

Kulturfairness in der Personalauswahl

W. Rainer Leenen, Alexander Scheitza & Siegfried Stumpf

Neben „Teilhabe“ und „interkultureller Kompetenz“ ist „Kulturfairness“ einer der Schlüsselbegriffe in der politischen Diskussion um verbesserte Integrationsangebote und um eine „interkulturelle Ausrichtung“ (Barbara John) der öffentlichen Verwaltung. In fast allen Bundesländern hat es in den letzten Jahren politische Initiativen zur sog. interkulturellen Öffnung der Verwaltung gegeben. Nordrhein-Westfalen hat mit dem „Gesetz zur Förderung der gesellschaftlichen Teilhabe und Integration in Nordrhein-Westfalen“ von 2012 und insbesondere mit der Landesinitiative „Mehr Migrantinnen und Migranten in den öffentlichen Dienst – Interkulturelle Öffnung der Landesverwaltung“ (Kabinettsbeschluss vom 21.12.2010) und jährlichen Umsetzungsberichten drei strategische Ziele festgeschrieben: „Der Anteil der Migrantinnen und Migranten im Öffentlichen Dienst soll erhöht, die interkulturelle Kompetenz der Landesbediensteten soll gesteigert und die interkulturelle Öffnung soll landesweit angestoßen werden.“ (MINISTERIUM FÜR ARBEIT, INTEGRATION UND SOZIALES DES LANDES NORDRHEIN-WESTFALEN 2013: S. 5). Da die Erhöhung des Anteils von Personen mit Migrationshintergrund an den Beschäftigten dezidiert „ohne verpflichtende Quote“ und ohne den Gleichbehandlungsgrundsatz verletzende Maßnahmen erreicht werden soll, ist es politisch von besonderem Interesse, Zugangshindernisse und „mögliche Hemmnisse bei der Auswahl und Einstellung von Menschen mit Migrationshintergrund“ (ebenda S. 6) identifizieren und ggf. auch abbauen zu können.

In diesem Beitrag stellen wir zunächst den aktuellen Forschungsstand zum Thema „Kulturfairness“ dar. Anschließend diskutieren wir sowohl verschiedene Erklärungsmöglichkeiten für unterschiedliche Erfolgsquoten von BewerberInnen mit

und ohne Migrationshintergrund als auch Ansatzpunkte zur Verringerung solcher „Spreizungen“. Mit der Empfehlung verschiedener Maßnahmen zur Verbesserung der Kulturfairness in Auswahlverfahren schließen wir unseren Beitrag ab.

1 Forschungskonzepte zu „Kulturfairness“

In der Forschung wird das Problem, dass Test- und Auswahlverfahren aus verschiedenen Gründen Verzerrungen aufweisen können, die die Chancen bestimmter sozialer Gruppierungen positiv bzw. negativ beeinflussen, mit den Begriffen „adverse impact“ und „cultural bias“ diskutiert. Bezeichnenderweise liegen Forschungsergebnisse hierzu vor allem aus Ländern mit einem entwickelten Bewusstsein für die Probleme kultureller Diversität vor (siehe beispielsweise für die Niederlande TE NIJENHUIS, J. & VAN DER FLIER, H. 1999).

„Adverse impact“ ist ein Begriff aus dem US-amerikanischen Arbeitsrecht. Nach den *Uniform Guidelines on Employee Selection Procedures*, einem Dokument der EQUAL EMPLOYMENT OPPORTUNITY COMMISSION ET AL. 1978, wird darunter eine deutlich unterschiedliche Auswahlquote im Rahmen von Einstellungs- oder Beförderungsentscheidungen verstanden, die zu Lasten der Mitglieder einer bestimmten Rasse, einer ethnischen Gruppe oder eines Geschlechts geht¹. „Adverse impact“ meint nicht, dass ein Mitglied der Mehrheit gegenüber einem Mitglied aus

¹ In den *uniform guidelines* heißt es wörtlich: „A substantially different rate of selection in hiring, promotion, or other employment decision which works to the disadvantage of members of a race, sex, or ethnic group.“ Siehe <http://www.uniformguidelines.com>.

einer Minderheitengruppe bevorzugt wird. Es bezeichnet vielmehr das Phänomen einer nicht intendierten Schlechterstellung einer Gruppe durch ein Verfahren, das identische Standards auf jeden in gleicher Weise anwendet. Ein beliebtes Beispiel zur Veranschaulichung ist die in einem Einstellungsverfahren für Feuerwehrleute gestellte Anforderung, einen 50 kg schweren Sack drei Stockwerke hochzutragen. Diese Auswahlanforderung könnte sich zum Beispiel in einer niedrigeren Auswahlquote von Frauen niederschlagen. Der entstehende „adverse impact“ ist nun nicht an und für sich unrechtmäßig. Er wird es erst dann, wenn der Arbeitgeber keinen Bezug zwischen dieser Anforderung im Auswahlverfahren und den Berufsanforderungen herstellen kann. Nach den *Uniform Guidelines* sollte sich ein Arbeitgeber dann (mit diesem Bezug zum Berufserfolg) rechtfertigen müssen, wenn ein ‚substantiell unterschiedlicher‘ Erfolg im Auswahlverfahren festzustellen war. Dieser wurde angenommen, wenn die sog. 80%-Regel verletzt wurde. Wenn in einem Auswahlverfahren zum Beispiel 40 % der männlichen, aber nur 20% der weiblichen Bewerber erfolgreich sind, dann liegt der Quotient aus beiden Erfolgsraten bei 50% und unterschreitet damit die geforderte 80%-Schwelle (vgl. ZEDECK 2010).

