals) aufgeführten Informationen. Die Operationalisierung dieser Anforderungen erfolgt mit der „Checkliste 1“ der Publikation „DIN Screen“ von Kersting (2008) (Download: [http://www.kersting-internet.de/DIN-Buch/din-buch\\_downloads.html](http://www.kersting-internet.de/DIN-Buch/din-buch_downloads.html)). Diese Checkliste gilt offiziell als „Standard zur Information und Dokumentation von Instrumenten zur Erfassung menschlichen Erlebens und Verhaltens“ des Diagnostik- und Testkuratoriums der Föderation Deutscher Psychologinnenvereinigungen. Obwohl die Qualitätsforderungen an die Verfahrenshinweise aus der DIN 33430 übernommen wurden, gelten sie – wie das gesamte TBS-TK Testbeurteilungssystem – nicht nur für die Eignungsdiagnostik, sondern für Tests aus allen Anwendungsbereichen (z. B. auch der klinischen oder pädagogischen Psychologie). 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.

Der zweite Schritt beinhaltet die eigentliche Bewertung des Tests anhand der Besprechungs- und Beurteilungskategorien des TBS-TK durch die Rezensent/-innen. 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. Die Beurteilung gliedert sich in sieben „Besprechungs- und Beurteilungskategorien“ und in eine zusammenfassende Abschlussbewertung, die in ein vorbereitetes Formular-Dokument eingetragen werden. Für jede Kategorie ist eine frei formulierte Bewertung vorgesehen. Für vier der sieben Beurteilungskategorien erfolgt darüber hinaus eine formalisierte Bewertung auf einer vierstufigen Skala. („Der Test erfüllt die Anforderungen (...) (a) ‚voll‘, (b) ‚weitgehend‘, (c) ‚teilweise‘ und (d) ‚nicht‘“). Darüber hinaus soll eine Abschlussbewertung/Empfehlung formuliert werden.

Die nach diesem Muster erstellten Rezensionen werden frei zugänglich veröffentlicht (siehe <http://www.zpid.de/Testkuratorium>). Tabelle 9.1 zeigt beispielhaft die Bewertungen zum „Golden Profiler of Personality“, eine erweiterte Form des Myers-Briggs Typen Indikator (MBTI), einem in der Personalauswahl häufig eingesetzten Persönlichkeitsfragebogen. Bezüglich der Gültigkeit (Validität) erfüllt der Test die Anforderungen nach Ansicht der Rezensenten nicht, da für diesen Test empirische Studien zur Konstrukt- und Kriteriumsvalidität fehlen.

**Tabelle 9.1:** Rezension des Golden Profiler of Personality (GPOP). Deutsche Adaptation des Golden Personality Type Profiler (Rezension von Höft & Muck, 2009)

TBS-TK Rezension	Golden Profiler of Personality (GPOP)	Die TBS-TK-Anforderungen sind erfüllt.			
		voll	weitgehend	teilweise	nicht
Testbeurteilungssystem des Testkuratoriums der Föderation deutscher Psychologenvereinigungen	Allgemeine Informationen, Beschreibung und diagnostische Zielsetzung			•	
	Objektivität			•	
	Zuverlässigkeit			•	
	Validität				•

Bislang liegen nur sehr wenige Testrezensionen nach dem neuen TBS-TK-System vor und selbst in vielen Jahren wird es nicht so sein, dass man zu jedem Test eine unabhängige Begutachtung einsehen kann. Dazu gibt es zu viele Tests, allein in Deutschland sind es mehrere Tausend Verfahren. Ein Blick in die TBS-TK-Beurteilungsrichtlinie und in beispielhafte Rezensionen lohnt dennoch, da dadurch das Prinzip der Testbewertung verständlich und für den eigenen Anwendungsfall analogisierbar wird. Vor allem die Checkliste mit den Fragen an Testanbieter kann in

Bezug auf jeden Test genutzt werden, um zwischen seriösen und unseriösen Anbietern zu differenzieren. Seriöse Testanbieter stellen die laut Checkliste geforderten Informationen in schriftlicher Form zur Verfügung. Sie weisen darüber hinaus auf die Grenzen der Verfahren hin.

Anders als in den oben dargestellten COTAN- und EFPA-Systemen sieht das TBS-TK-System keine Bewertung der Tests und Fragebogen allein aufgrund der numerischen Ausprägung der Testgütekriterien vor. 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 (siehe z. B. Schmidt, 1992 für ein Beispiel zur Schätzung der Kriteriumsvalidität). Eine ausführliche Darstellung der Argumente findet sich bei Kersting (2006).

Mit dem TBS-TK-System sollen Tests nicht nur im Nachhinein beurteilt werden, sondern das System soll bereits bei der Testentwicklung Berücksichtigung finden. Tatsächlich wirkt das TBS-TK-System kulturprägend. Dies kommt etwa darin zum Ausdruck, dass in den Verfahrenshinweisen (Manualen) von Neupublikationen von Tests mit einem hohen Qualitätsanspruch standardmäßig bereits Übersichten darüber vorgehalten werden, auf welcher Seite des Testmanuals die nach TBS-TK geforderten Informationen zu finden sind (Beispiele hierfür sind die Verfahrenshinweise zum BIP-6F, d2-R, FAKT, START-P und WIT-2).

### 9.3 Normen zur Personalauswahl

Gleichwohl die DIN 33430 zur Personalauswahl sowie die ISO-Norm zur Eignungsdiagnostik ebenfalls Qualitätsstandards (und Richtlinien) darstellen, sollen sie hier aufgrund ihrer besonderen Bedeutung gesondert thematisiert werden. Der Normansatz unterscheidet sich von den übrigen Qualitätsstandards vor allem dadurch, dass diese Normen unter dem Dach eines etablierten Normungsinstitutes außerhalb der psychologischen Berufsverbände formuliert wurde. Die DIN 33430 und die ISO 10667-1/ISO 10667-2 sind arbeitsfeldbezogen und nicht berufsgruppenorientiert. Dies sollte den Wirkungsgrad des Qualitätsstandards erhöhen. Während sich Nicht-Psychologen von Berufsstandesregeln für Psychologen nicht angesprochen fühlen, setzen DIN und ISO allen Verantwortlichen – ungeachtet ihrer Berufsgruppenzugehörigkeit – Qualitätsstandards. Einer Norm wohnt bei der Formulierung von Qualitätsstandards eine autoritative Wirkung inne, die eine berufsständisch formulierte Richtlinie nicht erlangen kann. Hinzu kommt eine gewisse rechtliche Bedeutung. Normen sind zwar keine Gesetze, indem sie über den aktuellen Stand von Wissenschaft und Praxis informieren, können sie aber gleichwohl Bedeutung bei einer richterlichen Entscheidungsfindung erlangen.

Normen finden bei Gericht insofern Beachtung, um den herrschenden Standard in einem Gebiet zu ermitteln. Die DIN 33430 kann in Deutschland 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 siehe u. a. Kapitel 28, Abeln & Reimann, 2004 sowie Wegener, 2003). Notwendige, umfassende Aufklärungen und Vertragsvereinbarungen können durch den Hinweis auf die DIN 33430 und/oder die ISO 10667-1/ISO 10667-2 verkürzt werden. Die Normen ermöglichen eine Bewertung externer und interner Anbieter von Personaldienstleistungen.

