

3.2 Qualitätssicherung und -verbesserung: Zur Überprüfung der Gültigkeit berufsbezogener Eignungsbeurteilungen

Martin Kersting

Die Deutsche Gesellschaft für Personalwesen (DGP) ist ein großer Anbieter von Personalauswahl und Personalentwicklung in Deutschland. Jährlich werden ca. 20.000 Bewerber getestet. Da sie seit über 50 Jahren Personalbeurteilungen durchführt, sind Untersuchungen zu Evaluation von Personalbeurteilungen aufgrund von Tests und anderen Instrumenten in großen Umfang möglich. Über Methoden und Ergebnisse dieser Evaluationsstudien wird berichtet.

3.2.1 Kurzvorstellung der Institution

Am Praxisbeispiel der Deutschen Gesellschaft für Personalwesen (DGP) erfolgt eine Darstellung der Qualitätssicherung und -optimierung beruflicher Eignungsbeurteilungen. Konkret wird auf die Möglichkeiten der Überprüfung der kriterienbezogenen Gültigkeit solcher Beurteilungen eingegangen. Die DGP ist die traditionsreichste und erfahrenste psychologische Personalberatung Deutschlands. Auf Initiative der amerikanischen Besatzungsmacht wurde die DGP am 1. April 1949 mit dem Ziel gegründet, den demokratischen Wiederaufbau der Bundesrepublik durch professionelle Personalarbeit zu unterstützen. Zunächst hat die DGP moderne Verfahren zur berufsbezogenen Eignungsbeurteilung von den Vereinigten Staaten übernommen, später dann in enger Kooperation mit Universitäten eigene

2004

Verfahren entwickelt. Eine Darstellung der Geschichte der DGP findet sich bei Althoff (1990 und 1999) sowie bei Greif (1995).

Die untrennbar zusammengehörenden Arbeitsgebiete der DGP sind die Personalauswahl und die Personalentwicklung. Qualitativ hochwertige berufsbezogene Eignungsbeurteilungen bleiben nicht bei der Feststellung des Ist-Zustandes und dem Abgleich mit Soll-Vorstellungen stehen, sondern münden – zumindest bei internen Personalentscheidungen – in nützlichen Vorschlägen zur persönlichen Weiterentwicklung. Dazu müssen die Potenziale der Person ebenso berücksichtigt werden, wie die konkreten Rahmenbedingungen der Organisation. Um eine förderorientierte (und nicht defizitorientierte) Eignungsbeurteilung zu gewährleisten, bedarf es daher Berater, die nicht nur Spezialisten für Eignungsbeurteilung sind, sondern auch über umfassende Kenntnisse und Erfahrungen im Bereich der Personalentwicklung verfügen. Auf die Angebote der DGP im Bereich von Training und Coaching wird hier aber nicht weiter eingegangen.

Insgesamt engagieren sich in den vier Geschäftsstellen der DGP in Berlin, Düsseldorf, Hannover und Leipzig aktuell über 50 Diplom-Psychologen für eine kompetente Beratung der Kunden bei allen Fragen des Personalwesens. Der vorliegende Bericht konzentriert sich auf die Qualitätssicherung und -optimierung der Eignungsbeurteilung. Weitere Informationen über die DGP können unter <http://www.DGP.de> abgerufen werden.

3.2.2 Berufsbezogene Eignungsbeurteilungen: Typische Fragestellungen

Von in Großgruppen durchgeführten psychologischen Eignungstests bis zu Einzel-Assessments – aufbauend auf einer systematischen Anforderungsanalyse der Zielposition oder -ebene – sind die Verfahren der DGP standardisiert und maßgeschneidert zugleich. Typische Aufträge der DGP sind die Eignungsbeurteilung für eine Ausbildung in Kommunal-, Landes- und Bundesverwaltungen (z. B. Auswärtiges Amt, Bundesbank, Bundesversicherungsanstalt für Angestellte) sowie für verschiedene Ausbildungsgänge im kaufmännischen, gewerblichen, medizinischen und sozialen Bereich. Weitere Tätigkeitsschwerpunkte

stellen die Eignungsbeurteilung bei der Besetzung von Führungspositionen sowie die Potenzialanalyse bei der systematischen Personalentwicklung dar.

Die Entwicklung von berufsbezogenen Testverfahren gehört zu den zentralen Kompetenzen der DGP. Prof. Irle (1989, S. 110) hat die DGP zu Recht als „Brutstätte quantitativer Testdiagnostik“ bezeichnet: Zahlreiche „Klassiker“ der Eignungsbeurteilung – wie der Intelligenz-Struktur-Test (I-S-T), der Wilde-Test (WIT), der Berufs-Interessen-Test (BIT) oder der Rechtschreibungstest (R-T) – wurden in der DGP entwickelt. Greif (1995, S. 34) spricht von der DGP als „Nukleus anwendungsorientierter Forschung und Entwicklung“. Für die Etablierung dieser anwendungsorientierten Wissenschaft stehen nicht zuletzt ehemalige DGP-Führungsspitzen wie Prof. R. Amthauer, Prof. K. Wilde, Prof. A.O. Jäger und Dr. K. Althoff.

Aktuell beurteilt die DGP pro Jahr ca. 20.000 Bewerber. Dazu werden DGP-eigene Leistungs- und Wissenstests eingesetzt (confidential tests). Die Tests werden nicht isoliert vertrieben, sondern den Kunden ausschließlich als Gesamtpaket angeboten, bestehend aus Testdurchführung, -auswertung und -interpretation. Dies entspricht der DIN 33430, die als Prozessnorm eine isolierte Bewertung der Tests als „Produkt“ ablehnt. Mit den schriftlichen Tests werden – je nach Anforderungsprofil – die Dimensionen „Verarbeitungskapazität“, „Merkfähigkeit“ und „Bearbeitungsgeschwindigkeit“ erfasst. Überprüfungen der Effizienz des Büroarbeitsverhaltens, der Rechtschreibung und der Kenntnisse in verschiedenen Wissensdomänen (z. B. Wirtschaftskennntnisse) können bei Bedarf ergänzt werden. Die computergestützte Testauswertung sichert die Objektivität der Testauswertung und ermöglicht in Form von Datenbanken Testanalysen, kundenspezifische Normierungen usw. Neben schriftlichen Leistungs- und Wissenstests verfügt die DGP über ein eigenes Postkorbverfahren und ein umfangreiches Repertoire an Assessment-Center-Übungen sowie Eignungsinterview-Modulen zur Erfassung der „soft skills“ wie z. B. soziale Kompetenz, Belastbarkeit, Kommunikationsfähigkeit und Motivation.

3.2.3 Qualitätssicherung

Die Qualitätssicherung in der DGP betrifft den gesamten Prozess der Eignungsbeurteilung, also (1) die Arbeits- und Anforderungsanalyse, (2) die entsprechende Auswahl der diagnostischen Strategie und der diagnostischen Verfahren, (3) die Durchführung und Auswertung von Verfahren einschließlich der Interpretation der Ergebnisse sowie (4) die empirische Evaluation und Qualitätssicherung. Außerdem berücksichtigt die DGP (5) die Perspektive der zu beurteilenden Personen (soziale Akzeptanz).

