

2.2 Anforderungen an die Qualität der Verfahren

Martin Kersting und Helmut Heyse

Für die berufsbezogene Eignungsbeurteilung sollen nur solche Verfahren genutzt werden, die nachweislich einen Bezug zu den Anforderungen haben. Die Auswahl der Verfahren ist sorgfältig zu planen, die Anwendung, Auswertung und Interpretation der Verfahren müssen regelgerecht und professionell erfolgen. Qualitätskriterien sind u. a. die Objektivität, Zuverlässigkeit und Gültigkeit. Diese Qualitätskriterien werden im folgenden Beitrag erläutert. Außerdem thematisiert der Beitrag die Qualität von Norm- und Referenzkennwerten sowie die Aktualität von Gütekriterien und Normen.

Die Güte einer Eignungsbeurteilung hängt zum einen von der sorgfältigen Planung und Durchführung des gesamten Ablaufs einer Eignungsuntersuchung ab (s. Kapitel 2.1), zum anderen von der Qualität der eingesetzten Verfahren/Instrumente und deren Anwendung.

Die Verfahren müssen drei Hauptkriterien genügen:

1. Objektivität in Durchführung, Auswertung und Interpretation,
2. Zuverlässigkeit in der Erfassung von Eignungsmerkmalen,
3. Gültigkeit in der Aussage.

Die folgenden knappen Erläuterungen sollen ein allgemeines Verständnis der Gütekriterien ermöglichen. Vertiefende Darstellungen finden sich z. B. bei Amelang und Zielinski 2002, Fisseni 1997, Jäger und Petermann 1995, Kubinger 1996, Kubinger und

2004

Jäger 2002, Lienert und Raatz 1994, Rost 1996, Schuler 1996, Wotawa 1980.

In der Norm finden sich dezidierte Regelungen zu diesem Thema in den Abschnitten 4.2.2 bis 4.2.5., in Anhang A.6 und A.7 sowie im Glossar.

2.2.1 Objektivität

„Grad, in dem die mit einem Verfahren zur Eignungsbeurteilung erzielten Ergebnisse unabhängig vom Auftragnehmer und seinen Mitwirkenden sowie von weiteren irrelevanten Einflüssen sind“ (Norm 3.8).

Die DIN-Norm fordert eine größtmögliche Objektivität. Dies bedeutet:

- Größtmögliche Unabhängigkeit der Ergebnisse des Verfahrens von den Personen und situativen Bedingungen bei der Durchführung (Durchführungsobjektivität). Dies wird u. a. sichergestellt durch das Einhalten von standardisierten Durchführungsbestimmungen.
- Größtmögliche Unabhängigkeit der Ergebnisse des Verfahrens von den Personen, die die Verfahren auswerten (Auswertungsobjektivität). Die Auswertungsobjektivität wird erhöht, wenn die Auswertung nach festgelegten, sinnvollen Regeln durchgeführt wird, z. B. mit Schablonen, Kategoriensystemen oder computergestützten Auswertungssystemen.
- Größtmögliche Unabhängigkeit der Interpretation der Ergebnisse des Verfahrens von den Personen, die diese Interpretation vornehmen (Interpretationsobjektivität). Auch die Objektivität der Interpretation wird durch verbindliche Regelungen gefördert, z. B. durch Angabe von Grenzwerten für bestimmte Kategorisierungen.

Eine gezielte Schulung der Anwender kann sich auf alle genannten Aspekte der Objektivität positiv auswirken. Weitere Ausführungen zur Objektivität finden sich in Abschnitt 4.2.2 sowie im Glossar der Norm.

2.2.2 Zuverlässigkeit

„Grad der Genauigkeit eines Verfahrens, mit dem es das gemessene Merkmal erfasst“ (Norm 3.11).

Dabei ist es gleichgültig, ob das Verfahren dieses Merkmal überhaupt zu messen beansprucht — dies ist eine Frage der Gültigkeit.

Die Bestimmung der Zuverlässigkeit kann auf mehreren Wegen erfolgen:

- durch wiederholte Messung des interessierenden Merkmals an derselben Stichprobe zu zwei oder mehreren unterschiedlichen Zeitpunkten (Wiederholungszuverlässigkeit oder Retest-Reliabilität) oder
- durch den Vergleich unabhängiger Schätzungen/Messungen des interessierenden Merkmals, die zu einem Zeitpunkt erfolgen. Beispiele sind für schriftliche Testverfahren bei mehrmaliger Testung die so genannte „Paralleltest-Reliabilität“ oder – bei einmaliger Testvorgabe – die „Split-Half-Reliabilität“ und die „Interne Konsistenz“.