Nicht nur in juristischer, sondern auch in marktwirtschaftlicher Hinsicht kann die Wirkung einer Norm weit über die Wirkung der bisherigen berufsgruppenintern 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- und ISO-Ansatzes der Qualitätssicherung in der Personalauswahl ist die Möglichkeit von Ist-Soll-Vergleichen: Während mit anderen Qualitätsstandards oft lediglich Soll-Werte formuliert wurden, ist die DIN 33430 Grundlage von Personenlizenzen sowie von Organisationszertifizierungen. Die Personenlizenzierung wird weiter unten erläutert. Inhaber von Personenlizenzen sind – laut dem DIN 33430 Portal ([www.din33430portal.de](http://www.din33430portal.de); Zugriff 26. Mai 2012) – beispielsweise Mitarbeiter der Personalabteilungen der Deutschen Telekom, der E.ON, der ÖBB oder der Merck AG oder der Beratungsunternehmen CNT, DGP, Geno, Gideon, IfP, Kienbaum, KIM und Obermann Consulting. Bei der Organisationszertifizierung wird die Übereinstimmung des in einer Organisation geübten Qualitätsmanagementsystems mit der Norm durch ein Zertifikat bestätigt. Den Hintergrund für diese Zertifizierung liefert die ISO 9000-Reihe, die Prüfung auf Konformität erfolgt durch eine externe Auditierung. Eine solche Organisationszertifizierung erfolgt durch private Unternehmen, die sich bei einem Ausschuss, der von mit Qualitätssicherung befassten Verbänden und Gesellschaften gebildet wird, akkreditieren lassen können. Zertifiziert nach DIN 33430 sind beispielsweise das Berufsförderungswerk Oberhausen, der Bildungsanbieter „JobColleg“, eine Abteilung der Commerzbank AG, das Personalberatungsunternehmen „Mercuri Urval“ und die Vivento Customer Services GmbH (VCS). Über die vom TÜV durchgeführte Prozesszertifizierung eines Personalauswahlverfahrens für Führungskräfte der Commerzbank AG berichten Dries, Krumm und Lux (2006). Für den gesamten Prozess vom Erstgespräch bis zur Übergabe des Zertifikats wurden neun Monaten benötigt, der finanzielle Aufwand für den Zertifizierer sowie ein beauftragtes Beratungsunternehmen wird in der Publikation mit 10.000 Euro angegeben.

Alternativ zur Personenlizenzierung oder zur Organisationszertifizierung kann man auch eine „Selbsterklärung“ abgeben und verkünden, dass man seine Personalauswahl-Dienste nach den Regeln der DIN oder ISO Personalauswahl-Normen anbietet. Entsprechende Selbsterklärungen finden sich mittlerweile auf den Web-

seiten zahlreicher Personalberatungsunternehmen. So erklärt beispielsweise Frau Dr. Joder auf ihrer Website <http://www.karinjoder.de> (Zugriff: 26. Mai 2012) in werblicher Absicht: „Meine diagnostischen Dienstleistungen sind qualitätsgesichert nach DIN 33430“. Eine nicht wahrheitsgemäße Selbsterklärung kann im Rahmen des Wettbewerbsrechts angegangen werden.

### 9.3.1 Die DIN 33430

Im Juni 2002 wurde die DIN 33430 verabschiedet, sie formuliert „Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen“. Die 15 Seiten Text und sieben Seiten Glossar umfassende Norm ist u. a. bei Kanning (2004) sowie bei Westhoff, Hagemeyer, Kersting, Lang, Moosbrugger, Reimann und Stemmler (2010) im vollständigen Wortlaut abgedruckt. Checklisten zur DIN 33430 (Kersting, 2008, vgl. Abb. 9.1) ermöglichen ein rasches Screening, ob die Prozesse den Anforderungen der DIN 33430 genügen.

#### DIN 33430:

Gegenstand der DIN 33430 ist der gesamte Prozess der Eignungsbeurteilung, der als eine Einheit angesehen wird; die DIN 33430 ist eine Prozess- und keine Produktnorm.

<i>Checkliste 4.2: Auswertung, Interpretation und Urteilsbildung</i>		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"/>		
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"/>		
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"/>		

Abbildung 9.1: Auszug aus der Checkliste zur DIN 33430 (Kersting, 2008)

Obwohl die DIN 33430 eine Prozessnorm ist, brüsten sich verschiedene Testanbieter damit, dass ihre Tests den Standards der DIN 33430 genügen würden – und demonstrieren damit ihre Unkenntnis. Die DIN 33430 formuliert lediglich Anforderungen an Verfahrenshinweise (Testmanuale), dieser Ausschnitt der DIN 33430 ist in das Testbeurteilungssystem des Diagnostik- und Testkuratoriums aufgenommen worden (siehe oben). Man kann nicht sinnvoll formulieren, dass ein Test der DIN 33430 entspricht, gleichwohl kann man aber konstatieren, dass ein Prozess der Eignungsbeurteilung nicht der DIN 33430 entspricht, weil bereits ein Prozesselement (die Anforderungen an die Verfahrenshinweise des eingesetzten Testes) nicht den DIN 33430-Anforderungen genügt.

Die DIN 33430 sieht ausdrücklich keine isolierte Betrachtung der Qualität von Verfahren (wie z. B. Tests) vor, sondern umfasst Festlegungen und Leitsätze für den Einsatz dieser Verfahren bei berufsbezogenen Eignungsbeurteilungen (DIN, 2002, S. 3 f.). Sie bezieht sich auf (1) die Planung von berufsbezogenen Eignungsbeurteilungen, (2) die Auswahl, Zusammenstellung, Durchführung und Auswertung von Verfahren sowie (3) auf die Interpretation der Verfahrensergebnisse und die Urteilsbildung. Außerdem formuliert sie (4) Anforderungen an die Qualifikation der an der Eignungsbeurteilung beteiligten Personen. Für die genannten Bereiche werden Standards der zur Zielerreichung (Eignungsbeurteilung) erforderlichen Sorgfalt des Verhaltens und der Kenntniserwerbformuliert, wobei der jeweilige Auftrag und die Rahmenbedingungen simultan zu betrachten sind.

Eine berufliche Eignungsbeurteilung nach DIN 33430 setzt eine sorgfältig durchgeführte Arbeits- und Anforderungsanalyse der Ausbildung, des Berufs oder der beruflichen Tätigkeit voraus. „Für die berufsbezogene Eignungsbeurteilung dürfen nur solche Verfahren eingeplant werden, die nachweislich einen Bezug zu den Anforderungen haben.“ (DIN, 2002, S. 6). Jeder Anforderungsdimension, die erhoben werden soll, muss mindestens ein Verfahren zugeordnet werden, mit dem diese Dimension erfasst wird. Und jedem der eingesetzten Verfahren muss mindestens eine Anforderungsdimension zugeordnet werden, zu deren Erfassung es eingesetzt wird. Der zuletzt genannte Aspekt wird häufig vernachlässigt. Ein Persönlichkeitsfragebogen, der erfasst, ob die befragte Person eher „gelb“, „grün“, „blau“ oder „rot“ ist, hat – ungeachtet der Frage seiner psychometrischen Güte – in einem Personalauswahlverfahren nichts zu suchen, wenn nicht vorab diese „Farben“ als Anforderungsdimensionen definiert wurden. Außerdem müssen diese Verfahren für die jeweilige Eignungsbeurteilung qualifiziert sein. Qualitätskriterien sind Objektivität, Zuverlässigkeit (Reliabilität) und Gültigkeit (Validität). Sofern für die Eignungsbeurteilung Verfahren eingesetzt werden, die den Vergleich mit Normwerten (im Sinne von Referenzkennwerten) vorsehen, kommt der Qualität der Normwerte ebenfalls Bedeutung zu. Sie müssen der Fragestellung und der Referenzgruppe der Kandidaten entsprechen (DIN, 2002, S. 8). Die Angemessenheit der Normwerte sowie der Kennwerte für die Zuverlässigkeit und Gültigkeit eines Verfahrens müssen spätestens alle acht Jahre überprüft werden.