(1) Arbeits- und Anforderungsanalyse

Konzeptioneller Ausgangspunkt von Personalentscheidungen ist das aus der Arbeits- und Anforderungsanalyse abgeleitete Anforderungsprofil (s. „Grundsätze“ der DIN, DIN 2002, S. 6). Als Beispiel für das Vorgehen der DGP können die Anforderungsanalysen für den Polizeidienst herangezogen werden. Zur Realisierung von Methodenvielfalt wurden bei der Erarbeitung der Anforderungsprofile drei verschiedene Ansätze genutzt, die arbeitsplatzanalytisch-empirische, die personbezogen-empirische und die erfahrungsgelitet-intuitive Methode (Schuler 1988, S. 202f). Ergebnisse zur arbeitsplatzanalytischen Methode lagen aus der Studie von Althoff (1974) vor, bei der die deutsche Bearbeitung des Position Analysis Questionnaire (Frieling und Hoyos 1978) zur Untersuchung von Polizeiarbeitsplätzen eingesetzt wurde. Dieser verhaltensorientierte Ansatz wurde durch Ergebnisse mehrerer polizeispezifischer DGP-Bewährungskontrollen (z. B. Althoff 1977, Greif 1972, Jäger und Althoff 1994, Kersting 1999, Kersting 2001, Wolf, 1990) um einen personenzentrierten Ansatz ergänzt. Auf der Grundlage der Studien wurden die Merkmale erfolgreicher Polizisten ermittelt. Diese Merkmale wurden dann in ein Anforderungsprofil für zukünftige Polizisten transformiert. Um auch mögliche Veränderungen der Anforderungen einzubeziehen, wurden die Ansätze schließlich noch um die erfahrungsgelitet-intuitive Methode ergänzt. Dazu wurde das erstellte Anforderungsprofil in Workshops polizeiinternen Experten zur Begutachtung und Modifikation vorgelegt.

(2) Auswahl der diagnostischen Strategie und diagnostischen Verfahren

Anhand der Kenntnisse und Erfahrungen der Berater sowie vor allem auf Basis der Ergebnisse empirischer Bewährungskontrollen wird für jeden Auftrag festgelegt, welche Strategie und welche Verfahren zur Erfassung der jeweiligen Anforderungsmerkmale und zur Beantwortung der Eignungsfrage eingesetzt werden. Im Interesse eines effizienten und kostengünstigen Verfahrens, das die beurteilten Personen möglichst gering belastet, wird häufig eine zweistufige Strategie gewählt. Zunächst werden Tests in Gruppenverfahren durchgeführt. Teilnehmer, die hier einen festgelegten Mindestwert nicht erreichen, werden zum zweiten, mündlichen Teil (Eignungsinterview, Assessment-Center-Übungen usw.) nicht mehr eingeladen. Diese Pre-reject-Strategie spart der Organisation Kosten und entspricht den in der DIN 33430 formulierten Leitsätzen (S. 12f), wonach Kandidaten nicht mehr beansprucht werden sollten, als für den Untersuchungszweck erforderlich ist.

(3) Durchführung und Auswertung von Verfahren einschließlich der Interpretation der Ergebnisse

Die Logistik der DGP ermöglicht es, die Eignung von Hunderten von Personen zeitgleich zu untersuchen, wobei die Untersuchungen „vor Ort“, also beim Kunden, stattfinden können. Die Erfahrung der Berater und die Standardisierung des Verfahrensablaufs stellen sicher, dass der Blick für die einzelnen Personen trotz der hohen Effizienz der Vorgehensweise nicht verloren geht. Dabei schafft geschulten Beratern gerade die hohe Standardisierung der Situation die notwendigen Freiräume, um auf Einzelpersonen achten und reagieren zu können. Weil die Reihenfolge der Verfahren, die Instruktionen, die zeitlichen Vorgaben, die Reaktion auf mögliche Rückfragen oder vorhersehbare Störungen, das komplette Vorgehen der Durchführung und der Auswertung vorab festgelegt und eingeübt wurden, bleibt der Kopf der Berater frei für das individuelle Gegenüber und für die Aufgabe der Eignungsbeurteilung. Aus diesem Grunde – sowie

zum Zwecke der Fehlerminimierung – erfolgen möglichst viele Schritte der Eignungsbeurteilung bei der DGP computergestützt. Um bei der Auswertung ein Höchstmaß an Effizienz und Auswertungsobjektivität zu gewährleisten, werden für die schriftlichen Tests computerlesbare Antwortbogen und computergesteuerte Auswertungsprogramme genutzt. Die Interpretationsobjektivität wird erhöht, indem der Computer einen nach dem Modell der statistischen Urteilsbildung generierten Vorschlag zur Bewertung der Testergebnisse errechnet (Expertensystem). Bei Aufgaben, die der Computer nicht wahrnehmen kann, wird die Objektivität durch das Mehraugenprinzip gesichert. So werden die im Kontext von Verfahren zur mündlichen Informationsgewinnung, Verhaltensbeobachtung und -beurteilung anfallenden Bewertungen in der Regel von zwei Beratern abgegeben (Prinzip der kontrollierten Subjektivität).

(4) Evaluation und Qualitätssicherung

Die DGP verwirklicht eine „kontrollierte Praxis“, durch die berufliche Eignungsbeurteilungen begründbar und gegenüber Dritten legitimierbar werden. Da die Auswertung der Verfahren und die Datenverwaltung computergestützt erfolgen und Jahr für Jahr über 20.000 Personen untersucht werden, ist es leicht möglich, jederzeit aktuell die in der DIN thematisierten Gütekriterien (z. B. die Zuverlässigkeitskennwerte) sowie zahlreiche andere Qualitätskennziffern der Verfahren zu berechnen und zu dokumentieren. Die entsprechenden Kennwerte werden analysiert und – den Regelungen der DIN entsprechend – veröffentlicht (z. B. Kersting und Beauducel 1997, 2001). Die Berechnungen sind kein wissenschaftlicher Selbstzweck, sondern die in der DIN geforderten Analysen sind Ausgangspunkt regelmäßiger Verfahrensoptimierungen. Insbesondere für die Berechnung solider, differenzierter und aktueller Normen (s. Kapitel 5.1) die den Vergleich einzelner Ergebnisse mit Referenzpopulationen ermöglichen, ist es notwendig, auf die (anonymisierten) Daten einer großen Anzahl an Personen zurückgreifen zu können. Besonderen Wert legt die DGP auf die empirische Überprüfung der Treffsicherheit ihrer Eignungsaussagen. Solche Überprüfungen werden viel zu

selten durchgeführt. Diesbezüglich stellt die DGP nach Hossiep (1995, S. 76) eine „rühmliche Ausnahme“ dar: Keine andere Organisation Deutschlands hat so viele empirische Bewährungs-kontrollen durchgeführt wie die DGP. In Abschnitt 3.2.5 wird eine Treffsicherheitsüberprüfung exemplarisch dargestellt.

(5) Berücksichtigung der Perspektive der zu beurteilenden Personen

Die wichtigsten Personen bei der Eignungsbeurteilung sind die internen oder externen Bewerber. Durch Professionalität und Transparenz sorgt die DGP für eine möglichst hohe soziale Akzeptanz der Auswahlverfahren. So erhalten z. B. alle Teilnehmer vor dem Verfahren eine umfangreiche Broschüre mit allgemeinen Informationen über Eignungsbeurteilungen sowie einen beispielhaften Ablaufplan des schriftlichen und mündlichen Verfahrens. Der Großteil der Broschüre ist Beispielaufgaben zum schriftlichen Test vorbehalten und regt die Teilnehmer zur eigenständigen Übung an. Die Broschüre informiert abschließend darüber, welche Rechte Bewerber haben und wie die Bewerber ihre Rechte wahren können. Auszüge aus der Broschüre können unter der Web-Adresse <http://www.DGP.de> eingesehen werden. Die Broschüre zielt auf die Steigerung der sozialen Akzeptanz, sie dient aber auch der Erhöhung der Chancengleichheit, da sich mit der Broschüre alle Bewerber – und nicht nur die Leser so genannter Testknackerbücher – vorbereiten können. Aufgrund der eindeutig positiven Resonanz zu dieser Informationsbroschüre kann das Vorgehen der ausführlichen Vorabinformation als Standard empfohlen werden.