„**Cultural bias**“ bezeichnet Verzerrungen, die in Test- und Auswahlverfahren bei einem ethnisch bzw. kulturell diversen Bewerbersample durch selbstverständlich unterstellte kulturelle Annahmen entstehen. Dies passiert typischerweise, wenn vor dem Hintergrund einer bestimmten Kultur ein Verhalten als „normal“ und erwartbar unterstellt wird, das nicht in jedem kulturellen Umfeld als normal angesehen wird. Im Auswahlverfahren für die Feuerwehr könnte eine solche Verzerrung z.B. durch die Anforderung, ein Schwimmzeugnis vorzulegen, entstehen. Schwimmen ist eine nicht in jeder Herkunftskultur von jedermann erwartete Kompetenz. Und erst recht ist der Besitz eines Schwimmzeugnisses überhaupt nicht selbstverständlich. Eine Verzerrung kann sich hier also dadurch einschleichen, dass in einem Auswahl-

verfahren Belege eingefordert werden, die von bestimmten Gruppen des Samples aus systemischen Gründen schwerer zu beschaffen sind (z.B. Urkunden aus dem Herkunftsland) oder in deren kulturellem Milieu als weniger wichtig angesehen werden.

Die Forschungsliteratur unterscheidet grundsätzlich drei Typen von Fehlern bzw. Verzerrungen, die in Test- und Assessmentprozessen unterlaufen können (vgl. VAN DE VIJVER & TANZER 2004: 124 ff). Von einer **Konstrukt-Verzerrung (construct bias)** wird gesprochen, wenn bestimmte Konstrukte, die im Auswahlverfahren gemessen werden sollen, von Mitgliedern unterschiedlicher Kulturen völlig anders verstanden und bewertet werden. Wenn man das Konstrukt ‚Unhöflichkeit‘ beispielsweise durch das Ausmaß an Gesprächsunterbrechungen definiert, dann wird diese Operationalisierung zwischen den USA und Brasilien nicht kulturell äquivalent sein (vgl. LEUNG & VAN DE VIJVER 2008: 150). Oder das Konstrukt „kritisches Denken“ oder „Adaptionsbereitschaft“ könnte von in Europa aufgewachsenen Personen anders verstanden und bewertet werden als von in China sozialisierten Personen.

Methoden-Verzerrung (method bias) ist ein Assessment-Problem, das durch die im Verfahren eingesetzten Methoden ausgelöst wird. Dazu zählt auch die im Verfahren geforderte Sprachfertigkeit. Aus der psychologischen Diagnostik bei Einwandererkindern kennt man den Effekt, dass bilinguale Kinder bestimmte Aufgaben schlechter lösen als einsprachig aufgewachsene Kinder (vgl. SCHÖLMERICH & LEYENDECKER 2009: 434). Der Effekt kann verstärkt werden, wenn Fragen und Arbeitsanweisungen in einer artifiziellen Sprachlichkeit gefasst sind, die dem Sprachstand der ‚second language speaker‘ nicht entsprechen. Generell beeinflusst also die **Sprachlastigkeit** von Auswahlverfahren die Chancen von Minoritätsangehörigen im Vergleich zu Majoritätsangehörigen. Analog zur Sprachlastigkeit kann auch die **Kulturlastigkeit** des Verfahrens kritisch reflektiert werden. Ver-

steckte kulturelle Anforderungen können sich beispielsweise in den in Tests verwendeten Karikaturen oder Zeichnungen verbergen oder in den geforderten Antwortschemata, wenn sie dem üblichen Kommunikationsstil der Respondenten nicht entsprechen. Generell kann auch die allgemeine Vertrautheit mit Testverfahren je nach kultureller Gruppierung stark variieren und zu Ergebnisverzerrungen führen.

Schließlich wird von einer **Item-Verzerrung (item bias)** gesprochen, wenn ein Auswahltest aus einer anderen Sprache übersetzt worden ist und bestimmte Ausdrücke in ihrer ursprünglichen Bedeutung nicht erfassbar sind.