Die DIN schreibt vor, dass für jedes Verfahren (also nicht nur für Tests, sondern z. B. auch für Eignungsinterviews oder Assessment Center) sogenannte Verfahrenshinweise vorliegen, und stellt – wie oben bereits erläutert – in einem umfang-

und detailreichen normativen Anhang hohe Anforderungen an die Gestaltung dieser Verfahrenshinweise.

Eine sorgfältige Planung der berufsbezogenen Eignungsbeurteilung setzt nach DIN 33430 voraus, dass alle Aspekte der Durchführung und Auswertung der Verfahren zur Eignungsbeurteilung sowie zur Interpretation der Ergebnisse vorab festgelegt und dokumentiert werden. Bei wiederholten Auswahlprogrammen sind diese Regeln spätestens alle drei Jahre zu überprüfen. Die Vorschriften der DIN 33430 gehen bis ins Detail, beispielsweise wird für Eignungsinterviews und Assessment Center verlangt, dass den Beurteilungskategorien vorab Beispielaussagen und Beispielverhaltensweisen der Kandidaten zugeordnet werden, bei schriftlichen Tests muss vorab geklärt werden, wie bei der Auswertung mit nicht bearbeiteten Items u. Ä. umgegangen wird usw. Der gesamte Prozess der Eignungsbeurteilung ist so zu dokumentieren, dass die Eignungsbeurteilung nachvollzogen werden kann.

Eine Eignungsbeurteilung nach DIN 33430 setzt nicht nur qualifizierte Verfahren, sondern auch kompetente Anwender für die Auswahl, Durchführung, Auswertung und Interpretation der Verfahren voraus. Die DIN 33430 unterscheidet zwischen dem Auftragnehmer einerseits und dem sogenannten Mitwirkenden andererseits. Als Auftragnehmer wird eine (organisationsinterne oder -externe) Person bezeichnet, die sich verpflichtet, für einen Auftraggeber eine berufsbezogene Eignungsbeurteilung im Sinne der DIN 33430 durchzuführen. Mitwirkende sind an der Eignungsbeurteilung beteiligt, z. B. als Assessoren im Assessment Center oder als Mitwirkende an Eignungsinterviews. Die Mitwirkenden stehen unter Anleitung, Fachaufsicht und Verantwortung des Auftragnehmers. Der Umfang der in der DIN 33430 geforderten Kenntnisse und angeleiteten Praxiserfahrungen variiert in Abhängigkeit von der Funktion, welche die Person im Rahmen der Eignungsbeurteilung wahrnimmt. Dabei werden drei, aufeinander aufbauende Funktionen unterschieden (vgl. Abb. 9.2). Personen, die als Auftragnehmer oder Mitwirkender an der Durchführung und Auswertung von Verhaltensbeobachtungen und -beurteilungen beteiligt sind, benötigen (1) Kenntnisse über Rahmenbedingungen von Verfahren zur mündlichen Informationsgewinnung, über einschlägige Evaluationen sowie über spezifizierte Bereiche aus dem Themenkontext Beobachtung und Beurteilung (z. B. Beobachtungseinheiten, Ratingverfahren, Beobachtungsfehler usw.). Ist man als Auftragnehmer oder Mitwirkender an Eignungsinterviews beteiligt, benötigt man zusätzlich (2) Kenntnisse zu den Themen (2a) Interviewklassifikationen, (2b) Handhabung von Interviewleitfäden, (2c) Frage- und Formulierungstechniken, (2d) interviewbezogene Beurteilungskriterien sowie (2e) Fragebereiche und ihre rechtliche Zulässigkeit. Der Auftragnehmer muss über die genannten Kenntnisse hinaus noch weitere Qualitätsanforderungen erfüllen, im Einzelnen werden u. a. (3) Kenntnisse zu (3a) Anforderungsanalysen (z. B. Methoden), (3b) Verfahren (z. B. Testtheorien, Gütekriterien) und (3c) zu Eignungsbeurteilungen (z. B. Geltungsbereiche) erwartet (vgl. Abb. 9.2).

Mit den sogenannten „Leitsätzen“, die keinen normativen Charakter haben, stellt die DIN 33430 weitere Hinweise für die Vorgehensweise bei berufsbezogenen Eignungsbeurteilungen zur Verfügung. Für einzelne Bereiche wie z. B. die

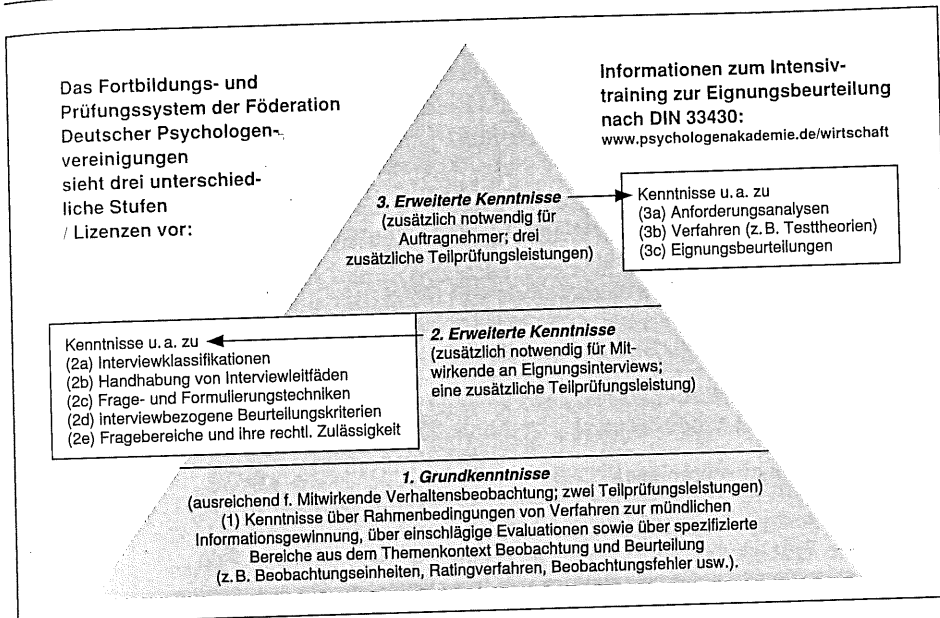


Abbildung 9.2: Qualitätsanforderungen der DIN 33430 an Auftragnehmer und Mitwirkende

Anforderungsanalyse, die Vorauswahl und die Untersuchungssituation werden Vorgehensweisen beschrieben, die als best practice gelten können. Die DIN 33430 empfiehlt beispielsweise, die Kandidaten vorab über den Arbeitsplatz und über den Ablauf der Eignungsuntersuchung zu informieren.

Es wird häufig darüber geklagt, dass Eignungsbeurteilungen von mangelhaft qualifizierten Personen durchgeführt werden. Durch Richtlinien, in denen die erforderlichen Kompetenzen der Anwender beschrieben werden sowie durch den Qualifizierungsansatz, sollte hier Abhilfe geschaffen werden (siehe oben). Mit der DIN 33430 sind die Qualitätsanforderungen an den Auftragnehmer und an die Mitwirkenden von berufsbezogenen Eignungsbeurteilungen spezifiziert. Auf dieser Basis können Fortbildungen und Prüfungen angeboten werden. Die Föderation Deutscher Psychologenvereinigungen hat 2004 eine Fortbildungs- und Prüfungsordnung zur Personenlizenzierung für berufsbezogene Eignungsbeurteilungen nach DIN 33430 erlassen, die 2011 geringfügig überarbeitet wurde ([http://www.din33430portal.de/lizenzpruefungsordnung\\_din\\_33430\\_rev24jun11.pdf](http://www.din33430portal.de/lizenzpruefungsordnung_din_33430_rev24jun11.pdf)). Auftragnehmer und Mitwirkende können Fortbildungen besuchen, die in einer engen Theorie- und Praxisverzahnung die geforderten Kenntnisse und Fertigkeiten (Qualitätsanforderungen) handlungsorientiert vermitteln. Zielgruppe sind alle Personen, die im Rahmen der Personalauswahl Aufgaben übernehmen, ungeachtet ihrer ursprünglichen Berufsqualifikation. Teilnehmer der Seminare sind häufig Personalberater und Mitarbeiter der Personalabteilungen von großen Organisationen, Beispiele sind Firmen wie AVIS, Deutsche Bank, Deutsche Telekom, MAN, T-Systems, Vivento usw. Für die Fortbildung sind sechs ein-

bzw. zweitägige Module vorgesehen. Die Fortbildung für Mitwirkende an Verhaltensbeobachtungen (MV) besteht aus den Modulen 1 und 2, für Mitwirkende an Eignungsinterviews (ME) kommt Modul 3 hinzu. Die Fortbildung für Auftragnehmer (A) besteht aus den Modulen 1 bis 6.