In jedem Fall ist in den Verfahrenshinweisen anzugeben, nach welcher Methode und anhand welcher Untersuchungsgruppe(n) die Zuverlässigkeit bestimmt wurde. Außerdem ist die Angemessenheit der gewählten Bestimmungsmethode zu erläutern.

Beispiele für die Bestimmung der Zuverlässigkeit bei Assessment Centern oder Eignungsinterviews

Ein Anhaltspunkt für die Zuverlässigkeit von Assessment Centern oder Eignungsinterviews lässt sich u. a. dadurch gewinnen, dass man berechnet, inwieweit die Urteile der beteiligten Beobachter oder Interviewer übereinstimmen. Die Übereinstimmung dürfte umso höher sein, je mehr die Assessment-Center-Aufgaben und Interviewfragen vorstrukturiert und je präziser Regeln für die Auswertung der beobachteten Reaktionen vereinbart sind.

Die Zuverlässigkeit von Verhaltensbeobachtungen durch verschiedene Beobachter kann u. a. durch Beobachterschulungen erhöht werden. Auch eine genaue Definition und Kategorisierung der interessierenden Verhaltensweisen oder Verhaltensklassen können die Zuverlässigkeit erhöhen.

Zur Bewertung der Zuverlässigkeit

In jedem Fall resultiert ein Zuverlässigkeitskoeffizient zwischen 0,0 (Zufall; keinerlei Genauigkeit der Messung) und 1,0 (perfekte Genauigkeit). Dieser Koeffizient erlaubt eine Aussage darüber, wie hoch z. B. die Übereinstimmung zwischen der ersten und zweiten Messung bzw. zwischen den Parallelformen oder zwischen verschiedenen Beurteilern ist.

Die Anforderungen an die Zuverlässigkeit der ausgewählten Verfahren variieren in Abhängigkeit von der jeweiligen Fragestellung. Wenn es z. B. bei einer ersten Vorauswahl (Screening) lediglich darum geht, extrem leistungsstarke Kandidaten von extrem schwachen Kandidaten zu unterscheiden, kann auch ein Verfahren mit geringerer Zuverlässigkeit ein qualitativ zufriedenstellendes Ergebnis liefern.

Die Beurteilung der Zuverlässigkeit eines Verfahrens kann nicht schematisiert erfolgen, indem man die numerische Höhe der Zuverlässigkeitskoeffizienten verschiedener Verfahren mit einem bestimmten Richtwert vergleicht. Entscheidend ist vielmehr die Qualität der Untersuchungen, mit denen die Zuverlässigkeit bestimmt wurde. Qualitätsmerkmale solcher Untersuchungen sind beispielsweise Größe, Repräsentativität (für die Zielgruppe), Aktualität der Untersuchung sowie vor allem die Ergebnisse von unabhängigen Vergleichs- und Wiederholungsuntersuchungen.

Wird beispielsweise ein Verfahren zur Diagnose von Eignungsmerkmalen von Managern in Führungspositionen angeboten, so spricht es nicht für das Verfahren, wenn die Zuverlässigkeit lediglich anhand einer Gruppe von Studenten bestimmt wurde — selbst wenn sich dabei hohe Zuverlässigkeitswerte ergeben haben sollten.

Bei einem stabilen Merkmal wie Intelligenz würde es z. B. gegen die Zuverlässigkeit eines Intelligenztests sprechen, wenn

sich bei denselben Personen mit demselben Verfahren unter vergleichbaren Untersuchungsbedingungen nach einem Jahr ganz andere Intelligenzwerte ergeben würden als bei der ersten Messung.

Da es im Kontext der Berufseignung häufig um die Erfassung stabiler Merkmale geht, räumt die Norm der Wiederholungszuverlässigkeit eine besondere Bedeutung ein.

Sie fordert, dass die Zuverlässigkeit von Verfahren, die zumindest partiell zeit- und situationsstabile Eignungsmerkmale erfassen sollen, über die Methode der Messwiederholung (Wiederholungszuverlässigkeit) bestimmt wird. Da diese Methode aber etwas aufwändiger ist, wird sie leider zu selten angewandt.