Verzerrungen dieser Art können nicht nur in PC-gestützten oder schriftlichen Befragungen, sondern auch im Assessment-Center auftauchen, in dem die BewerberInnen meist mit simulierten Anforderungssituationen konfrontiert werden. Das können besonders irritierende Situationen oder auch Konkurrenz- oder Stresssituationen sein. Der ‚method bias‘ hinsichtlich der Sprachfertigkeiten könnte im Assessment-Center beispielsweise durch Übungen entstehen, die eine besondere Wendigkeit, Reaktionsschnelligkeit und Treffsicherheit in der Assessmentsprache voraussetzen (z.B. in Rollenspielen oder simulierten Streitgesprächen). Auch über Beobachtungen und Bewertungen der Rater kann sich ein „bias“ in das Verfahren einschleichen. Die Rater müssen interkulturell geschult sein, um sprachliche Transferfehler bei mehrsprachigen BewerberInnen, um unterschiedliche Standards der Selbstdarstellung und Präsentation von Kompetenz, ein kulturspezifisches Erleben, Wahrnehmen und Verarbeiten von Situationen (insbesondere von kritischen Situationen) sowie kulturabhängige Rollen- und Kommunikations-erwartungen erkennen und kulturfair beurteilen zu können. Damit hängt die Qualität des Verfahrens nicht zuletzt auch von der interkulturellen Auswahlkompetenz der Gutachter ab.

2 Hypothesen zu den Ursachen unterschiedlicher Erfolgsquoten

Es ist ausgesprochen verführerisch, empirische Befunde zu unterschiedlichen Erfolgsquoten in Auswahlverfahren direkt in Aussagen über dessen vorhandene oder fehlende „Kulturfairness“ umzumünzen. Das ist natürlich voreilig, weil die Unterschiede allein noch nichts über kausale Zusammenhänge aussagen. Detailliertere Informationen über Besonderheiten der (unterschiedlichen) Bewerber und BewerberInnen sowie über die Spezifik des Auswahlverfahrens können allerdings **Hypothesen** mehr oder weniger stark stützen bzw. zweifelhaft erscheinen lassen.

Eine **erste Hypothese**, die sich bei unterschiedlichen Erfolgsquoten nahelegt, würde die unterschiedlichen **Bildungsvoraussetzungen** von BewerberInnen mit und ohne Migrationshintergrund in den Blick nehmen. Migrationshintergrund wird in unserer Gesellschaft oft auch mit schlechteren Bildungschancen und schlechteren Bildungsabschlüssen in Verbindung gebracht. Die Frage wird also gestellt werden, ob nicht Bildungsunterschiede in einem ursächlichen Zusammenhang zu den unterschiedlichen Erfolgsquoten stehen. Hier wäre demnach zu überprüfen, ob Migrationshintergrund ein von Bildung unabhängiger Kausalfaktor ist.

Eine **zweite Hypothese** lenkt den Blick auf die Tatsache, dass Personen mit Migrationshintergrund in aller Regel zweisprachig oder mehrsprachig sein werden und ein großer Teil dieser Gruppe **Deutsch als Zweitsprache** erworben haben wird. Die Frage, ob Sprache als Einflussfaktor in Frage kommt, lässt sich anhand der Vergleiche von Subtests überprüfen. Sind die Mittelwertunterschiede zwischen den Gruppen bei Subtests mit komplexen sprachlichen Anforderungen deutlich größer als bei Testverfahren, die wenig und einfaches sprachliches Material verwenden, spricht viel für die Sprachfähigkeits-hypothese.

Erklären ließe sich das möglicherweise mit dem **kulturellen Milieu**, in dem Sprache erworben wird. Das betrifft nicht nur die zweisprachig Sozialisierten, sondern auch einsprachig aufgewachsene Personen. In bestimmten sozialen bzw. kulturellen Milieus sind kompliziertere Sprachwendungen (wie der Gebrauch des Konjunktivs oder verschachtelter Satzkonstruktionen) weniger üblich. Auch bestimmte poetische oder altmodische Ausdrücke aus der Hochsprache (wie etwa Feldrain, Aue, Gestade) kommen wenig oder gar nicht zur Anwendung. An diesem Punkt setzte eine behördenübergreifende Arbeitsgruppe an, die das Innenministerium des Landes Niedersachsen 2007 mit dem Auftrag eingesetzt hat, das Thema MigrantInnen im Auswahlverfahren zu analysieren. Besonders die Überprüfung des vierstündigen Eignungstests durch Arbeitsgruppenmitglieder mit Migrationshintergrund war aufschlussreich: Bestimmte Testbestandteile setzten ein kulturelles Sprachrepertoire voraus, über das MigrantInnen weniger als deutsche BewerberInnen verfügten. So stieß man in der Untersuchung des Testverfahrens zum Beispiel auf Worte wie „Lunte“, „Hirse“, „Docht“, „Gesinde“ oder „Spelunke“, die sich für die ‚migrantischen‘ BewerberInnen als Stolperstein erwiesen. In den Tests zur Merkfähigkeit wurden auch alte deutsche Straßennamen wie „Erlengrund“, „Schnekensee“ oder „Weseraue“ verwendet und von den BewerberInnen mit Migrationshintergrund schlechter memoriert als von den deutschen TeilnehmerInnen. BewerberInnen mit Migrationshintergrund bzw. aus bildungsfernen Familien haben nach den Schlussfolgerungen der Arbeitsgruppe auch größere Schwierigkeiten mit Worten wie beispielsweise „Mulde“, „schraffieren“ oder „skurril“². Ganz pragmatisch wurden solche Ausdrücke in den Items des Fragenkatalogs des Einstellungsverfahrens 2009 durch kulturneutrale Begriffe ersetzt. Die Projektgruppe führt in ihrem Ergebnisbericht eine leichte Reduktion der Differenz (z.B. bei den männlichen Bewerbern von 27,2 % im Jahr 2008 auf 20,7 % im Jahr 2009) auf die bereits eingeleiteten Veränderungen zurück