**Tabelle 9.2:** Anzahl und thematische Inhalte von Modulen zur Fortbildung nach DIN 33430 (Föderation Deutscher Psychologinnenvereinigungen, 2011)

Modul	Inhalt	Zielgruppe <sup>1</sup>	Tag(e)
1	Einführung in die DIN-Norm 33430	A, MV, ME	1
2	Verhaltensbeobachtung und Verhaltensbeurteilung	A, MV, ME	2
3	Eignungsinterviews	A, ME	2
4	Anforderungsanalyse, Konstrukte und Prozeduren der Eignungsbeurteilung	A	2
5	Psychometrische Grundlagen der Eignungsbeurteilung	A	1
6	Evaluation der Eignungsbeurteilung	A	2

*Anmerkung:* <sup>1</sup>A=Auftragnehmer, MV=Mitwirkende Verhaltensbeobachtung, ME Mitwirkende an Eignungsinterviews

Die für die Module definierten Kenntnisse und Fertigkeiten bilden die Grundlage zum Erwerb der Personenlizenz für die berufsbezogene Eignungsbeurteilung nach DIN 33430. Lizenzen werden nach bestandener Lizenzprüfung ausgestellt. Es können drei Lizenzen erworben werden: (1) Lizenz A für Auftragnehmer, (2) Lizenz MV für Mitwirkende an Verhaltensbeobachtungen und (3) Lizenz ME für Mitwirkende an Eignungsinterviews. Die Zulassung zur Lizenzprüfung ist nicht an bestimmte Ausbildungsabschlüsse gebunden, der vorherige Besuch von Fortbildungsveranstaltungen ist nicht vorgeschrieben. Allerdings müssen Personen, die eine Lizenz A anstreben, über angeleitete Praxiserfahrungen verfügen. Für den erstmaligen Erwerb einer Lizenz müssen mehrere Teilprüfungsleistungen erbracht werden. Die gebührenpflichtige Prüfung besteht jeweils aus einem unter Aufsicht zu bearbeitenden Multiple Choice-Test. Die Teilprüfungsleistungen entsprechen den in Tabelle 9.2 gelisteten Ausbildungsmodulen und Inhalten (vgl. Abb. 9.2): (1) Lizenz MV: die Inhalte der Module zur Fortbildung 1 und 2 (zwei Teilprüfungsleistungen), (2) Lizenz ME: die Inhalte der Module zur Fortbildung 1 bis 3 (drei Teilprüfungsleistungen) und (3) Lizenz A: die Inhalte der Module zur Fortbildung 1 bis 6 (sechs Teilprüfungsleistungen). Grundlage der Prüfungsfragen ist das vom Diagnostik- und Testkuratorium herausgegebene Buch „Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33430“ (Westhoff et al., 2010). Eine ausführliche Darstellung der Fortbildung sowie der Lizenzprüfung findet sich bei Kersting (2010b).

### 9.3.2 Die ISO-Normen 10667-1 und 10667-2

Die Norm „Assessment Service Delivery: Procedures and methods to assess people in work and organizational settings“ wurde 2011 von der internationalen Organisation für Normung (ISO) veröffentlicht. Die Normen formulieren Anforderungen und Leitlinien für Verfahren und Methoden der berufsbezogenen Eignungsbeurteilung von Personen, Gruppen und Organisationen. Die Norm umfasst zwei Teile. Teil 1, die ISO 10667-1 (ISO, 2011a), richtet sich an den Auftraggeber (client), Teil 2, die ISO 10667-2 (ISO, 2011b), richtet sich an den Auftragnehmer (service provider). Im Folgenden wird für die beiden Normen zusammen der Begriff ISO 10667 genutzt, gleichwohl es streng genommen nur die ISO 10667-1 und die ISO 10667-2 gibt. Bei der Erarbeitung dieser ISO-Normen wurden verschiedene nationale und transnationale Qualitätsstandards, u. a. auch die DIN 33430 (siehe oben) zu Beratungszwecken herangezogen.

Die ISO 10667 betont und formalisiert deutlicher als alle anderen Qualitätsstandards die Kommunikation zwischen den an der Eignungsdiagnostik beteiligten Personen und Organisationen. Damit unterstreicht die Norm, dass die Definition von Qualität ein soziales Produkt ist und die Durchsetzung von Qualität eine Frage von Verhandlungen. Entsprechend regelt die Norm vor allem das Verständigungsverfahren (aggreement procedure) zwischen Auftragnehmer und (internem oder externem) Auftraggeber. Im Rahmen des Verständigungsverfahrens sollen Auftraggeber und Auftragnehmer die erwünschten Ergebnisse der Eignungsbeurteilung antizipieren und die Risiken für unerwünschte Nebenwirkungen realisieren. Innovativ an der ISO 10667 ist, dass sie nicht nur Anforderungen an den Auftragnehmer formuliert, sondern auch an den Auftraggeber (und sogar an den Teilnehmer/die diagnostizierte Person). Die ISO 10667 geht im Allgemeinen und hinsichtlich der Gestaltung eines Dienstleistungsvertrages im Besonderen nicht nur auf die Leistungen ein, die der Auftragnehmer zu erbringen hat, sondern auch auf die Leistungen, die der Auftraggeber einer gelungenen Eignungsuntersuchung schuldet. Beispielsweise wird der Auftraggeber auf eine evidenzbasierte sowie rechtlich einwandfreie Vorgehensweise verpflichtet. Er soll darüber hinaus darüber nachdenken, ob und wie er zusätzliche Forschungsaktivitäten wie z. B. Normierungs- und Validitätsstudien unterstützen kann und dieses Engagement, sofern es geplant ist, im Dienstleistungsvertrag dokumentieren.

Der Abschnitt über das Verständigungsverfahren ist zwar nominell der dritte Abschnitt, das Vorwort, die durchaus umfassende Einleitung sowie die Abschnitte 1 (Anwendungsbereich) und 2 (Begriffe) sind aber, ebenso wie das Literaturverzeichnis am Ende, eher formeller Natur, so dass die Ausführungen zur Auftragsklärung in Abschnitt 3 den eigentlichen Auftakt der Norm darstellen.

Im vierten Abschnitt „pre-assessment procedures“ geht es um die Ziele und Zielgruppen der Eignungsdiagnostik. Die Norm fordert, dass der Ablauf und Inhalt der Eignungsdiagnostik in einem „assessment service agreement“ festgehalten werden. Hierzu zählen, neben Fragen der eingesetzten Instrumente, auch Regeln zum Feedback und eine Kostenaufstellung. Die Norm empfiehlt, bereits im Rahmen dieser Absprache ein „post-assessment review“ zu vereinbaren (siehe dazu weiter unten).