Wie in den Leitsätzen der DIN 33430 (S. 13) empfohlen, können Bewerber und Berater im Laufe des Verfahrens miteinander kommunizieren. Alle Aufgaben werden zunächst ausführlich erklärt, bei Testaufgaben wird das Verständnis durch Beispiele gesichert.

Nach dem Verfahren können sich die Bewerber auf Wunsch gegen eine geringe Gebühr ihr Verfahrensergebnis von einem Berater in einem persönlichen Feedbackgespräch ausführlich erläutern lassen. Dies sichert die Transparenz und die soziale

Akzeptanz, ohne dass der Auftraggeber/Kunde mit zeitlich und finanziell aufwendigen Feedbacks belastet wird.

Die soziale Akzeptanz der DGP-Verfahren wird regelmäßig in empirischen Untersuchungen überprüft (z. B. Kersting 1998).

Qualifiziertes Personal und wissenschaftliche Fundierung

Voraussetzung für die Qualität in allen fünf genannten Phasen des Prozesses der Eignungsbeurteilung ist einerseits qualifiziertes Personal und andererseits eine wissenschaftliche Fundierung der Arbeit.

Alle Berater der DGP sind Diplom-Psychologen. Diese Qualifikation wird durch spezifische Einarbeitungsprogramme sowie durch regelmäßige Fortbildungsveranstaltungen ergänzt.

Die Arbeit der DGP beruht auf einer engen Beziehung zwischen Anwendung und Wissenschaft. So kommt es, dass DGP-Arbeiten auch über die DGP hinaus Bedeutung erlangt haben. Beispielsweise hat Prof. A. O. Jäger seine Untersuchungen über die „Dimensionen der Intelligenz“ (Jäger 1962) in der DGP begonnen. Jägers „Berliner Intelligenzstrukturmodell“ ist gerade für die Eignungsbeurteilung wichtig. Schuler (1996, S. 105) führt aus, dass „...sich die Vielfalt der in der eignungsdiagnostischen Praxis eingesetzten Fähigkeits- und Leistungstests auf die Grundfaktoren des Berliner Intelligenzstrukturmodells“ zurückführen lassen. Der aus Jägers Untersuchungen hervorgehende Test gilt Amelang und Zielinski (2002, S. 20) als Meilenstein in der Geschichte der psychologischen Diagnostik. In der DGP wurden – so zwei weitere Beispiele – umfassende empirische Vergleiche verschiedener Intelligenztests vorgenommen (Greif 1972) und zahlreiche empirische Untersuchungen zur Treffsicherheit eignungsbeurteilender Verfahren realisiert. Die wissenschaftliche Orientierung der DGP kommt den Kunden und den beurteilten Personen unmittelbar zugute. In der Praxis häufig gestellte Fragen, wie z. B. die nach der Bedeutung von Schulnoten für den Ausbildungs- und Berufserfolg (Althoff 1984) oder die Fairness gegenüber Bewerbern aus Ost- und Westdeutschland (Kersting, 1995, 1996), können aufgrund entsprechender Forschungsarbeiten empirisch fundiert beantwortet werden. Ermöglicht wird die wis-

senschaftliche Fundierung der DGP-Arbeit durch enge Kooperationen mit Universitäten. Bis zu seinem Tod hat die DGP mit Prof. Jäger (Freie Universität Berlin) kooperiert, aktuell bestehen Kooperationen mit den Universitäten Aachen, Gießen und Osnabrück.

3.2.4 Kriterienbezogene Gültigkeit berufsbezogener Eignungsbeurteilungen

Verfahren zur berufsbezogenen Eignungsbeurteilung gibt es viele; empirische und nachprüfbare Beweise der Treffsicherheit der aufgrund dieser Verfahren formulierten Eignungsbeurteilungen sind hingegen rar. Tatsächlich gibt es Unternehmen/Organisationen, die seit Jahrzehnten Verfahren zur Eignungsbeurteilung einsetzen, ohne die Aussagekraft dieser Verfahren jemals geprüft zu haben. Man verlässt sich hier auf das subjektive Evidenzgefühl. Das aber ist irrational und widerspricht nicht nur wissenschaftlichen, sondern auch wirtschaftlichen Gesichtspunkten (s. Kapitel 2.3). Konsequenterweise fordert die DIN 33430 (S. 16), die Kennwerte zur Gültigkeit in den Verfahrenshinweisen bzw. in einer Auftraggebern, Kandidaten und interessierten Fachvertretern zugänglichen Publikation zu dokumentieren.

Seit 1960 (eine Vorhersage der Ergebnisse in einem Refa-Lehrgang anhand der Ergebnisse im Auswahlverfahren; Jäger 1960) bis heute (Überprüfung der Aussagekraft eines zur Personalauswahl bei Führungskräften eingesetzten Assessment Centers; Kersting, im Druck) überprüft die DGP kontinuierlich die Gültigkeit ihrer Verfahren. Im Folgenden soll eine der zahlreichen Bewährungskontrollen beispielhaft dargestellt werden.

Die Logik von Überprüfungen der Kriteriumsgültigkeit ist leicht nachzuvollziehen. Kriteriumsgültigkeit wird in der DIN 33430 (S. 21) definiert als Grad der Übereinstimmung des mit einem eignungsbeurteilenden Verfahren erzielten Ergebnisses einer Person mit der Ausprägung dieses Kriteriums in einem empirisch ermittelten Kriterium. Ein „Kriterium“ ist der DIN 33430 (S. 22) zufolge ein Maßstab, anhand dessen das infragestehende Personenmerkmal im Alltag beurteilt wird. Man schaut also, wie die

Eignung einer Person im Eignungsinterview, im Assessment Center oder im Test beurteilt wurde (das ist der so genannte Prädiktorwert) und vergleicht diese Einschätzung mit einer unabhängigen Beurteilung des berufsbezogenen Erfolgs dieser Person (Kriterium). Beispielsweise stellt man die Ergebnisse des Auswahlverfahrens der späteren Produktivitätsrate oder der Leistungsbeurteilung durch den Vorgesetzten gegenüber. Statistisch wird der Grad der Übereinstimmung in einer Korrelation ausgedrückt, die eine Je-desto-Aussage darstellt – z. B. „je positiver die Eignungsbeurteilung ausfällt, desto höher auch die Zufriedenheit des Vorgesetzten mit dem neuen Mitarbeiter“.