² Im Auswahlverfahren NRW konnten wir vergleichbare Schwierigkeiten bei den Worten „Kahn“ oder „Fibel“ feststellen.

(vgl. NIEDERSÄCHSISCHES MINISTERIUM FÜR INNERES, SPORT UND INTEGRATION 2009: 30).

Erfolgsunterschiede in den Subtests eines Auswahlverfahrens lassen sich allerdings kaum auf einzelne unverständliche Worte oder sprachliche Wendungen zurückführen. Häufig ist die Gesamtanlage eines Subtests bzw. ein bestimmter Denk- oder Problemlösungsansatz der Stolperstein für die ‚migrantischen BewerberInnen‘. DE SOËTE, LIEVENS & DRUART (2012: 400) konstatieren in diesem Zusammenhang, dass es für bestimmte kognitive Testverfahren geradezu kennzeichnend ist, dass sie vergleichsweise hohe Spreizeffekte in der multikulturellen Personalauswahl erzeugen. In der Forschungsliteratur spricht man von einem Dilemma, da sich diese Instrumente generell durch eine hohe prognostische Validität auszeichnen.

Diversity-Validity-Dilemma

wird der Zielkonflikt beim Einsatz von Auswahlinstrumenten genannt, bei dem Instrumente mit valider Erfolgsprognostik gleichzeitig starke Unterschiede im Bewältigungserfolg durch kulturell und ethnisch unterschiedliche Gruppen erzeugen. Solche Instrumente könnten auf der einen Seite eine zuverlässige Hilfe bei der Auswahl von für die Berufsanforderungen geeigneten BewerberInnen sein, hindern den Arbeitgeber aber auf der anderen Seite daran, eine diverse Mitarbeiterschaft zu rekrutieren. Tests zu kognitiven Fähigkeiten gelten generell als die Testinstrumente mit der höchsten prognostischen Validität (vgl. SCHMIDT & HUNTER 1998); sie erzeugen aber gleichzeitig auch die größten Unterschiede zwischen BewerberInnen der Majorität und aus Minderheitsgruppen (vgl. z.B. den Überblick in DE SOËTE, LIEVENS & DRUART 2012). Hinzu kommt, dass gerade für die Berufsgruppe Polizei kognitive Testwerte geringere Zusammenhänge mit dem Berufserfolg aufweisen als dies für zahlreiche andere Berufsgruppen wie z.B. Manager der Fall ist (vgl. SALGADO ET AL. 2003; DE SOËTE, LIEVENS, OOSTROM & WESTERVELD 2013, S. 247).

Seit dem Überblicksartikel von PLOYHART & HOLTZ (2008) werden in der Forschung geeignete Strategien diskutiert, das Diversity-Validity-Dilemma zu unterlaufen oder abzuschwächen. Unter den verschiedenen Lösungsrichtungen, die aufgrund bisher vorliegender Forschungsbefunde besonders viel versprechend zu sein scheinen (vgl. ausführlicher dazu SCHMITT & QUINN 2010 sowie DE SOETE, LIEVENS & DRUART 2012) erscheinen uns die folgenden zwei Strategien am aussichtsreichsten, um die Unterschiedlichkeit der Leistungen von Angehörigen verschiedener kultureller Gruppen zu reduzieren und gleichzeitig die Validität des Verfahrens zu sichern.

(1) Stärkerer Einsatz von Simulationen (z.B. typische Assessment-Center-Aufgaben) oder simulationsnaher Befragungsmethoden (z.B. Situational-Judgment-Tests oder strukturierte Interviewformen wie das situative Interview)