Der fünfte Abschnitt thematisiert die eigentliche Eignungsbeurteilung in verschiedenen Phasen von der Planung bis zum Ergebnisbericht. Gefordert wird u. a. das Vorliegen eines detaillierten Plans für die Durchführung der Eignungsbeurteilung. Viel Wert wird außerdem auf den Schutz von Testmaterialien, aber auch von Bewertungsregeln sowie von persönlichen Daten gelegt. Ganz im Sinne der Kommunikation und Verständigung betont die ISO 10667 die Notwendigkeit einer umfassenden Information aller Beteiligten und Betroffenen, insbesondere des Teilnehmers. Dieser soll nach ausführlicher Information und Aufklärung sein Einverständnis mit dem eignungsdiagnostischen Prozess erklären (informed consent). Hinsichtlich der eigentlichen Durchführung der Eignungsbeurteilung wird u. a. die Notwendigkeit des Einsatzes qualifizierter Kräfte betont. Ein weiteres Thema ist die Qualität des Feedbacks. Der fünfte Abschnitt endet mit Vorschlägen zu einer ersten Evaluation des Prozesses, die am Ende der Eignungsbeurteilung stehen sollte.

Empfohlen wird, wie bereits oben angeführt, dass Auftraggeber und Auftragnehmer sich auf die Durchführung eines „post-assessment reviews“ einigen. Die Inhalte dieses Rückblicks auf die durchgeführte Eignungsbeurteilung werden im sechsten Abschnitt beschrieben. Dabei geht es vor allem um die Fragen, ob der gewünschte Zielzustand erreicht wurde, welche Ursachen eine etwaige Abweichung vom Zielzustand haben könnte sowie um Überlegungen, wie der Prozess weiter verbessert werden kann.

Vier Anhänge runden die ISO 10667 ab. Diese sind „informativer“ Art, was in der Sprache des Normungswesens bedeutet, dass die Aussagen in den so gekennzeichneten Anhängen keinen normativen Charakter haben. Im Anhang A geht es um die Teilnehmer. Dem Teilnehmer werden nicht nur Rechte eingeräumt, sondern auch Pflichten auferlegt, z. B. die Pflicht, sich zu informieren, den Instruktionen zu folgen usw. Anhang B widmet sich der Dokumentation, wobei es nicht nur um die Dokumentation der genutzten Prozeduren und Instrumente geht, sondern auch um die Dokumentation der Erkenntnisse, die im Sinne eines evidenzbasierten Vorgehens die Entscheidung für die gewählte eignungsdiagnostische Vorgehensweise untermauern. Auch die Forderung nach einer Dokumentation zu den Gütekriterien Objektivität, Reliabilität und Validität sowie der Fairness finden sich im Anhang B und sind somit nicht normativ. Abschließend wird gefordert, die Sinnhaftigkeit der gewählten eignungsdiagnostischen Vorgehensweisen regelmäßig zu überprüfen. Die Analyse und Interpretation der Ergebnisse der Eignungsdiagnostik sind Gegenstand des Anhangs C. Zum einen werden Anforderungen an die Qualifikation derjenigen Personen gestellt, welche die Ergebnisse analysieren und interpretieren. Zum anderen werden u. a. Fragen der Datenkombination und der Entscheidungsregeln thematisiert. Der letzte Anhang, D, stellt Anforderungen an die Ergebnisberichte, diese sollen beispielsweise zum diagnostischen Ziel (z. B. Auswahl vs. Entwicklung) „passen“ und auch die Grenzen der Aussagekraft explizieren.

Die beiden Normtexte ISO 10667-1 und ISO 10667-2 sind thematisch gleich und folgen derselben Gliederung, Part 1 behandelt die Anforderungen an den Auftraggeber von Personaldienstleistungen, Part 2 greift dieselben Thematiken auf, um die entsprechenden Anforderungen an den Auftragnehmer zu formulieren. Viele Abschnitte in den beiden Texten sind identisch.

### 9.3.3 Das Binnenverhältnis der DIN- und ISO-Normen zur Personalauswahl

Die ISO 10667 ersetzt nicht die DIN 33430, sondern die DIN 33430 bleibt als nationale Norm erhalten und steht neben der ISO-Norm. Die nationale DIN-Arbeitsgruppe hat aber die Veröffentlichung der ISO 10667 zum Anlass genommen, eine Überarbeitung der DIN 33430 zu beschließen. Ziel der Modifikation, mit der im Jahre 2011 begonnen wurde und die zum Zeitpunkt der Publikation des vorliegenden Beitrags noch nicht abgeschlossen war, ist es, die Schnittstellen zur ISO 10667 herauszuarbeiten und ggf. einzelne Inhalte der ISO 10667 in die DIN 33430 aufzunehmen. Die ISO führt also nicht zu einem Bruch mit dem etablierten DIN 33430 System. Wohl aber gibt die ISO Anlass, die DIN 33430 fortzuschreiben.

## 9.4 Fazit

Der Ausgangspunkt jeglicher Qualitätssicherung und -optimierung in der Personalauswahl ist eine Verständigung darüber, welche konkreten Ziele man mit der Personalauswahl verfolgt und welche Kriterien Qualität indizieren. Orientierung bieten Qualitätsstandards. Der zentrale Qualitätsstandard ist die DIN 33430. Bei der Erstellung der DIN 33430 wurden andere Qualitätsstandards berücksichtigt. Soweit die DIN 33430 inhaltlich noch Lücken lässt (beispielsweise bei der Adaption von diagnostischen Verfahren in andere Sprach-/Kulturräume sowie bezüglich der computergestützten Personalauswahl), behalten die anderen Standards ihre Bedeutung, um den Stand wissenschaftlicher Erkenntnis oder in Fachkreisen anerkannter Verfahrenspraxis zu repräsentieren. Mittlerweile hat sich die DIN 33430 fest etabliert. In Ausschreibungen (z. B. der Bundesbank oder der Wasser- und Schifffahrtsverwaltung des Bundes) zu personaldiagnostischen Dienstleistungen wird die Beachtung der DIN 33430 zur Voraussetzung der Auftragsvergabe erklärt. Viele Beratungsunternehmen und fast jedes aktuelle Testmanual referenzieren die DIN 33430. Die DIN 33430 hat zahlreiche Begleit- und Folgemaßnahmen nach sich gezogen. Beispiele dafür sind die im vorliegenden Artikel behandelten Checklisten zur DIN 33430 (Kersting, 2008), das neue Testbeurteilungssystem des Diagnostik- und Testkuratoriums (2010) sowie die Organisationszertifizierungen, Fortbildungen und Lizenzprüfungen. Einige Universitätsinstitute integrieren die Inhalte der DIN 33430 in das Studium und bieten den Studenten die Möglichkeit, die Personenlizenzen zur DIN 33430 studienbegleitend zu erwerben (siehe Dormann, Moosbrugger, Stemmler & Maier, 2009). 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 explizit die DIN 33430 in Anspruch nimmt. Besonders hervorzuheben ist die 2011 erfolgreich abgeschlossene Entwicklung einer ISO-Norm, bei der die DIN 33430 zu Beratungszwecken herangezogen wurde. Die ISO 10667 ergänzt die DIN

33430 um eine internationale Perspektive und ist somit insbesondere für Organisationen interessant, die sich mit der Auswahl von Personal im internationalen Kontext befassen. Die ISO 10667 liefert außerdem wertvolle Hinweise für die Auftragsklärung und für die Gestaltung des Dienstleistungsvertrages zwischen Auftraggeber und (internen oder externen) Auftragnehmern und ist vorbildlich hinsichtlich der hohen Anforderungen an die Informationspolitik und Transparenz.