Wenn vermeintliche Untersuchungen zur Kriteriumsgültigkeit so kompliziert dargestellt sind, dass man nicht recht weiß, was eigentlich genau geprüft und gerechnet wurde, so sollte dies nicht als Indikator für Wissenschaftlichkeit interpretiert werden, sondern Anlass zu Misstrauen geben. Abstand nehmen sollte man von Unternehmen/Organisationen, die eine Überprüfung der Qualität ihrer Verfahren behaupten, aber nicht belegen können. Beliebt sind allgemeine Formeln wie „die Treffsicherheit konnte immer wieder nachgewiesen werden“, „... wurde in zahlreichen Studien überprüft“ oder „... das Verfahren hat sich bewährt“. Solche Allgemeinplätze sind belanglos. Teilweise werden sogar Zahlen (Trefferquoten, prozentuale Übereinstimmungen und Korrelationskoeffizienten) berichtet, ohne dass die zugrunde liegende Auswertung irgendwo einsehbar dokumentiert ist. (Oder es wird auf unzugängliche firmeninterne „Arbeitsberichte“, „Präsentationen“ und „Rechenschaftsberichte an die Geschäftsleitung“ sowie auf unzugängliche Diplomarbeiten verwiesen). Um diesem Misstand Einhalt zu gebieten und den Anwendern eine Kontrolle der Qualität von Angeboten zu ermöglichen, fordert die DIN 33430 (S. 15f), dass die Kennwerte zur Gültigkeit in einer zugänglichen Publikation zu dokumentieren sind und dass die Darstellung der zugrunde liegenden empirischen Arbeit den üblichen Kriterien für wissenschaftliche Publikationen entspricht.

3.2.5 Eine konkrete Bewährungskontrolle

Anhand der Eignungsbeurteilung von Bewerbern für eine Ausbildung im Finanzbereich soll die Logik der Überprüfung der Kriteriumsgültigkeit beispielhaft verdeutlicht werden. (Weitergehende Details der Studie finden sich bei Kersting 1999.) Die aufgrund der Anforderungsanalyse bei den Bewerbern als erforderlich angesehenen Fähigkeiten wurden mit einem schriftlichen Testverfahren (BIS-r-DGP Test; Kersting und Beauducel 2001) sowie mit Assessment-Center-Übungen und Eignungsinterviews geprüft. Die Leistungen der Bewerber im Auswahlverfahren wurden auf der „DPG-Empfehlungsgradskala“ festgehalten. Diese Skala reicht von dem negativen Wert „1“ („den Anforderungen nicht entsprechend“) bis zum positiven Wert „5“ („den Anforderungen entsprechend“). Zwischenstufen sind die Werte „1,5“, „2“, „2,5“, „3“, „3,5“, „4“ bis „4,5“, so dass sich insgesamt neun Skalenstufen ergeben. Bei der Bestimmung des Empfehlungsgrades werden die im Auswahlverfahren geprüften einzelnen Leistungsdimensionen entsprechend dem Ergebnis der Anforderungsanalyse unterschiedlich gewichtet.

Zur Abschätzung der Kriteriumsgültigkeit der Eignungsbeurteilungen wird die Verteilung der Ergebnisse des Auswahlverfahrens (DGP-Empfehlungsgrade) der Verteilung des Kriteriums (dem Erfolg in der Zwischenprüfung) in Form einer Kreuztabelle deskriptiv gegenübergestellt (Tabelle 3.2.1). Für jede „Zelle“ der Tabelle sind drei unterschiedliche Angaben verzeichnet, die am Beispiel der ersten Zelle (DGP-Empfehlungsgrad „teilweise anforderungsgerecht (3,5)“ und dem Ergebnis „nicht bestanden“) kurz erläutert werden sollen. Die Angabe in der ersten Reihe dieser Zelle bezieht sich auf die absolute Stichprobengröße von einer Person. Die erste Zahl ist eine absolute Nennung, die in Klammern aufgeführte Zahl der entsprechende prozentuale Wert. Also: *Eine der insgesamt 132 Personen hat im Auswahlverfahren einen DGP-Empfehlungsgrad von „nur teilweise anforderungsgerecht“ erzielt und die Zwischenprüfung „nicht bestanden“. Das sind 0,8 % der Gesamtstichprobe von 132 Personen.* Die Zahl darunter, rechts neben dem Ergebnis der Zwischenprüfung, bezieht sich nur auf die Personen, die in der Zwischenprüfung „durchgefallen“ sind: *11,1 % derjenigen, die in der Zwischenprüfung*

„durchgefallen“ sind, erreichten im DGP-Verfahren einen Empfehlungsgrad von „3,5“. Die letzte, grau hinterlegte Angabe pro Zelle bezieht sich nur auf die Teilnehmer mit einem DGP-Empfehlungsgrad von „3,5“: 50 % derjenigen, die in dem Auswahlverfahren einen Empfehlungsgrad von „3,5“ erzielten, haben die Zwischenprüfung nicht bestanden.

Tabelle 3.2.1: Kreuztabelle

N = 132		Ergebnis des Auswahlverfahren: DGP-Empfehlungsgrad			
Kriterium: Ergebnis der Zwischenprüfung	„3,5“ (nur teilweise anforde- rungsgerecht)	„4“ (weit- gehend anforde- rungsgerecht)	„4,5“	„5“ (anforde- rungsgerecht)	
„nicht bestanden“	1 (0,8%)	3 (2,3%)	5 (3,8%)		
	11,1%	33,3%	55,6%		
	50%	8,6%	7,2%		
„bestanden“	1 (0,8%)	32 (24,2%)	64 (48,5%)	26 (19,6%)	
	0,8%	26,0%	52,0%	21,1%	
	50%	91,4%	92,8%	100%	
	2 (100%)	35 (100%)	69 (100%)	26 (100%)	

Um die Aussagekraft der Eignungsbeurteilung zu bestimmen, ist die zuletzt erläuterte Zahl in den grau hinterlegten Feldern am interessantesten. Die Organisation will erfahren, inwieweit sie sich auf die Eignungsbeurteilungen verlassen kann. Genügt ein Bewerber, der im Auswahlverfahren als „den Anforderungen entsprechend“ klassifiziert wurde, auch tatsächlich den Anforderungen der Ausbildung? Die Kreuztabelle ist ein deutlicher Beleg für die hohe Treffsicherheit der Eignungsbeurteilung: Von den 26 Personen, die im Auswahlverfahren als „den Anforderungen ent-

sprechend“ (Empfehlungsgrad „5“) beurteilt wurden, gab es keine, die den Anforderungen der Zwischenprüfung dann entgegen der Aussage des Auswahlverfahrens nicht genügte. Alle diese Personen bestanden die Zwischenprüfung. Die Eignungsbeurteilung „anforderungsgerecht“ impliziert also quasi eine Garantie für ein erfolgreiches Abschneiden bei der Zwischenprüfung. Eine solche Garantie gibt es natürlich für Personen mit „leichten Einschränkungen“ nicht. Die Empfehlungsgrade „4“ und „4,5“ („entspricht weitgehend den Anforderungen“) signalisieren leichte Einschränkungen in Bezug auf den zu erwartenden Ausbildungserfolg. In der dargestellten Untersuchung entsprachen noch 92 % der Teilnehmer mit diesem Empfehlungsgrad den Anforderungen der Zwischenprüfung.

Eine hohes Risiko ergibt sich bei dem Empfehlungsgrad „3,5“ „nur teilweise den Anforderungen entsprechend“. Aufgrund der Ergebnisse im Auswahlverfahren bestanden hier Bedenken gegenüber einer Zulassung. Die so beurteilten Bewerber sind nicht durchgängig leistungsschwach (das sind die Bewerber mit den Empfehlungsgraden „2“ und „1“, die aber natürlich erst gar nicht eingestellt wurden, s. Abschnitt 3.2.7 zu Streuungseinschränkungen), sie weisen aber neben Stärken auch Schwächen auf. Eine Prognose ist hier schwierig, es sind „unsichere Kandidaten“. Tatsächlich konnte nur eine der beiden so (mit dem Empfehlungsgrad „3,5“) beurteilten Personen die Zwischenprüfung bestehen. Die andere Person hat die Zwischenprüfung im Übrigen nur äußerst knapp (mit 6,87 Punkten) bestanden.