Bei *Situational-Judgment-Tests* und *situativen Interviews* werden BewerberInnen gefragt, wie sie in bestimmten „berufstypischen“ Situationen handeln würden. Generell gelangt man mit diesen Verfahren zu Spreizungswerten, die deutlich unter dem Niveau von kognitiven Fähigkeitstests liegen. Dennoch sind aber auch hier Spreizungen vorhanden, die bezogen auf Cohen's d auf geringem bis mittlerem Niveau liegen. Als Schlüsselement kann hierbei die Realitätsnähe („Fidelity“) der eingesetzten Verfahren betrachtet werden, was sowohl die Aufgaben- bzw. Reizseite betrifft als auch die Reaktionsseite (Stimulus-Fidelity vs. Response-Fidelity). So produzieren z.B. Situational-Judgment-Aufgaben geringere Spreizungen, wenn die zu lösende Situation nicht in Textform beschrieben wird, sondern als Videofilm dargestellt ist (Stimulusseite), und zudem offene Antwortformate statt z.B. Multiple-Choice-Varianten (Reaktionsseite) vorliegen. DE SOETE, LIEVENS, OOSTROM UND WESTERFELD (2013) haben einen nach diesen Prinzipien konstruierten Situational-Judgement-Test im Anwendungsfeld der Polizistenauswahl in den Niederlanden eingesetzt. Ihre Studie zeigt, dass dieses Verfahren nur sehr geringe

Mittelwertunterschiede zwischen BewerberInnen mit vs. ohne Migrationshintergrund produziert (Cohen's $d = 0.14$). Zudem gibt es Hinweise auf die prognostische Validität des Verfahrens. Die empirische Relation zwischen Testleistung und Validitätskriterium scheint nicht von der kulturellen Zugehörigkeit der BewerberInnen (mit vs. ohne Migrationshintergrund) abzuhängen. Dies spricht für die Fairness des Verfahrens.

(2) Reduktion von nicht leistungsbezogener (konstruktirrelevanter) Varianz im Auswahlverfahren (Strategien gegen Sprach- und Kulturlastigkeit)

Hier geht es grundsätzlich darum, diejenigen Varianzanteile in den Prädiktormessungen (z.B. einem Testverfahren), die in keinem Zusammenhang mit dem zu messenden Kriterium (der Bewährung am Arbeitsplatz) stehen, so weit wie möglich zu eliminieren. Unterschiede zwischen den Leistungen im Auswahlverfahren sollten möglichst eng mit Performanz-Unterschieden am Arbeitsplatz korrelieren. Forschungsergebnisse legen nahe, dass der Anteil der Varianz an den Testergebnissen, die irrelevant für die Leistung am Arbeitsplatz ist, durchweg beträchtlich ist (OUTTZ & NEWMAN 2010). Als wichtiger „Treiber“ für konstruktirrelevante Varianz kann die Sprachlastigkeit der eingesetzten Verfahren („verbal load“) angesehen werden. Minoritäten weisen oftmals schlechtere Werte in der verbalen Leistungsfähigkeit auf, so z.B. beim Leseverständnis und beim Ausdruck in Wort und Schrift. Folglich sollten die verbalen Anforderungen, die das Auswahlinstrument den Kandidaten auferlegt, streng auf das Ausmaß beschränkt werden, das die spätere berufliche Tätigkeit erfordert. Die Schwierigkeit liegt hier darin, die tatsächlichen Voraussetzungen für ein effektives berufliches Handeln von den in einer Organisation oder in einem Berufsfeld kolportierten Anforderungen zu unterscheiden. Die Praktiker in einem Berufsfeld erliegen zuweilen einem Mythos besonders erwünschter oder unumgänglich notwendiger beruflicher Anforderungen, der einer nüchternen Anforderungsanalyse nicht standhält. Vorliegende

Untersuchungsergebnisse zeigen, dass Ansätze zur Reduktion der Sprachlastigkeit der Verfahren (z.B. Videodarstellungen statt schriftlicher Situationsbeschreibungen, vgl. Punkt (1)) Leistungsunterschiede zwischen unterschiedlichen kulturellen Gruppen tatsächlich reduzieren können.

Ein weiterer Treiber konstruktivirrelevanter Varianz kann in der ‚Kulturlastigkeit‘ des Auswahlverfahrens („cultural load“) bestehen, das ja oftmals aus Sicht der kulturellen Mehrheit konstruiert wird. Die vorliegenden Befunde zur Wirksamkeit von Maßnahmen zur Reduktion von ‚Kulturlastigkeit‘ geben ein noch unklares Bild ab und weitere Forschungen sind hier erforderlich.

Die Optimierung eines Auswahlverfahrens unter Berücksichtigung der Gesichtspunkte von Adverse Impact und Fairness kann nur dann wirklich umfassend erfolgen, wenn neben differenzierten Analysen zum Abschneiden unterschiedlicher kultureller Gruppen im Auswahlverfahren auch Daten zur Kriteriumsvalidität dieses Verfahrens verfügbar sind und damit die Zusammenhänge zwischen Leistung im Auswahlverfahren und Erfolg am Arbeitsplatz bekannt sind. Liegen Validitätsdaten vor, so kann man z.B. prüfen, inwieweit das Weglassen ganzer Verfahren oder einzelner Subtests, die starke Mittelwertunterschiede hervorrufen, die Validität des Gesamtverfahrens belasten würde.