#### **Fallbeispiel: Problemlösung**

Zur Konfliktlösung richtete das Unternehmen eine Arbeitsgruppe „Qualität in der Personalauswahl“ ein, zu der auch ein externer Berater hinzugezogen wurde. Zunächst einigte man sich darauf, welche Kriterien die Personalauswahl erfüllen muss, um von allen Beteiligten als qualitativ hochwertig eingestuft zu werden. Alle Beteiligten verständigten sich darauf, die Personalauswahl an dem Qualitätsstandard DIN 33430 auszurichten. Anhand der Checklisten zur DIN 33430 wurde eine Bestandsaufnahme der Stärken und Schwächen der bisherigen Praxis vorgenommen. Dabei zeigte sich vor allem ein mangelhafter Anforderungsbezug der eingesetzten Prozesse und Verfahren sowie ein Vakuum hinsichtlich der Regeln zur Entscheidungsfindung (wie kommt man von den Ergebnissen in den Teilverfahren zu einem Gesamtergebnis). Zahlreiche Aspekte der Durchführung und Auswertung der eingesetzten Verfahren zur Eignungsbeurteilung sowie zur Interpretation der Ergebnisse waren – entgegen den Forderungen der DIN 33430 – nicht geregelt. Auch die Dokumentation erwies sich als Schwachstelle des bisherigen Verfahrens. Bezüglich der eingesetzten Verfahren stellte sich heraus, dass die vom GPOP Persönlichkeitsfragebogen erfassten Dimensionen nicht den erfolgsrelevanten Dimensionen laut Anforderungsprofil entsprachen und der Test außerdem die nach DIN 33430 gestellten Anforderungen an den Informationsgehalt von Verfahrenshinweisen nach Einschätzung von Höft und Muck (2009) nicht gut erfüllt. Da die DIN den Einsatz von Verfahren mit einer möglichst hohen Gültigkeit (Validität) fordert (DIN, 2002, S. 7), wurde die Gestaltung des AC vor dem Hintergrund aktueller wissenschaftlicher Studien (siehe Kasten, S. 246) reflektiert. Obwohl das bisherige AC mehrere Tage dauert, kommen nur wenige Übungen (z. B. lange Gruppendiskussionen) zum Einsatz. Im Zuge der Qualitätsoptimierung wurde beschlossen, dass AC in Zukunft kürzer, aber zugleich auch hinsichtlich der eingesetzten Instrumente (z. B. Übungen) heterogener zu gestalten. Außerdem soll im Rahmen des AC zukünftig auch ein Intelligenztest eingesetzt werden. Ein halbstandardisiertes Interview ersetzt zukünftig das bisherige freie Gespräch. Das neue Personalauswahlverfahren wird von einem konstanten Team gleichartig in Berlin und München durchgeführt. Im Anschluss an jedes einzelne Verfahren wird ein „post-assessment review“ im Sinne der ISO 10667 durchgeführt, eine erste Evaluation des Verfahrens erfolgt nach einem Jahr.

## Tipps für die Praxis

### Tipps:

1. Qualität in der Personalauswahl setzt eine Reflexion über den Qualitätsbegriff voraus (wer versteht was unter dem Begriff „Qualität“?). Ausgangspunkt der Qualitätssicherung und Optimierung ist die Einigung auf und verbindliche Festlegung von Qualitätsstandards (vgl. Kapitel 9.1).
2. Wichtig ist die Auftragsklärung. Welche Ziele sollen mit der Personalauswahl erreicht werden? Woran wird der Erfolg der Personalauswahl im Nachhinein festgemacht? Welche Aufgaben übernimmt der Auftraggeber, welche der Auftragnehmer? (siehe Auftragsklärung im Kontext der ISO 10667, vgl. Kapitel 9.3.2).
3. Grundlage jeder Personalauswahl ist die Anforderungsanalyse. Jeder Anforderungsdimension, die erhoben werden soll, muss mindestens ein Verfahren zugeordnet werden, mit dem diese Dimension erfasst wird. Und jedem der eingesetzten Verfahren muss mindestens eine Anforderungsdimension zugeordnet werden, zu deren Erfassung es eingesetzt wird (vgl. Kapitel 9.3.1).
4. Es reicht nicht aus, die Anforderungen abstrakt zu beschreiben (z. B. „soziale Kompetenz“), sondern den Anforderungsdimensionen müssen vorab Beispielaussagen und Beispielverhaltensweisen der Kandidaten zugeordnet werden (vgl. Kapitel 9.3.1).
5. Alle Aspekte der Durchführung und Auswertung der Verfahren zur Personalauswahl sowie zur Interpretation der Ergebnisse müssen vorab festgelegt und dokumentiert werden (vgl. Kapitel 9.3.1).
6. Personen, die im Rahmen der Personalauswahl Aufgaben übernehmen, müssen für diese Aufgaben qualifiziert sein. Einzelpersonen können durch Personenzertifikate nachweisen, dass sie über die nach DIN 33430 für bestimmte Aufgaben in der Eignungsdiagnostik geforderten Kenntnisse verfügen (vgl. Kapitel 9.3.1).
7. Organisationen können ein Zertifikat erwerben, das die Konformität des Qualitätsmanagementsystems mit der Norm bestätigt (vgl. Kapitel 9.3.1).
8. Voraussetzung für die Qualitätsüberprüfung von Tests und Fragebogen ist, dass genügend Informationen (z. B. zur empirischen Bewährung) des Verfahrens zur Verfügung stehen. Das Diagnostik- und Testkuratorium hat auf der Basis der DIN 33430 eine Checkliste mit Mindestanforderungen an den Informationsgehalt von Verfahren entwickelt. Bevor man ein Verfahren einsetzt, sollte man prüfen, ob diese Informationen vorliegen.
9. Um einen „deutschen“ Test oder Fragebogen in einem anderen Sprach- und Kulturraum einsetzen zu können, reicht es nicht aus, dass eine Übersetzung des Verfahrens vorliegt. Die Übersetzung muss bestimmten Standards entsprechen, außerdem ist neben der Sprache die „kulturelle Distanz“ zu berücksichtigen, die sich in Bereichen wie Sprache, Familienstruktur, Religion, Lebensstil und Wertekanon manifestieren (vgl. Kapitel 9.1). Vergleichbares gilt für den internationalen Einsatz von Assessment Centern.

10. Computer und Internet können für die Personalauswahl genutzt werden. Hierbei sind zahlreiche technische Standards sowie Anforderungen an den Datenschutz zu beachten (vgl. Kapitel 9.1).

## Lernquiz

Bei jeder Frage können gegebenenfalls auch mehrere Antworten richtig sein. Die Lösungen finden Sie unter [www.hogrefe.de/buecher/lehrbuecher/psychlehrbuchplus](http://www.hogrefe.de/buecher/lehrbuecher/psychlehrbuchplus).