Fragwürdig sind Verfahren möglicherweise auch dann, wenn die Untersuchungen, in denen sich eine gute Zuverlässigkeit ergab, nur vom Verfahrensanbieter und/oder -entwickler selbst und nicht zusätzlich von unabhängigen Organisationen durchgeführt wurden.

Weitere Ausführungen zur Zuverlässigkeit finden sich in der Norm unter Abschnitt 4.2.3 und A.6 sowie im Glossar.

2.2.3 Gültigkeit

„Genauigkeit, mit der ein Verfahren tatsächlich das misst oder vorhersagt, was es messen oder vorhersagen soll (z. B. ein Merkmal oder eine Verhaltensweise einer Person)“ (Norm 3.5).

Nach Lienert und Raatz (Lienert und Raatz 1994, S. 10) ist ein Verfahren dann gültig, wenn seine Ergebnisse einen sicheren Rückschluss auf den Ausprägungsgrad des zu erfassenden Eignungsmerkmals zulassen. Es darf z. B. nicht passieren, dass ein Test, der die Rechenfertigkeit erfassen soll, eigentlich die Verständnisfähigkeit für sprachliche Textaufgaben prüft.

Die Bestimmung der Gültigkeit kann auf unterschiedliche Weise erfolgen:

- **Inhaltsgültigkeit:** Hier geht es um den Nachweis, dass das Verfahren Aufgaben enthält, die eine repräsentative Stichprobe des interessierenden Inhaltsbereichs darstellen. Voraussetzung ist, dass dieser interessierende Inhaltsbereich genau definiert ist. Ein inhaltsgültiger Vokabeltest zu einer bestimmten Unterrichtseinheit muss beispielsweise eine repräsentative Auswahl der Vokabeln enthalten, die zu lernen waren.

Häufig werden zur Beurteilung der Inhaltsgültigkeit auch Experten aus dem betreffenden Fachgebiet herangezogen. In diesem Fall müssen der fachbezogene Ausbildungsstand, die Erfahrung und die Qualifikation dieser Experten beschrieben werden. Darüber hinaus ist anzugeben, „wie die Experten zu einer Einschätzung gekommen sind“ und „inwieweit sie in ihrem Urteil übereinstimmen“ (Norm, Anhang A).

Näheres zur Inhaltsgültigkeit unter A.7.3 der Norm.

- **Kriteriumsgültigkeit:** Grad der Übereinstimmung zwischen der mit Hilfe eines Verfahrens erfassten Merkmale einer Person und der Einschätzung oder Einordnung dieser Person bzw. dieser Merkmale durch ein unabhängiges Außenkriterium, z. B. Ausbildungserfolg, Produktivitätsrate, Leistungsrate, Vorgesetztenbeurteilung.
- **Konstruktgültigkeit:** Ein Konstrukt ist ein gedankliches Konzept, das aus Überlegungen und Erfahrungen abgeleitet worden ist, um bestimmtes beobachtbares Verhalten zu erklären, z. B. Intelligenz oder Angst. Ein Verfahren, das ein derartiges Konstrukt oder Aspekte daraus (z. B. logisches Denken als Teil des Konstrukts „Intelligenz“) erfassen soll, beruht auf Überlegungen, worin sich logisches Denken zeigt und worin nicht.

Es gibt unterschiedliche Methoden, um die Konstruktgültigkeit eines Verfahrens zu überprüfen, z. B.:

- Analyse und Nachweis der theoretischen Stringenz,
- Zusammenhang mit anderen Beobachtungen, z. B. Ergebnissen in anderen Tests.

So sollten die Leistungen in einem Intelligenztest beispielsweise mit den Leistungen in einem anderen vergleichbaren Intelligenztest, nicht aber mit den Leistungen in einem Stimmungsfragebogen kovariieren.

Wie bereits für die Zuverlässigkeit ausgeführt, gilt auch für die Gültigkeit, dass die diesbezügliche Beurteilung eines Verfahrens nicht schematisiert erfolgen kann. Auch hier sind die Qualität der zugrunde liegenden Untersuchung sowie die Angemessenheit des gewählten Gültigkeitsansatzes zu berücksichtigen.

In jedem Fall sind die Gültigkeitshinweise und die zur Bestimmung der Gültigkeit eingesetzten statistischen Analysemethoden in einer zugänglichen Publikation zu dokumentieren.