Im abschließenden Kapitel sollen nun einige generelle Empfehlungen zur Vermeidung von Adverse Impact und Unfairness gegeben werden. Grundsätzlich sollte jedes Auswahlverfahren im Hinblick sowohl auf Validität als auch Fairness evaluiert werden und die hierfür erforderlichen Daten erhoben werden.

3 Empfehlungen zur Sicherung von mehr Kulturfairness in Auswahlverfahren

So viel Sprache wie nötig

Kommunikation – schriftlich wie mündlich – ist für die meisten Verwaltungsberufe das zentrale Werkzeug. Eine gute und sichere Ausdrucksfähigkeit ist für VerwaltungsmitarbeiterInnen für den Aktenverkehr innerhalb und zwischen Behörden unabdingbar. Unverzichtbar ist auch die Fähigkeit, durch Kommunikation mit einem Gegenüber in Kontakt zu treten und Ziele durch den differenzierten Einsatz kommunikativer Mittel zu erreichen. Die häufig in Auswahlverfahren eingesetzten Rechtschreibtests bilden nur einen geringen Teil der im Berufsalltag erforderlichen sprachlichen Fähigkeiten ab. Angesichts der automatischen Rechtschreibprüfung, die fast alle Computerprogramme inzwischen bieten, ist der hohe Stellenwert, der orthografischen Fähigkeiten in vielen Auswahlverfahren beigemessen wird, kaum mehr nachzuvollziehen. Darüber hinaus ist zu hinterfragen, inwiefern die Kenntnis von Begriffen außerhalb des beruflichen Alltagssprachgebrauchs in einem Auswahlverfahren für einen Verwaltungsberuf notwendig ist. Eine Überprüfung von Wortschatz und Rechtschreibfähigkeit sollte daher maßvoll sein und sich an den Aufgaben und Bedingungen einer zukünftigen Tätigkeit orientieren.

So wenig Sprache wie möglich

Unsere Forschungsergebnisse weisen darauf hin, dass die Ergebnisse der Tests zur Erfassung von Intelligenz oder zu Teilbereichen kognitiver Leistungsfähigkeit mitunter stark durch Sprachfähigkeit konfundiert sind. Sprachlastige Testverfahren erfassen offensichtlich weniger das analytische Leistungsniveau eines Kandidaten oder einer Kandidatin, sondern vielmehr seine/ihre Fähigkeit, eine Aufgabenstellung sprachlich schnell und richtig zu erfassen. Wir konnten erkennen, dass sich die Unterschiede zwischen BewerberInnen mit und ohne Migrationshintergrund bei Verfahren mit kurzen, einfachen bzw. selbst-

erklärenden Aufgabenstellungen und mit sprachfreiem Material deutlich verringern. Für die Erfassung aller Anforderungsmerkmale jenseits von Sprachfähigkeit empfehlen wir daher den Einsatz von sprachfreien bzw. sprachreduzierten Verfahren. Unverzichtbare sprachliche Instruktionen und Stimuli sollten eine einfache und klare Sprache verwenden. Um dies zu gewährleisten, bietet es sich an, bei der Konstruktion bzw. Überarbeitung von Instrumenten Nicht-Muttersprachler in die Formulierungen einzubeziehen.

Verhaltensüberprüfung statt schriftlicher Testverfahren

Schriftliche Testverfahren sind besonders beliebt, weil sie erlauben, größere Bewerberzahlen mit überschaubarem Aufwand einer Auswahl zu unterziehen. Sie können auch recht gut Aufschluss über das analytische Leistungsniveau einer Person geben und lassen Wahrscheinlichkeitsaussagen über die zu erwartende Bewältigung beruflicher Anforderungen zu. Einen unmittelbaren Eindruck über die Fähigkeiten einer Person gewinnt man jedoch, wenn reale und typische Anforderungen einer Tätigkeit simuliert und der Kandidat bei der Bewältigung dieser Aufgaben beobachtet wird. Die praktischen und interaktiven Teile eines Assessment-Centers bieten diese Möglichkeit.

Anders als schriftliche Verfahren (traditionell mit Papier-und-Bleistift oder am PC) können in einem Assessment-Center auch Aspekte von Kommunikation erfasst werden, die über Rechtschreibung hinausgehen: die Ausdrucksfähigkeit eines Kandidaten, seine Fähigkeit, mit anderen Personen ins Gespräch zu kommen, auf sie einzugehen und Gespräche auch bei Hindernissen zielführend zu gestalten.