Die Organisation hätte sich die Investition in diese beiden Auszubildenden sparen können, wäre sie dem im Auswahlverfahren ermittelten Empfehlungsgrad gefolgt. Da es insgesamt mehrere hundert Bewerber mit einem Empfehlungsgrad von „3,5“ und schwächer gab (s. unten), ist davon auszugehen, dass durch das Auswahlverfahren Fehlinvestitionen in Millionenhöhe vermieden wurden (s. Kapitel 2.3).

Während in der Tabelle 3.2.1 das Kriterium in dichotomisierter Form („bestanden“ versus „nicht bestanden“) analysiert wurde, werden die Zwischenprüfungsergebnisse in Tabelle 3.2.2 in den differenzierten Notenstufen „sehr gut“ bis „ungenügend“ betrachtet.

Tabelle 3.2.2: Kreuztabelle

N = 132		Ergebnis des Auswahlverfahren: DGP-Empfehlungsgrad			
Kriterium: Ergebnis der Zwischenprüfung		„3,5“ (nur teilweise anforde- rungsgerecht)	„4“ (weit- gehend anforde- rungsgerecht)	„4,5“	„5“ (anforde- rungsgerecht)
„gut“			2 (1,5%)	6 (4,5%)	5 (3,8%)
			15,4%	46,2%	38,5%
			5,7%	8,7%	19,2%
„befriedigend“			14 (10,6%)	39 (30,0%)	18 (13,6%)
			19,7%	54,9%	25,4%
			40,0%	56,5%	69,2%
„ausreichend“			1 (0,8%)	17 (12,8%)	20 (15,2%)
			2,4%	41,5%	48,8%
			50%	48,6%	29,0%
„mangelhaft“			1 (0,8%)	2 (1,5%)	4 (3,0%)
			14,3%	28,6%	57,1%
			50%	5,7%	5,8%
			2 (100%)	35 (100%)	59 (100%)

Tabelle 3.2.2 zeigt, dass ein besserer Empfehlungsgrad mit einer besseren Note in der Zwischenprüfung einhergeht. 84,7 % aller Personen, die in der Zwischenprüfung eine Spitzenleistung erbracht haben (Note „gut“), konnten einen Empfehlungsgrad von „4,5“ oder „5“ aufweisen. 5,8 % aller Personen mit dem Empfehlungsgrad von „4,5“ erzielten jedoch in der Zwischen-

prüfung die Note „mangelhaft“ – diese Note kommt unter den Personen mit dem besten Empfehlungsgrad, nämlich „5“, nicht vor. Umgekehrt gibt es keine einzige Person mit dem Empfehlungsgrad „3,5“, die in der Zwischenprüfung eine gute oder auch nur befriedigende Note erzielt hat.

In einem weiteren Analyseschritt wurde der Zusammenhang zwischen dem Auswahlverfahren und der Zwischenprüfung als Korrelation bestimmt. Gegenüber der Kreuztabelle 3.2.2 wird das Präzisionsniveau der Analyse noch einmal erhöht, da nicht nur die Note, sondern der feiner gestufte gerundete Punktwert der Zwischenprüfung (1 [schlechtester Wert] bis 15 Punkte) analysiert wird. Der als Maß des Zusammenhangs bestimmte Korrelationskoeffizient kann von $-1,0$ (vollständig negativer Zusammenhang) bis $+1,0$ (vollständig positiver Zusammenhang) variieren. Ein Koeffizient von $0,0$ zeigt an, dass kein Zusammenhang zwischen den Indikatoren besteht. In der vorliegenden Untersuchung war der Empfehlungsgrad zu $r = .36$ statistisch bedeutsam mit dem in der Zwischenprüfung erzielten Punktwert assoziiert.

3.2.6 Bewährungskontrollen als Voraussetzung für Verfahrensoptimierungen

Empirische Kontrollen der Aussagekraft von Eignungsbeurteilungen sind keine lästige, von der DIN aufgebürdete Pflicht, sondern sie sind vor allem der Garant dafür, dass Fehler im Verfahren erkannt und beseitigt werden, dass Veränderungen zeitnah berücksichtigt werden und dass das Verfahren kontinuierlich verbessert wird.

Bei den Feinalysen zu dem hier skizzierten Beispiel erwiesen sich einzelne Dimensionen des Leistungstests als prognostisch bedeutsam, andere hingegen nicht. Entgegen den Befunden der Anforderungsanalyse war beispielsweise die Einfallsmenge der Bewerber vergleichsweise irrelevant für den Ausbildungserfolg. Auch die Fähigkeit zum Umgang mit figuralem Aufgabenmaterial war für den Erfolg in der Zwischenprüfung weitgehend unbedeutend. Da sich diese Befunde auch in anderen Bewährungskontrollen mit ähnlichen Berufsbildern (zum Teil aber unter

Verwendung anderer Kriterienmaße) abzeichneten, konnte der Leistungstest aufgrund der Analysen optimiert werden. Die neueste Version des Testverfahrens, der so genannte „A-1“-Test, verzichtet auf Aufgaben zur Einfallsmenge und auf Aufgaben mit figuralem Material. Dadurch wird der Test zugleich kürzer und vorhersagegenauer.

Empirische Bewährungskontrollen – so das Fazit – sind eine wesentliche Voraussetzung für die Qualität der Verfahren und für Verfahrensoptimierungen. Leider wird von diesem Instrument der Qualitätssicherung viel zu selten Gebrauch gemacht. Nach Hossiep (1995, S. 73f) wurden in den sechs einschlägigen deutschsprachigen Fachzeitschriften innerhalb eines Zeitraumes von 44 Jahren lediglich 82 Arbeiten zum Thema empirische Kontrollen der Gültigkeit berufseignungsbeurteilender Verfahren veröffentlicht. Die DGP allein hat im gleichen Zeitraum die Ergebnisse von 24 empirischen Bewährungskontrollen publiziert und die Analyseergebnisse für die kontinuierliche Verbesserung ihrer Verfahren genutzt. Keine andere private deutsche Organisation kann hinsichtlich Bewährungskontrollen ein der DGP vergleichbar hohes Engagement aufweisen.

3.2.7 Zur Interpretation von Befunden zur kriterienbezogenen Gültigkeit berufsbezogener Eignungsbeurteilungen

Anhand des konkreten Beispiels sollen einige Rahmenbedingungen erläutert werden, die bei der Interpretation von Befunden zur kriterienbezogenen Gültigkeit berufsbezogener Eignungsbeurteilungen berücksichtigt werden müssen.