1. Im Sinne des professionellen Qualitätsmanagements, im Sinne von DIN und ISO-Normen ist Qualität definiert als ...
  - 1 statistisches Qualitätsziel (es dürfen maximal 0,62 % aller Prozesse fehlerhaft sein)
  - 2 der Grad, in dem ein Satz inhärenter Merkmale Anforderungen erfüllt
  - 3 höchste Güte
  - 4 eine Beschaffenheit
  
2. Höft und Muck (2009) haben im Auftrag des Diagnostik- und Testkuratoriums der Föderation deutscher Psychologenvereinigungen gemäß den TBS-TK-Richtlinien im Jahre 2009 eine Rezension zum Golden Profiler of Personality (GPOP) von Golden, Bents und Blank erstellt/publiziert. Ihrer Ansicht nach werden die TBS-TK Anforderungen bezüglich der Validität des Tests nicht erfüllt. Welche der folgenden Aussagen sind richtig?
  - 1 Höft und Muck sind das Beurteilungsobjekt
  - 2 Golden, Bents und Blank sind das Beurteilungsobjekt
  - 3 Die Aussage, dass der GPOP die TBS-TK-Anforderungen bezüglich der Validität nicht erfüllt, ist das Beurteilungsergebnis
  - 4 Die DIN 33430 ist eine Norm zur Beurteilung von Tests und Fragebogen
  
3. Becker, Höft, Holzenkamp und Spinath (2011) führten eine Metaanalyse zur Kriteriumsvalidität von Assessment Centern (AC) durch. Ihren Moderatoranalysen zufolge
  - 1 schneiden kürzere AC (max. ein Tag) unter Validitätsgesichtspunkten besser ab als längere AC
  - 2 sind AC, die einen Intelligenztest inkludieren, aussagekräftiger als AC ohne Intelligenztest
  - 3 sind AC im deutschsprachigen Raum valider als im englischsprachigen Raum
  - 4 lassen sich harte Erfolgskriterien (wie Verkaufszahlen) besser vorherhersagen als „weiche“ (z. B. Vorgesetztenurteile)

4. Qualitätsstandards der Personalauswahl ...
  - ① zielen auf die Vereinheitlichung der Personalauswahlprozesse
  - ② erfüllen (standes-)ethische Funktionen
  - ③ formulieren Regeln und Prinzipien, deren Beachtung zur Qualitätssicherung und -optimierung beiträgt
  - ④ zielen auf die Steigerung der Gewinne der Anbieter entsprechender Dienstleistungen (die bei Gewährleistung des Standards Monopolpreise erzielen können)
5. Eine Bewertung von Tests und Fragebogen allein aufgrund der numerischen Ausprägung der Testgütekriterien ist u. a. deshalb problematisch, weil
  - ① ... Testgütekennwerte nicht nur vom Test, sondern auch von der Untersuchungsgruppe abhängen
  - ② ... auch Tests von Anbietern ohne wissenschaftlichen Hintergrund manchmal hohe Kennwerte aufweisen
  - ③ ... sich die Qualität von Tests im Wesentlichen dadurch zeigt, dass die getesteten Personen die Gutachten plausibel finden
  - ④ ... Kennwerte lediglich Einzelschätzungen darstellen
6. Bei der Organisationszertifizierung im DIN/ISO-Kontext ...
  - ① wird die Übereinstimmung des in einer Organisation geübten Qualitätsmanagementsystems mit der Norm durch ein Zertifikat bestätigt
  - ② führen private Unternehmen zur Prüfung auf Konformität Auditierungen durch
  - ③ führen staatliche Organisationen zur Prüfung auf Konformität Auditierungen durch
  - ④ wird darauf geachtet, dass nur Psychologen die Personalauswahl durchführen
7. Gegenstand der DIN 33430 sind ...
  - ① spezielle Testverfahren
  - ② Personalentscheidungen
  - ③ die Qualifikation der Auftraggeber
  - ④ die Qualifikation der Auftragnehmer
8. Zum Thema Anforderungsanalyse führt die DIN 33430 aus:
  - ① Jeder Anforderungsdimension, die erhoben werden soll, muss mindestens ein Verfahren zugeordnet werden, mit dem diese Dimension erfasst wird
  - ② In einem Personalauswahlverfahren dürfen nicht mehr als sieben Anforderungsdimensionen erhoben werden
  - ③ Jedem der eingesetzten Verfahren muss mindestens eine Anforderungsdimension zugeordnet werden, zu deren Erfassung es eingesetzt wird
  - ④ Jede Anforderungsanalyse muss von zwei unabhängig voneinander arbeitenden Personen durchgeführt werden

9. Die DIN 33430 unterscheidet zwischen dem Auftragnehmer einerseits und dem sogenannten Mitwirkenden andererseits. Welche der folgenden Aussagen trifft zu?
- 1 Assessoren im AC sind Mitwirkende im Sinne der DIN 33430
  - 2 Zum Interview hinzugezogene Experten der Fachabteilung sind Mitwirkende im Sinne der DIN 33430
  - 3 Bei Auftragnehmern im Sinne der DIN handelt es sich um organisationsexterne Personen
  - 4 Bei der DIN fallen Auftraggeber und Auftragnehmer zusammen, da beide Verantwortung für die Qualität tragen
10. Die ISO 10667-1 und ISO 10667-2 formulieren Anforderungen ...
- 1 an Testverfahren
  - 2 an den Auftragnehmer
  - 3 an den Auftraggeber
  - 4 an den Teilnehmer

## Übung

Das Bundesland xy stellt regelmäßig Anwärter (m/w) für den gehobenen Polizeivollzugsdienst ein. Nach einer Vorauswahl aufgrund von Zeugnisnoten erfolgt ein Online-Test zur weiteren Vorauswahl. Die dritte Phase der Auswahl sieht ein ein-tägiges Assessment Center vor. Sie werden beauftragt, die Qualität der Personalauswahl zu prüfen und Optimierungsvorschläge einzubringen. Entwickeln Sie einen Plan für das Vorgehen. Lassen Sie sich dabei von folgenden Fragen leiten:

- Welche weiteren Informationen würden Sie erfragen?
- Welche Aspekte der Personalauswahl würden Sie näher untersuchen?
- Welche Qualitätsstandards (Soll-Werte) legen Sie zugrunde?
- Wie genau nehmen Sie den Abgleich von Soll-Werten (Qualitätsstandards) und Ist-Werten (Praxis der Personalauswahl) vor?

## Weiterführende Literatur

DIN. (2010). DIN 33430: Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen. In K. Westhoff, C. Hagemeyer, M. Kersting, F. Lang, H. Moosbrugger, G. Reimann. & G. Stemmler (Hrsg.), *Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33430* (3., überarb. Aufl., S. 232–259). Lengrich: Pabst Science Publishers.

Kersting, M. (2008). *Qualität in der Diagnostik und Personalauswahl: Der DIN Ansatz*. Göttingen: Hogrefe.

## 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: American Educational Research Association.
- Anderson, N., Herriot, P. & Hodgkinson, G.P. (2001). The practitioner-researcher divide in industrial, work and organizational (IWO) psychology: Where are we now, and where do we go from here? *Journal of Occupational and Organizational Psychology*, 74, 391–441.
- Arbeitskreis Assessment Center. (2004). *Standards der Assessment Center Technik*. Hamburg: Autor.
- Arbeitskreis Assessment Center. (2008). *Interview-Standards*. Hamburg: Autor.
- Association of Test Publishers (ATP) (2002). *Guidelines for computer-based testing*. Washington: ATP.
- Bartram, D. (1996). Test qualifications and test use in the UK: The competence approach. *European Journal of Psychological Assessment*, 12, 62–71.
- Bartram, D. & Hambleton, R. (Eds.). (2006). *Computer-based testing and the internet: Issues and advances*. New York: John Wiley.
- Becker, N., Höft, S., Holzenkamp, M. & Spinath, F.M. (2011). The predictive validity of assessment centers in German-speaking regions: A meta-analysis. *Journal of Personnel Psychology*, 10, 61–69.
- 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: University of Illinois Press.
- DIN. (2002). *DIN 33430: Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen. Ref. Nr. DIN 33430:2002–06*. Berlin: Beuth.
- Dormann, C., Moosbrugger, H., Stemmler, G. & Maier, G.A. (2009). Erwerb von Personenzulassungen zur DIN 33430 im Rahmen des Psychologiestudiums. *Psychologische Rundschau*, 60, 23–26.
- Dries, C., Krumm, S. & Lux, V. (2006). Prozesszertifizierung der Führungskräfteauswahl nach DIN 33430 in einer deutschen Großbank. In K. Westhoff (Hrsg.), *Nutzen der DIN 33430 – Praxisbeispiele und Checklisten* (S. 140–147). Lengerich: Pabst.
- 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.
- Evers, A. (1996). Regulations concerning test qualifications and test use in the Netherlands. *European Journal of Psychological Assessment*, 12, 153–159.
- Evers, A. (2012) The Internationalization of Test Reviewing: Trends, Differences and Results. *International Journal of Testing*, 12, 136–156.
- Fischer, G.H. (1968) *Einführung in die Theorie psychologischer Tests: Grundlagen und Anwendungen*. Bern: Huber.
- Föderation Deutscher Psychologenverbände. (1986). Beschreibung der einzelnen Kriterien für die Testbewertung. *Diagnostica*, 32, 358–360.