Weitere Vorschriften in Bezug auf die Gültigkeit finden sich in der Norm unter Abschnitt 4.2.4, in Anhang A.7 sowie im Glossar, insbesondere hinsichtlich statistischer Adjustierung, Optimierung und Generalisierung.

2.2.4 Normwerte und Referenzkennwerte

Besondere Aufmerksamkeit muss den Bezugsgrößen gewidmet werden, mit denen die individuell erreichten Ergebnisse von Verfahren verglichen werden. Dies sind zum einen an Gruppen (Referenzgruppen) gewonnene Vergleichswerte (Normwerte). Ohne solche Normwerte ist keine Aussage darüber möglich, wie stark ein Merkmal einer Person im Vergleich mit einer für die Eignungsaussage relevanten Stichprobe anderer Personen ausgeprägt ist.

Der Vergleichsmaßstab kann zum anderen auch in personen- bzw. gruppenunabhängigen Kennwerten bestehen, die eine Merkmalsausprägung in Bezug auf ein bestimmtes Kriterium anzeigen, z. B. die Anschlageschwindigkeit beim Schreiben (so genannte „kriteriumsorientierte Messung“, s. z. B. Hilke 1980). Dazu macht die Norm jedoch keine näheren Aussagen.

Die Normierung eines Verfahrens kann auf der Basis einer undifferenzierten Gesamtgruppe (z. B. Erwachsene) und/oder anhand von Einzelgruppen erfolgen (z. B. differenziert nach Alter, Geschlecht, Vorbildung u. Ä.; Grob- und Feinnormen).

Entscheidend für die Frage, ob sich ein Verfahren für den aktuellen Einsatz eignet, ist, ob die Merkmale der Gruppe, anhand derer die Normen berechnet wurden, den Merkmalen der interessierenden (Bewerber-)Gruppe entsprechen. Beispielsweise können die von Senior-Managern erzielten Ergebnisse nicht sinnvoll mit den Normwerten verglichen werden, die bei Studenten oder jugendlichen Azubis gewonnen wurden.

Auch Vergleichswerte, die an nur kleinen Gruppen berechnet wurden und/oder nicht mehr aktuell sind, da die entsprechenden Daten vor Jahrzehnten gewonnen wurden, können zu Fehlinterpretationen führen (siehe Kapitel 5.1).

Sofern für die Eignungsbeurteilung Verfahren eingesetzt werden, die den Vergleich mit Normwerten vorsehen, schreibt die DIN-Norm daher vor, dass diese Normwerte der Fragestellung und den Stichprobenmerkmalen der Kandidaten entsprechen müssen. Gruppenspezifische Normierungen müssen begründet und in ihrer Besonderheit beschrieben werden.

Die Verwendbarkeit von vorgegebenen Normen hängt auch von der Art und Weise ab, wie ein Verfahren durchgeführt wird. Wird z. B. ein so genanntes Papier- und Bleistiftverfahren auf den Computer übertragen und computergestützt angewendet, muss die Übertragbarkeit von einem Medium auf das andere nachgewiesen sein (s. Norm 4.2.5). Andernfalls müssen entweder spezifische Normen für die computergestützte Darbietung vorliegen oder neue Normen berechnet werden.

Auch für die Ergebnisse mündlicher Verfahren oder von Verhaltensproben (Eignungsinterviews, Arbeitsproben, Assessment Center usw.) sind Vergleichsmaßstäbe hilfreich, z. B. für die Einschätzungen von „Teamfähigkeit“ oder „Zuverlässigkeit“ eines Kandidaten.

Da mündliche Verfahren aber oft positions- und zielgruppenspezifisch sind, wird die zum Aufbau von gruppenbezogenen Referenzkennwerten notwendige Personenanzahl häufig nicht erreicht.

Hier bieten sich bestimmte Beurteilungsverfahren an, beispielsweise:

- Häufigkeitslisten, bei denen die Verhaltensweisen nach der Häufigkeit ihres Auftretens eingestuft werden, z. B. „fast nie“ bis „fast immer“,
- Checklisten, in denen das Vorkommen bestimmter Verhaltensweisen markiert wird,
- verhaltensorientierte Einstufungsverfahren, d. h. Schätzskaleten in Bezug auf differenzierte Verhaltensdimensionen und deren Ausprägung, z. B. Verhaltensbeispiele für schlechtes, mittelmäßiges und gutes Führungsverhalten (Behaviorally anchored Rating Scales, BARS).