Interkulturelle Schulung von AC-BeurteilerInnen

Auch Assessment-Center sind nicht kulturfrei: In einem AC spiegeln sich idealerweise nicht nur die Anforderungen einer zu besetzenden Stelle wider, sondern auch die von der Organisation für richtig und wichtig gehaltenen Verhaltensweisen und Handlungsorientierungen bei der Berufs-

ausübung. Dies betrifft beispielsweise den Kommunikationsstil, das Auftreten vor Publikum (besonders in Belastungssituationen), den Einsatz von Gesten, Mimik sowie Blickkontakt, aber auch die von einem/einer KandidatIn präsentierten Lösungswege für den Umgang mit Problemen. Werden organisationsinterne BeurteilerInnen eingesetzt, ist zu vermuten, dass diese aufgrund ihrer betrieblichen Sozialisationserfahrung vertraute Verhaltensweisen und Handlungsstrategien bei KandidatInnen anders – und zwar besser – bewerten, als eher ungewöhnliche. Wird die Reproduktion gewohnter Standards zur Messlatte für Erfolg, gerät die Organisation unweigerlich in eine „Homogenisierungsfalle“, die Innovationen und Anpassungen an neue Bedingungen schwer machen. Prozesse wie die interkulturelle Öffnung einer Organisation im Allgemeinen oder eine Steigerung der kulturellen Vielfalt in der Mitarbeiterschaft im Speziellen, sind auf diese Weise kaum realisierbar.

Um kulturelle Vielfalt zu erreichen, ist es nötig, dass im AC das Erreichen der Organisationsziele und weniger die Einhaltung bekannter Lösungswege oder das Präsentieren organisationsüblicher Verhaltensweisen positiv bewertet wird. Aufgrund ihrer (zumindest in Teilen) anderen Sozialisationserfahrung werden BewerberInnen und Bewerber mit Migrationshintergrund mitunter anders mit Problemstellungen umgehen und andere Lösungswege suchen. Auch werden sie sich dabei gelegentlich anders präsentieren und ausdrücken als ihre „bio-deutschen“ Konkurrenten. Um diese Unterschiede nicht als Seltsamkeit und fälschlicherweise als Defizit zu bewerten, benötigen die BeurteilerInnen in ACs Wissen über Kultur und kulturelle Unterschiede. Eine interkulturelle Schulung sollte also Teil der Ausbildung zum/zur BeurteilerIn sein und – wie andere AC-relevante Aspekte auch – einer regelmäßigen Auffrischung unterzogen werden.

Kompensationsmöglichkeiten

Trotz der dargestellten unterschiedlichen Möglichkeiten, Auswahlinstrumente kulturfairer zu machen, wird es kaum gelingen, eine gewisse

Benachteiligung von Personen zu vermeiden, die Deutsch nicht als Erstsprache bzw. „Muttersprache“ gelernt haben. Es stellt sich die Frage, ob sich dieser Nachteil auf irgendeine Art und Weise wieder ausgleichen lässt.

Für alle Tätigkeiten, in denen Fremdsprachenkenntnisse von Vorteil sind, liegt eine Möglichkeit auf der Hand: Viele BewerberInnen mit Migrationshintergrund verfügen über Sprachkenntnisse in so genannten Bedarfssprachen der Öffentlichen Verwaltung. Gegenüber BewerberInnen, die nur in einer Sprachumgebung groß geworden sind (und allenfalls über Schulkenntnisse in sog. Bildungsniveau-Sprachen wie Englisch oder Französisch verfügen), haben BewerberInnen, in deren Herkunftsfamilie italienisch, griechisch, russisch, türkisch oder arabisch gesprochen wurde, hier einen beruflich verwertbaren Vorteil. Dieses Potenzial in den Blick zu nehmen und entsprechende Fremdsprachenkenntnisse im Rahmen der Personalauswahl zu erfassen, bietet vielen BewerberInnen mit Migrationshintergrund die Möglichkeit, eine gewisse Benachteiligung durch die Sprache des Auswahlverfahrens auszugleichen. Wir empfehlen daher, im Anforderungsprofil bei allen BewerberInnen auch Fremdsprachenkenntnisse in unterschiedlichen Bedarfssprachen zu berücksichtigen.

Monitoring der Verfahren

Die Auseinandersetzung mit kultureller Fairness hat in Deutschland noch keine lange Tradition. Für die meisten auf dem Markt befindlichen Auswahlinstrumente liegen keine Untersuchungen zur Kulturfairness vor. Ebenso wenig sind die von größeren Organisationen eingesetzten meist „maßgeschneiderten“ Verfahren auf ihre Adverse-impact-Effekte bei BewerberInnen anderer kultureller Herkunft untersucht. Bis das entsprechende Know-how vorhanden ist und Eingang in die Entwicklung neuer Verfahren gefunden hat, können Organisationen die von uns dargestellten Empfehlungen aufgreifen und ihre aktuellen Auswahlinstrumente entlang unserer Vorschläge überarbeiten. Es sollte dabei genau überprüft werden, wie sich die Ergebnisse eines interkulturell modifizierten Auswahlverfahrens von denen vorhergehender Durchläufe unterscheiden. Die Ergebnisse dieses Monitorings ermöglichen eine fortschreitende Optimierung des Auswahlverfahrens. Darüber hinaus liefert eine systematische Erfassung relevanter Daten hilfreiche Erkenntnisse über die Anfälligkeit von Tests, Subtests und Items für kulturelle Verzerrungseffekte.