Streuungseinschränkungen

Die in den Tabellen 3.1.1 und 3.1.2 dargestellte Verteilung der Empfehlungsgrade verdeutlicht ein grundsätzliches Problem von Bewährungskontrollen. Obwohl es neun Kategorien von Empfehlungsgraden gibt, sind im Wesentlichen nur die drei Kategorien „4“, „4,5“ und „5“ besetzt. Dies ist verständlich, da es

sich bei Bewährungskontrollen in der Regel nicht um die Gesamtgruppe der zum Auswahlverfahren eingeladenen Bewerber handelt, sondern um eine einseitige Auswahl aus dieser Gruppe. Schließlich stellt die Organisation nur Erfolg versprechende Kandidaten ein. In der Begrifflichkeit der Statistik spricht man von einer „abgeschnittenen“ Verteilung und meint damit, dass die als „leistungsschwach“ beurteilten Bewerber nicht in die Analyse einbezogen werden. Vermutlich wären aber gerade die deutlich leistungsschwachen Bewerber in der Ausbildung, im Beruf „gescheitert“. Dies ist ein Grund dafür, dass die berichteten Zusammenhänge zwischen den Ergebnissen von Eignungsbeurteilungen und den berufsrelevanten Kriterien den tatsächlich bestehenden Zusammenhang zumeist deutlich unterschätzen (Althoff 1984).

Einen Anhaltspunkt für die extreme Verteilungseinschränkung erhält man, wenn man die Verteilung der Testergebnisse der zugelassenen Bewerber mit den Ergebnissen in der nicht vorausgewählten Eichstichprobe vergleicht (Kersting und Beauducel 1997). Über die Hälfte (54 %) der Bewerber in der Eichstichprobe erzielten im Test einen Empfehlungsgrad von „3“ oder schlechter. Die Varianz der testbasierten Empfehlungsgrade betrug in der Eichstichprobe 2,047 gegenüber einer eingeschränkten Varianz von 0,154 in der analysierten Gruppe der zugelassenen Personen (diese beiden Varianzwerte werden weiter unten einer Korrektur der Gültigkeitsberechnung zugrunde gelegt). 49,1 % aller Bewerber der Eichstichprobe erzielten auf der von „70“ bis „130“ reichenden Testleistungsskala einen Wert kleiner oder gleich „100“. Dies dürfte auch für die Gesamtgruppe derjenigen Kandidaten gelten, deren Ergebnisse im vorliegenden Bericht analysiert wurden. Der Mittelwert der zur Ausbildung zugelassenen Gruppe liegt demgegenüber fast bei „110“!

Die tatsächliche Aussagekraft des Auswahlverfahrens kann mit der vorliegenden Untersuchung nicht demonstriert werden. Die eigentliche Leistung des Auswahlverfahrens besteht darin, innerhalb der großen Zahl von fast 1000 Bewerbern für 135 Stellen die aussichtsreichen Kandidaten auszuwählen. Da für alle testbasierten Werte die ursprüngliche Varianz aus der Eichstichprobe bekannt ist, lässt sich für die hier analysierte Gruppe mit der o. a. eingeschränkten Varianz eine Korrektur des Gültig-

keitskoeffizienten entsprechend den empirisch vorliegenden Varianzverhältnissen vornehmen (s. Formel 13.29) [correlation corrected for restriction of range, case 1] Guilford 1950, S. 349). Während die einfache Korrelation zwischen dem testbasierten Empfehlungsgrad (also ohne den mündlichen Verfahrensteil, daher ein anderer Wert und eine andere Stichprobengröße als oben berichtet) und dem Ergebnis der Zwischenprüfung $r = .29$ beträgt, ergibt sich eine entsprechend korrigierte Gültigkeit der Testempfehlungsgrade von $r = .74$ ($N = 135$).

Der enorme Unterschied zwischen den Werten $r = .29$ und $r = .74$ verdeutlicht die Sinnfälligkeit der in der DIN 33430 formulierten Regeln zu statistischen Adjustierungen (S. 16). Diesen Regeln zur Folge müssen bei Ergebnissen zur Gültigkeit sowohl die ursprünglich erhaltenen wie auch die korrigierten Kennwerte aufgeführt werden. Das Beispiel zeigt, dass der unkorrigierte Wert die tatsächliche Aussagekraft der Eignungsbeurteilung eher unterschätzt, der korrigierte Wert hingegen die Treffsicherheit der Eignungsbeurteilung unter optimierten Bedingungen beschreibt und somit idealisiert. Leider werden von unseriösen Anbietern entgegen den Regeln der DIN 33430 häufig nur die optimierten Schätzungen angegeben. Neben der Korrektur für Streuungseinschränkungen oder mangelhafter Zuverlässigkeit der Prädiktoren (z. B. Testergebnisse) und/oder Kriterien (z. B. Vorgesetztenbeurteilungen) können auch so genannte multiple Regressionskoeffizienten berechnet werden, die in der Regel höher ausfallen als die einfachen Korrelationen. Sofern die Ergebnisse der multiplen Regressionsanalyse aber nicht in Entscheidungsregeln umgesetzt werden, sind diese Daten auf einem Abstraktionsniveau angesiedelt, welches für die Praxis von bestenfalls sekundärer Bedeutung ist. Die Analyse gibt dann vor allem Auskunft über die theoretisch maximal mögliche Vorhersagbarkeit des Kriteriums, nicht aber über die Treffsicherheit des tatsächlich angewandten Verfahrens. Der einfache Korrelationskoeffizient liefert häufig konkretere Informationen über die Vorhersagegüte der in der Praxis eingesetzten Instrumente. Von daher ist es sinnvoll, dass die DIN zusätzlich zum Bericht der korrigierten Kennwerte auch die Dokumentation ursprünglich erhaltener Kennwerte vorschreibt.

Skepsis gegenüber Kennwerten zur Kriteriumsgültigkeit ist prima facie angezeigt, wenn die Werte deutlich außerhalb des Wertebereichs liegen, der sich bislang in Studien zur Treffsicherheit ergeben hat. Eine sehr große Anzahl nationaler und internationaler wissenschaftlicher Studien aus fast 100 Jahren entsprechender Forschung hat ergeben, dass die Prognose beruflicher Leistungen am besten mit professionellen Intelligenztestverfahren gelingt. Selbst damit sind aber keine perfekten Vorhersagen zu erwarten. In einer großen Untersuchung, die für das amerikanische Ministerium für Arbeit durchgeführt wurde, zeigte sich eine mittlere kriterienbezogene Gültigkeit der auf Grundlage von Intelligenztests vorgenommenen berufsbezogenen Eignungsbeurteilungen von 0,51. Die Werte streuten zwischen 0,23 und 0,58 (s. z. B. Hunter und Hunter 1984, Schmidt und Hunter, 1998 [die zuletzt genannte Übersicht von Schmidt und Hunter liegt auch in deutscher Übersetzung vor, s. Kleinmann und Strauß 1998]). Entsprechende Erfahrungswerte im Rahmen von 0,30 bis 0,55 werden auch im Anhang der DIN 33430 als Orientierungsrahmen berichtet. Werte zur kriterienbezogenen Gültigkeit berufsbezogener Eignungsbeurteilungen, die deutlich höher ausfallen als die bislang berichteten Erfahrungen, sind eher unwahrscheinlich. Im Einzelfall liegt bei extrem hohen Werten der Verdacht nahe, dass es sich hier um die Angabe von statistisch optimierten Schätzungen handelt und der nach der DIN 33430 geforderte Bericht der unkorrigierten Werte fehlt. Dies wäre ein eindeutiger Indikator für einen unseriös arbeitenden Anbieter.