Grundlage für die Entwicklung solcher Skalen können u. a. Beobachtungen und Erfahrungen mit Berufsausübenden sein. Informationen zu diesen Vorgehensweisen finden sich z. B. bei Schuler (1989).

Nähere Angaben zu Normwerten und Referenzkennwerten finden sich in der Norm unter 4.2.5 und B.34 und B.39.

2.2.5 Aktualisierung von Gütekriterien und Normen

Berufliche Tätigkeiten sind einer ständigen Entwicklung unterworfen. Dies kann dazu führen, dass sich die Anforderungen in einem Beruf ändern. Auch kann sich – z. B. aufgrund geänderter Sozialisationsbedingungen – der Grad, in dem eine bestimmte Kompetenz in der Bevölkerung allgemein oder in einer spezifischen Bevölkerungsgruppe vorhanden ist, verringern oder erhöhen. So könnte sich z. B. die durchschnittliche Ausprägung des Eignungsmerkmals „EDV-Kompetenz“ durch die Verbreitung von PCs in kürzerer Frist wesentlich ändern. Beides kann dazu führen, dass Normen nicht mehr anwendbar sind, weil früher gültige Referenzkennwerte heute nicht mehr zutreffen (s. Kapitel 5.1).

Auch die Zuverlässigkeit und Gültigkeit eines Verfahrens können zeitlichen Veränderungen unterworfen sein. Ob ein seit 15 Jahren unveränderter Test zur Eignungsbeurteilung technischer Zeichner heute noch gültig ist, dürfte bezweifelt werden.

Die DIN 33430 trägt solchen möglichen Entwicklungen Rechnung, indem eine regelmäßige Überprüfung der Angemes-

senheit von Normwerten, Gütekriterien und Entscheidungsregeln vorgeschrieben wird.

Hinsichtlich der Normwerte/Referenzkennwerte wird gefordert, dass ihre Angemessenheit spätestens alle acht Jahre überprüft werden muss (Norm 4.2.5).

Falls zwischenzeitlich Erkenntnisse darüber vorliegen, dass sich das in Frage stehende Eignungsmerkmal in der Bevölkerung verändert hat, sind die Normwerte früher zu aktualisieren (s. das o. g. Beispiel der „EDV-Kompetenz“).

Auch die Zuverlässigkeit und Gültigkeit eines Verfahrens müssen mindestens alle acht Jahre überprüft werden.

„Von den als Gültigkeitsbeleg geltenden Untersuchungen müssen einige relativ aktuellen Datums sein (nicht älter als 8 Jahre). Bei einer wesentlichen Änderung des Verfahrens zur Eignungsbeurteilung müssen neue Gültigkeitsbelege erbracht werden“ (Norm A.6 und A.7).

Wenn Verfahren oder Kombinationen von Verfahren häufiger für gleichartige Fragestellungen benutzt werden, z. B. Testbatterien für Routineanwendungen, ist die Gültigkeit der Regeln, anhand derer die Ergebnisse zur Eignungsbeurteilung führen, spätestens alle drei Jahre zu überprüfen.

2.2.6 Verfahrenshinweise

Um über die Verwendbarkeit eines Verfahrens zur Eignungsbeurteilung entscheiden zu können, müssen jedem Verfahren „Verfahrenshinweise“ beigegeben sein. Die Bedeutung der Verfahrenshinweise kommt u. a. darin zum Ausdruck, dass die Norm einen speziellen normativen Anhang „Anforderungen an Verfahrenshinweise“ enthält (Anhang A).

Diese müssen den Anwender in die Lage versetzen, das Verfahren kritisch zu bewerten und ordnungsgemäß anzuwenden. Zu den aufzuführenden Informationen zählen

- die Nennung von Zielsetzungen und Anwendungsbereichen,
- Darstellungen zur Konstruktion des Verfahrens,
- Wege zur Berechnung seiner Gütemerkmale,
- Erfahrungen mit der Anwendung,

- Weiterentwicklungen,
- Angaben zur Durchführung, Auswertung und Interpretation und
- eine Aufklärung über den Aufwand und Zeitbedarf.

Die Norm verlangt wahrheitsgetreue Informationen und belegbare Darstellungen, nicht nur in den Verfahrenshinweisen, sondern auch in anderen Publikationen oder Werbematerialien zu dem Verfahren.