LITERATUR

DE SOETE, BRITT, LIEVENS, FILIP & DRUART, CELINA (2012): An update on the Diversity-Validity Dilemma in personnel selection: A review. In: *Psychological Topics* 21, 3, 399-424.

DE SOETE, BRITT, LIEVENS, FILIP, OOSTROM, JANNEKE & WESTERVELD, LENA (2013): Alternative predictors for dealing with the diversity-validity-dilemma in personnel selection: The constructed response multimedia test. In: *International Journal of Selection and Assessment*, 21(3), 239-250.

EQUAL EMPLOYMENT OPPORTUNITY COMMISSION, CIVIL SERVICE COMMISSION, DEPARTMENT OF LABOR, & DEPARTMENT OF JUSTICE (1978): Uniform Guidelines on Employee Selection Procedures. 29 C.F.R. Part 1607, unter: <http://www.ecfr.gov/cgi-bin/textidx?c=ecfr&rgn=div5&view=text&node=29:4.1.4.1.8&idno=29> (abgerufen am 25.05.2014).

LEENEN, W. RAINER, STUMPF, SIEGFRIED & SCHEITZA, ALEXANDER (im Druck): „Interkulturelle Kompetenz“ in der Personalauswahl – Konzeptualisierung und Integration in bestehende Auswahlssysteme. In: K. v. Helmolt (Hg.): Inter-

kulturelle Arbeitsplätze, Arbeitssituationen und Berufsfelder. Stuttgart: ibidem-Verlag.

LEUNG, KWOK & VAN DE VIJVER, FONS J.R. (2008): Strategies for strengthening causal inferences in cross cultural research: The consilience approach, in: *International Journal of Cross Cultural Management* 8, 145-169.

MINISTERIUM FÜR ARBEIT, INTEGRATION UND SOZIALES DES LANDES NORDRHEIN-WESTFALEN (2013): Landesinitiative Nordrhein-Westfalen. Mehr Migrantinnen und Migranten in den Öffentlichen Dienst - Interkulturelle Öffnung der Landesverwaltung. Zweiter Umsetzungsbericht für den Zeitraum 31. Mai 2012 bis 30. Mai 2013. Düsseldorf.

NIEDERSÄCHSISCHES MINISTERIUM FÜR INNERES, SPORT UND INTEGRATION – LANDESPRÄSIDIUM FÜR POLIZEI, BRAND- UND KATASTROPHENSCHUTZ (2009): Weiterentwicklung der interkulturellen Kompetenz in der Polizei des Landes Niedersachsen. Abschlussbericht der Projektgruppe. Unveröff. Ms.

OUTTZ, JAMES L. & NEWMAN, DANIEL A. (2010): A theory of adverse impact. In: Outtz, J.L. (ed.), *Adverse impact: Implications for organizational staffing and high stakes selection*. New York/Hove: Routledge, 53-94.

PLOYHART, ROBERT E. & HOLTZ, BRIAN C. (2008): The diversity-validity-dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. In: *Personel Psychology*, 61, 153-172.

SALGADO, JESUS F., ANDERSON, NEIL, MOSCOSO, SILVIA, BERTUA, CRISTINA, DE FRUYT, FILIP / ROLLAND, JEAN-PIERRE (2003): A meta-analytic study of general mental ability validity for different occupations in the European community. In: *Journal of Applied Psychology*, 88, 1068-1081.

SCHMIDT, FRANK L. & HUNTER, JOHN E. (1998): Messbare Personmerkmale: Stabilität, Variabilität und Validität zur Vorhersage zukünftiger Berufsleistung und berufsbezogenen Lernens. In M. Kleinmann & B. Strauss (Hrsg.), *Potentialfeststellung und Personalentwicklung*. Göttingen: Hogrefe, 16-43.

SCHMITT, NEAL & QUINN, ABIGAIL (2010): Reductions in measured subgroup mean differences: What is possible? In: Outtz, J.L. (ed.), *Adverse impact: Implications for organizational staffing and high stakes selection*. New York/Hove: Routledge, 429-451.

SCHÖLMERICH, AXEL & LEYENDECKER, BIRGIT (2009): Psychologische Diagnostik bei Kindern aus zugewanderten familien, in: Irblich, D. & Renner, G. (Hrsg.), *Diagnostik in der klinischen Kinderpsychologie*, Göttingen, Bern, Wien usw.: Hogrefe.

TE NIJENHUIS, JAN & VAN DER FLIER, HENK (1999): Bias research in the Netherlands: Review and implications, in: *European Journal of Psychological Assessment*, Vol. 15, 2, 165-175.

VAN DE VIJVER, FONS J.R. & TANZER, NORBERT K. (2004): Bias and equivalence in cross-cultural assessment: an overview, in: *Revue européenne de psychologie appliquée* 54, 12, 119-135.

ZEDECK, SHELDON (2010): Adverse impact: History and evolution .In: Outtz, J.L. (ed.), *Adverse impact: Implications for organizational staffing and high stakes selection*. New York/Hove: Routledge, 3-28.