Art und Qualität des Kriteriums

Die kriterienbezogene Gültigkeit berufsbezogener Eignungsbeurteilungen hängt nicht nur von der Qualität des zu prüfenden Verfahrens ab. Auch die Art des Kriteriums und die Qualität der Kriterienmessung beeinflussen den empirischen Zusammenhang. Da die in der vorliegenden Untersuchung genutzten Zwischenprüfungsergebnisse keine perfekte Vorhersage des endgültigen Ausbildungserfolgs erlauben, weist das gewählte Kriterium bereits als Indikator des Ausbildungserfolgs Defizite auf. Mit dem Auswahlverfahren soll aber eigentlich nicht nur der Ausbildungs-

erfolg, sondern letztendlich auch der Berufserfolg vorhergesagt werden. Der aber ist ein überaus komplexes und sicherlich auch uneinheitlich definiertes Kriterium. Unbestreitbar sind gute Ausbildungsleistungen nicht mit dem Berufserfolg gleichzusetzen. Die Vorhersagequalitäten des Verfahrens können anhand des Kriteriums „Zwischenprüfung“ nur eingeschränkt bewertet werden (Kriteriumsdefizienz).

Grundsätzlich muss die Art und die Qualität des gewählten Kriteriums beachtet werden, wenn man die Ergebnisse zur Gültigkeit eines Verfahrens der Eignungsbeurteilung bewerten will. In der Regel gilt:

- Die Vorhersage von Ausbildungsleistungen gelingt besser als die Prognose beruflicher Leistungen (z. B. Hunter und Hunter 1984).
- Die Vorhersagbarkeit der Kriterien nimmt mit dem Subjektivitätsgrad der Kriterien ab.
- Je zuverlässiger die Kriterien sind, desto besser können diese vorhergesagt werden.

Anderen Faktoren kommt hingegen ein geringerer Einfluss zu als ursprünglich vermutet. Die lange präferierte Auffassung, die Kriteriumsgültigkeit sei zu einem hohen Grade (berufs-) situationsspezifisch, konnte beispielsweise im Laufe der Zeit widerlegt werden. Intelligenztests z. B. können bei praktisch allen bislang untersuchten Berufen zur Leistungsprognose beitragen. Auch die Annahme, es sei nur eine Vorhersage kurz- oder mittelfristiger Ereignisse möglich, findet in aktuellen Studien (Hossiep 1995) keine Bestätigung.

Die kriterienbezogene Gültigkeit berufsbezogener Eignungsbeurteilungen variiert nicht nur in Abhängigkeit von der Qualität der Verfahren, sondern auch in Abhängigkeit von der Selektionsquote, von der Qualität und Art der Kriterien usw. Aus diesem Grunde macht es keinen Sinn, „Mindestwerte“ für Gültigkeitskennwerte zu formulieren oder lediglich die Höhe der Gültigkeitskoeffizienten zweier Verfahren miteinander zu vergleichen. Man muss auch die jeweiligen Untersuchungsbedingungen würdigen, die zu den Ergebnissen geführt haben. Deshalb dienen die in der DIN 33430 im Anhang formulierten Richtwerte von 0,30 bis

0,55 für die kriterienbezogene Gültigkeit berufsbezogener Eignungsbeurteilungen auch nur der groben Orientierung und sind kein normativer Bestandteil des Regelwerks. Ein Verfahren, das unter widrigen Umständen noch eine Kriteriumsgültigkeit von immerhin 0,30 erzielt, kann qualitativ hochwertiger und weit aus nützlicher sein als ein Verfahren, das unter optimalen Bedingungen eine Gültigkeit von 0,55 nachweist.

3.2.8 Durch die DIN 33430 initiierte weitere Qualitätsverbesserungen

Die DGP sieht sich durch die DIN 33430 in ihrer methodisch fundierten und kontrollierten Praxis bestätigt. Gleichwohl fordert die DIN 33430 auch von der DGP Anstrengungen zur weiteren Qualitätsverbesserung.

Hinsichtlich der Bestimmung der Zuverlässigkeit gab die DIN 33430 den wesentlichen Ausschlag, auch die *Stabilität* der DGP Verfahren zu untersuchen. Wie die meisten Organisationen hat sich die DGP bislang aus Aufwandsgründen mit der Bestimmung von Zuverlässigkeitswerten aus einer einmaligen Verfahrensvorgabe (z. B. interne Konsistenz) begnügt. Die DIN 33430 (S. 16) fordert für die Erfassung von Eignungsmerkmalen, für die eine zumindest partielle Zeit- und Situationsstabilität angenommen wird, aber die Bestimmung der Retest-Reliabilität. Damit wird geprüft, ob Personen nach einiger Zeit mit Hilfe des Verfahrens noch annähernd genauso beurteilt werden wie beim ersten Mal. Dies ist eine sinnvolle Regelung, und die DIN 33430 hat den notwendigen Druck erzeugt, die hierzu notwendigen Studien in Angriff zu nehmen.

Weiteren Nachholbedarf hatte die DGP zum Zeitpunkt der Veröffentlichung der DIN 33430 hinsichtlich der in der Norm formulierten Anforderungen an die Dokumentation. Die erforderlichen Dokumente waren zum Teil unsystematisch verstreut, zum Teil überhaupt nicht vorhanden. Die seitens der DIN 33430 genannten Dokumentationspflichten werden dabei durchaus als umfangreich und aufwendig bewertet. Gleichwohl lohnt sich die Mühe: Wenn die Dokumentation einmal erreicht ist und standardisiert fortgeführt wird, spart dies langfristig Zeit (z. B. bei

der Einarbeitung neuer Berater oder Back-Office-Mitarbeiter, bei der Vorbereitung von Präsentationen usw.). Reanalysen einiger DGP-Bewährungskontrollen aus den Anfangsjahren der DGP sind aufgrund der mangelhaften Dokumentation und Datenspeicherung nicht mehr möglich. Dieser unwiederbringliche Erkenntnisverlust hätte vermieden werden können, wenn die DGP ihre Verfahren von Anfang an so dokumentiert hätte, wie es die DIN jetzt verbindlich vorschreibt. In einem stressigen Arbeitsalltag mit hoher Auslastung des Personals werden häufig nur die dringlichsten Dinge erledigt, die Dokumentation „bleibt liegen“. Mit der DIN 33430 sind alle Organisationen angehalten, dafür zu sorgen, dass nicht nur die dringlichen, sondern auch die wichtigen Aufgaben realisiert werden.

Häufig besteht auf Seiten der Kunden Skepsis gegenüber Anforderungsanalysen und Evaluationen. Das Wissen über die notwendigen Anforderungen und über die Qualität der Verfahren müssten die Berater doch bereits mitbringen, so die landläufige Annahme. Die DIN 33430 ist der DGP daher eine willkommene Argumentationshilfe, um die Kunden weiterhin von der Notwendigkeit und von den Vorteilen einer systematischen und regelmäßigen Anforderungsanalyse und Verfahrensevaluation zu überzeugen.

Die in der DIN 33430 formulierten Anforderungen an die Qualifikation der Berater sowie an Mitwirkende werden von der DGP erfüllt. Gleichwohl ist die DIN der DGP auch ein Ansporn zu weiterer Qualifizierung des Personals. Die von Hilke im vorliegenden Band (s. Kapitel 3.1) geschilderte Supervision der Fallarbeit im Psychologischen Dienst der Bundesanstalt für Arbeit ist diesbezüglich vorbildlich. Die DGP prüft aktuell, ob sich ein – gegenüber der Bundesanstalt sicherlich „abgespecktes“ – Programm der Supervision in der DGP etablieren lässt.