Soweit möglich sollten die Verfahrenshinweise auch Angaben zur Verfahrenspflege, zu Studien anderer Autoren bzgl. des Verfahrens sowie Hinweise auf seine Rezensionen enthalten. Sofern die Verfahrenshinweise sich auf eine empirische Arbeit beziehen, müssen diese eine kritische Würdigung der Ergebnisse hinsichtlich ihrer theoretischen und methodischen Grundlagen ermöglichen. Die Dokumentation solcher Arbeiten hat den üblichen Kriterien für wissenschaftliche Publikationen zu entsprechen.

Anhand der Verfahrenshinweise kann eine erste Einschätzung externer und interner Angebote in Bezug auf die Qualität der angebotenen berufsbezogenen Eignungsbeurteilungen vorgenommen werden. Vorsicht ist geboten, wenn Anbieter die Verfahrenshinweise auf nicht belegte Behauptungen wie z. B. „die Zuverlässigkeit und Gültigkeit des Verfahrens konnte immer wieder belegt werden“ begrenzen. *Fragwürdig sind auch Formulierungen wie „das Verfahren wurde in xy-tausend Fällen erfolgreich angewandt“*, da sich aus derartigen Aussagen kein Hinweis auf die Zusammensetzung und Größe der angewandten Referenzgruppen oder auf Studien und Ergebnisse zur Gültigkeitsprüfung ergibt.

Zusammenfassung

Sowohl die Verfahren als auch deren Einsatzbedingungen entscheiden über die Qualität der berufsbezogenen Eignungsbeurteilung. Der Beitrag konzentriert sich auf die direkt verfahrensbezogenen Qualitätsaspekte der Objektivität, Zuverlässigkeit und Gültigkeit. Darüber hinaus werden die Anforderungen der DIN

33430 an die Qualität von Norm- und Referenzkennwerten und an die Aktualität von Gütekriterien und Normen erläutert. Abschließend wird dargestellt, welche Erwartungen der Anwender an „Verfahrenshinweise“ stellen kann.

Literatur

- Amelang, M. & Zielinski, W. (2002). *Psychologische Diagnostik und Intervention*. Berlin: Springer.
- Fisseni, H.-J. (1997). *Lehrbuch der psychologischen Diagnostik*. Göttingen: Hogrefe.
- Hilke, R. (1980). *Grundlage normorientierter und kriteriumorientierter Tests*. Bern: Huber.
- Jäger, R.S. & Petermann, F. (Hrsg.) (1995). *Psychologische Diagnostik*. Weinheim: PVU.
- Kubinger, K.D. (1996). *Einführung in die Psychologische Diagnostik*. Weinheim: Beltz, PVU.
- Kubinger, K.D. & Jäger, R.S. (2002). *Stichwörter der psychologischen Diagnostik*. Weinheim: Beltz, PVU.
- Lienert, G.A. & Raatz, U. (1994). *Testaufbau und Testanalyse*. Weinheim: Beltz, PVU.
- Rost, J. (1996). *Testtheorie und Testkonstruktion*. Bern: Huber.
- Schuler, H. (1989). Leistungsbeurteilung. In: E. R. Roth (Hrsg.), *Enzyklopädie der Psychologie* (Bereich D: Praxisgebiete; Serie III: Wirtschafts-, Organisations- und Arbeitspsychologie; Bd. 3; S. 402-430). Göttingen: Hogrefe.
- Schuler, H. (1989). Leistungsbeurteilung. In: E. R. Roth (Hrsg.), *Enzyklopädie der Psychologie* (Bereich D: Praxisgebiete; Serie III: Wirtschafts-, Organisations- und Arbeitspsychologie; Bd. 3; S. 402-430). Göttingen: Hogrefe.
- Schuler, H. (1996). *Psychologische Personalauswahl. Einführung in die Berufseignungsdiagnostik*. Göttingen: Verlag für Angewandte Psychologie.
- Wottawa, H. (1980). *Grundriß der Testtheorie*. München: Juventa.

Lutz Hornke / Ulrich Winterfeld (Hrsg.)

**Eignungsbeurteilungen
auf dem Prüfstand:
DIN 33430 zur Qualitätssicherung**

Herausgegeben für den Berufsverband
Deutscher Psychologinnen und Psychologen
und die Deutsche Gesellschaft für Psychologie