- Föderation Deutscher Psychologenverbände. (2011). *Fortbildungs- und Prüfungsordnung der Föderation Deutscher Psychologenvereinigungen zur Personenlizenzierung für berufsbezogene Eignungsbeurteilungen nach DIN 33430*. Bonn: Autor.
- Gergen, K. J. (1985). The social constructionist movement in modern psychology. *American Psychologist*, 40, 266–275.
- Häcker, H., Leutner, D. & Amelang, M. (1998). *Standards für pädagogisches und psychologisches Testen*. Göttingen: Hogrefe.
- Hambleton, R. K. (2001). The next generation of the ITC test translation and adaptation guidelines. *European Journal of Psychological Assessment*, 17, 164–172.
- Herrmann, T. (1966). Sammelreferat: Zur Geschichte der Berufseignungsdiagnostik. *Archiv für die gesamte Psychologie*, 118, 253–278.
- Höft, S. & Muck, P. M. (2009). TBS-TK Rezension: Golden Profiler of Personality (GPOP). Deutsche Adaptation des Golden Personality Type Profiler von John P. Golden. *Report Psychologie*, 7/8, 322–323.
- Holzenkamp, M., Spinath, F. M. & Höft, S. (2010). Wie valide sind Assessment Center im deutschsprachigen Raum? Eine Überblicksstudie mit Empfehlungen für die AC-Praxis. *Wirtschaftspsychologie*, 12, 17–25.
- 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. überarbeitete Aufl., S. 255–283). Lengerich: Pabst.
- Hunter, J. E. & Schmidt, F. L. (2004). *Methods of metaanalysis* (2<sup>nd</sup> ed.) Thousand Oaks: Sage.
- ISO (2011a). *ISO 10667-1. Assessment service delivery – Procedures and methods to assess people in work and organizational settings – Part 1: Requirements for the client*. Genf: ISO.
- ISO (2011b). *ISO 10667-2. Assessment service delivery – Procedures and methods to assess people in work and organizational settings – Part 1: Requirements for service providers*. Genf: ISO.
- ITC (2000). *International guidelines for test use*. (Version 2000). USA: Author.
- ITC (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 (Ed.), *Handbook of employee selection* (pp. 593–625). New York: Routledge.
- Kanning, U. P. (2004). *Standards der Personaldiagnostik*. Göttingen: Hogrefe.
- Kanning, U. P. (2007). *Förderung sozialer Kompetenzen in der Personalentwicklung*. Göttingen: Hogrefe.
- Kanning, U. P. (2010). *Von Schädeldeutern und anderen Scharlatanen. Unseriöse Methoden der Psychodiagnostik*. Lengerich: Pabst.
- Kersting, M. (2006). Zur Beurteilung der Qualität von Tests: Resümee und Neubeginn. *Psychologische Rundschau*, 57, 243–253.
- Kersting, M. (2008). *Qualität in der Diagnostik und Personalauswahl: Der DIN Ansatz*. Göttingen: Hogrefe.
- Kersting, M. (2010a). Tests und Persönlichkeitsfragebogen in der Personalarbeit. *Personalführung*, 10/2010, 20–31.
- Kersting, M. (2010b). 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., Häcker, H. & Hornke, L.F. (2011). Qualitätsstandards in der Diagnostik. In L.F. Hornke, M. Amelang & M. Kersting (Hrsg.), *Grundfragen und Anwendungsfelder psychologischer Diagnostik* (Enzyklopädie der Psychologie, Serie Psychologische Diagnostik, Band 1, S. 1–86). Göttingen: Hogrefe.
- Kersting, M. & Westmeyer, H. (2013). Qualitätsbeurteilung management-diagnostischer Verfahren und Prozesse. In W. Sarges (Hrsg.), *Management-Diagnostik* (4. Auflage, S. 948–954). Göttingen: Hogrefe.
- Lindley, P. A. & Bartram, D. (2012). Use of the EFPA Test Review Model by the UK and Issues Relating to the Internationalization of Test Standards. *International Journal of Testing*, 12, 108–121.
- Muchinsky, P.M. (2004). When the psychometrics of test development meets organizational realities: A conceptual framework for organizational change, examples, and recommendations. *Personnel Psychology*, 57, 175–209.
- 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.
- Schmidt, F.L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in Psychology. *American Psychologist*, 47, 1173–1181.
- Schuler, H. (1991). Ethische Probleme der (sozial-)psychologischen Forschung. In H. Lenk (Hrsg.), *Wissenschaft und Ethik* (S. 331–355). Stuttgart: Reclam.
- Schuler, H. (1996). *Psychologische Personalauswahl. Einführung in die Berufseignungsdiagnostik*. Göttingen: Hogrefe.
- Schuler, H. (2002a). Emotionale Intelligenz – ein irreführender und unnötiger Begriff. *Zeitschrift für Personalpsychologie*, 3, 138–140.
- Schuler, H. (2002b). *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.
- Sektion Arbeits-, Betriebs- und Organisationspsychologie im Berufsverband Deutscher Psychologen. (1980). *Grundsätze für die Anwendung psychologischer Eignungsuntersuchungen in Wirtschaft und Verwaltung*. Bonn: Berufsverband Deutscher Psychologen.
- Sireci, S.G. (1998). The construct of content validity. *Social Indicators Research*, 45, 83–117.
- Society for Industrial and Organizational Psychology, Inc. (SIOP) (2003). *Principles for the validation and use of personnel selection procedures* (4<sup>th</sup> ed.). Bowling Green, Author.
- Spies, R.A., Carlson, J.F. & Geisinger, K.F. (2010). *The eighteenth mental measurements yearbook*. Nebraska: University of Nebraska Press.
- Task force on AC Guidelines (2009). Guidelines and ethical considerations for assessment center operations. *International Journal of Selection and Assessment*, 17, 243–254.
- Testkuratorium der Föderation Deutscher Psychologinnenvereinigungen (2010). TBS-TK. Testbeurteilungssystem des Testkuratoriums der Föderation Deutscher Psychologinnenvereinigungen (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.

- Weber, H. & Westmeyer H. (1999). Emotionale Intelligenz: Kritische Analyse eines populären Konstruktes. In *www.Literaturkritik.de, Nr. 2*.
- Wegener, M. (2003). Rechtliche Verbindlichkeit der DIN 33430 für Behörden und Gerichte. *DGP Informationen, 48*, 7–11.
- Westhoff, K., Hagemeister, C., Kersting, M., Lang, F., Moosbrugger, H., Reimann, G. & Stemmler, G. (Hrsg.). (2010). *Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33430* (3. Aufl.). Lengerich: Pabst Science Publishers.
- Westmeyer, H. (2011). Communicamus ergo sum oder Am Anfang stehen die Beziehungen: Hans Westmeyer über Kenneth Gergens Konstruierte Wirklichkeiten. In B. Pörksen (Hrsg.), *Schlüsselwerke des Konstruktivismus* (S. 411–424). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Woehr, D. J. & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology, 67*, 189–205
- Wottawa, H. & Oenning, S. (2002). Von der Anforderungsanalyse zur Eignungsbeurteilung: Wie praktikabel ist die neue DIN 33430 bei der Bewerberauswahl. *Wirtschaftspsychologie, 4*, 43–56.