Insgesamt wird sich die DGP weiterhin bemühen, nicht nur DIN-Standards einzuhalten, sondern aufgrund der gemachten Erfahrungen, der Anforderungen der Kunden und der Einsicht in wissenschaftliche Entwicklungen selbst Standards zu setzen. Dazu wird es notwendig, in einigen Bereichen über die in der DIN 33430 gesetzten Minimalanforderungen hinauszugehen.

Zusammenfassung

Der Beitrag stellt die Möglichkeiten der Qualitätssicherung und -optimierung beruflicher Eignungsbeurteilungen dar. Qualitätssicherung betrifft den gesamten Prozess der Eignungsbeurteilung, also (1) die Arbeits- und Anforderungsanalyse, (2) die entsprechende Auswahl der diagnostischen Strategie und der diagnostischen Verfahren, (3) die Durchführung und Auswertung von Verfahren einschließlich der Interpretation der Ergebnisse sowie (4) die empirische Evaluation und Qualitätssicherung. Die zuletzt genannte Evaluation steht im Mittelpunkt des Artikels. Anhand eines konkreten Praxisbeispiels wird aufgezeigt, wie die kriterienbezogene Gültigkeit von Eignungsbeurteilungen geprüft werden kann und was bei der Interpretation von Gültigkeitsbefunden beachtet werden muss.

Literatur

Althoff, K. (1974). Das Anforderungsprofil für den gehobenen Polizeivollzugsdienst – Schlußfolgerungen für den höheren Polizeivollzugsdienst? (Unveröffentlichtes Manuskript.) Hannover: Deutsche Gesellschaft für Personalwesen.

Althoff, K. (1977). Zusammenhänge zwischen Ergebnissen von Eignungstests und beruflicher Bewährung. *Schriftenreihe der Polizei-Führungsakademie*, 6-27.

Althoff, K. (1984). Zur Aussagekraft von Schulzeugnissen im Rahmen der Eignungsdiagnostik. *Psychologie und Praxis – Zeitschrift für Arbeits- und Organisationspsychologie*, 30, 77-85.

Althoff, K. (1990). Zur Geschichte der Deutschen Gesellschaft für Personalwesen e.V. und ihrer Bedeutung für die Entwicklung der Berufseignungsdiagnostik im Nachkriegsdeutschland. *Diagnostica*, 36, 249-263.

- Althoff, K. (1999). Aus der wechselvollen Geschichte der Deutschen Gesellschaft für Personalwesen e.V. (DGP). *DGP-Informationen**. Hannover: Deutsche Gesellschaft für Personalwesen, 56, 61-81.
- Amelang, M. & Zielinski, W. (2002). *Psychologische Diagnostik und Intervention* (3. Auflage). Berlin: Springer.
- DIN (2002). DIN 33430: Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen. Berlin: Beuth.
- Frieling, E. & Hoyos, C. Graf (1978). Fragebogen zur Arbeitsanalyse (FAA). Bern: Huber.
- Greif, S. (1972). Gruppenintelligenztests. Untersuchungen am WIT, IST, LPS und AIT. Frankfurt/M.: Lang.
- Greif, S. (1995). Geschichte der Organisationspsychologie. In: H. Schuler (Hrsg.). *Lehrbuch Organisationspsychologie* (S. 15-48). Bern: Huber.
- Guilford, J.P. (1950). *Fundamental statistics in psychology and education*. New York: McGraw-Hill.
- Hossiep, R. (1995). Berufseignungsdiagnostische Entscheidungen. Zur Bewährung eignungsdiagnostischer Ansätze. Göttingen: Hogrefe.
- Hunter, J.E. & Hunter, R.F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.
- Irle, M. (1989). Kurt Wilde, der Wegbereiter psychologischer Testdiagnostik nach 1945. In: T. Herrmann (Hrsg.). *Positionen in der Psychologie 1949 und 1989* (S. 104-111). Göttingen: Hogrefe.
- Jäger, A. O. (1960). Zum prognostischen Wert psychologischer Eignungsuntersuchungen. *Psychologische Rundschau*, 11, 160-178.
- Jäger, A.O. (1967). Dimensionen der Intelligenz. Göttingen: Hogrefe.

* (Hinweis: Die drei zitierten Arbeiten aus den „DGP-Informationen“ können bei der Dt. Gesellschaft für Personalwesen e.V. angefordert werden; Stamme Str. 40D; D-30459 Hannover; Tel.: 0511/9 43 93-0; E-Mail: DGPpersonal@aol.com).

- Kersting, M. (1995). Der Einsatz „westdeutscher“ Tests zur Personalauswahl in den neuen Bundesländern und die Fairneßfrage. Auswirkungen der Testleistungsdisparität zwischen Ost und West auf die Auswahlentscheidung. *Report Psychologie*, 20, 32-41.
- Kersting, M. (1996). Ost-West-Leistungsunterschiede in Berufseignungstests in Abhängigkeit von der kulturspezifischen Wirkung einiger Aufgabenmerkmale. *Zeitschrift für Arbeits- und Organisationspsychologie*, 40, 106-117.
- Kersting, M. (1998). Differentielle Aspekte der sozialen Akzeptanz von Intelligenztests und Problemlöseszenarien als Personalauswahlverfahren. *Zeitschrift für Arbeits- und Organisationspsychologie*, 42, 61-75.
- Kersting, M. (1999). Personalauswahl mit dem BIS-r-DGP-Test: Evaluation der prognostischen Validität. *DGP-Informationen**, 56, 43-60.
- Kersting M. & Beauducel, A. (1997). Der neue DGP-Leistungstest auf der Basis des Berliner Intelligenzstrukturmodells: Informationen zu ausgewählten Testgütekriterien und zur Normierung. *DGP-Informationen**, 55, 93-103. Hannover: Deutsche Gesellschaft für Personalwesen.
- Kersting, M. & Beauducel, A. (2001). Der BIS-r-DGP. Berliner Intelligenzstruktur-Test der Deutschen Gesellschaft für Personalwesen. In: W. Sarges & H. Wottawa (Hrsg.). *Handbuch wirtschaftspsychologischer Testverfahren* (S. 103-110). Lengerich: Pabst Science Publishers.
- Kersting, M. (in Druck). Assessment Center: Erfolgsmessung und Qualitätskontrolle. In: St. Höft & B. Wolf. *Instrumente der Personalauswahl und -entwicklung*. Hamburg: Windmühle.
- Kleinmann, M. & Strauß, B. (Hrsg.) (1998). Potentialfeststellung und Personalentwicklung. Göttingen: Verlag für Angewandte Psychologie.
- Schmidt, F.L. & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.

* (Hinweis: Die drei zitierten Arbeiten aus den „DGP-Informationen“ können bei der Dt. Gesellschaft für Personalwesen e.V. angefordert werden; Stamme Str. 40D; D-30459 Hannover; Tel.: 0511/9 43 93-0; E-Mail: DGPpersonal@aol.com).

Schmitt, N., Gooding, R.Z., Noe, R.A. & Kirsch, M. (1984). Metaanalysis of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology*, 37, 407-422.

Schuler, H. (1988). Berufseignungsdiagnostik. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 9, 201-213.

Schuler, H. (1996) *Psychologische Personalauswahl*. Göttingen: Verlag für Angewandte Psychologie.

Wolf, B. (1990). Zusammenhänge zwischen Ergebnissen von eignungsdiagnostischen Verfahren und Ausbildungserfolg – dargestellt am Beispiel der Auswahl von Bewerbern für den gehobenen Polizeivollzugsdienst (unveröffentlichte Diplomarbeit). Göttingen: Georg-August-Universität zu Göttingen.

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