

IV. 2021 - 23. Jahrgang - ISSN 1615-7729

Wirtschafts psychologie

Eignungsdiagnostik für die 2020er-Jahre



Stefan Höft, Siegfried Stumpf, Tim Warszta (Hrsg.)



Pabst Science Publishers

Adverse Impact und Kulturfairness in der Personalauswahl: Empirische Befundlage und Leitlinien für die Praxis

Rainer Leenen¹, Siegfried Stumpf¹, Stefan Höft²

¹ Technische Hochschule Köln

² Hochschule der Bundesagentur für Arbeit (HdBA), Mannheim

Zusammenfassung

Adverse Impact liegt vor, wenn Personalauswahl eine beträchtlich unterschiedliche Auswahlquote zum Ergebnis hat, die zu Lasten der Mitglieder einer sozialen Gruppierung geht. Das Vorliegen von Adverse Impact ist nicht gleichzusetzen mit dem Verstoß gegen Fairnessgesichtspunkte, sollte aber als ernster Hinweis und Impuls verstanden werden, die Fairness des Auswahlverfahrens zu überprüfen. In diesem Beitrag stehen Unterschiede im Abschneiden unterschiedlicher ethnisch-kultureller Gruppen in der Personalauswahl im Mittelpunkt. Hierzu wird ein differenzierter Überblick zum Stand der Forschung gegeben. Es wird dargestellt, inwieweit ethnisch-kulturelle Subgruppendifferenzen auftreten, inwiefern diese von der Auswahlmethodik sowie der spezifisch betrachteten Ethnie/Kultur abhängen und durch welche Maßnahmen Subgruppendifferenzen reduziert und bessere Chancen insbesondere für Minoritäten geschaffen werden können, ohne die Validität des Auswahlverfahrens zu gefährden. Davon ausgehend werden Leitlinien für die Praxis der Personalauswahl formuliert, die dabei unterstützen sollen, die kulturelle Diversität der Belegschaft in einer Organisation zu erhöhen und das Auswahlverfahren fair und valide zu gestalten. Die Leitlinien sind prozessorientiert abgefasst, sie decken von der Anforderungsanalyse über die Konzeption und Durchführung des Verfahrens bis hin zur Evaluierung die erforderlichen Prozessschritte ab und geben Gestaltungsvorschläge zur Reduktion von Adverse Impact und zur Etablierung eines validen und kulturfaireren Auswahlverfahrens.

Schlüsselworte: Diversität, Kultur, Adverse Impact, Kulturfairness, Personalauswahl

Adverse impact and cultural fairness in personnel selection: Empirical findings and guidelines for practice

Abstract

Adverse impact occurs when personnel selection results in a significantly different selection rate to the detriment of the members of a social group. Adverse impact is not the same as a violation of fairness considerations, but it should be understood as a serious indication and work order to review the fairness of the selection process. This paper focuses on differences in the performance of different ethno-cultural groups in personnel selection. At first, a differentiated overview of the empirical findings is given. It is shown to what extent ethnic-cultural subgroup differences occur, to what extent they depend on the selection methodology as well as on the specific ethnicity/culture under consideration. After on, it is explained through which assessments subgroup differences can be reduced and better opportunities created, especially for minorities, without endangering the validity of the selection process. Based on this, guidelines for the practice of personnel selection are formulated to support cultural diversity of the workforce in an organization and to make the selection process fair and valid. The guidelines are formulated in a process-oriented manner. They cover the necessary process steps from the task analysis through to the conception and implementation of the procedure and finally to the evaluation. They provide design suggestions for reducing adverse impact and for establishing a valid and culturally fair selection procedure.

Keywords: Diversity, Culture, Adverse Impact, Cultural Fairness, Personnel Selection

1. Einleitung

Fragen der Kulturfairness gewinnen in Auswahlverfahren zunehmend an Bedeutung. Organisationen wollen nicht nur das für ihre Produktions- oder Dienstleistungsziele geeignete Personal einstellen – sie verfolgen oftmals auch Diversitätsziele: sie wollen kulturelle Minderheiten auf keinen Fall benachteiligen oder sie möchten sukzessive eine stärker kulturell divers zusammengesetzte Belegschaft aufbauen. Vor allem im Fokus der Öffentlichkeit stehende Organisationen, die sich verpflichtet fühlen, eine gesellschaftliche Vorbildfunktion zu übernehmen oder einen erkennbar großen Nachholbedarf bei der Rekrutierung und Einstellung kulturell diversen Personals haben (wie der Öffentliche Dienst in Deutschland; Ette et al. 2021, S. 24), setzen sich aus legitimatorischen Gründen Diversitätsziele. Auch funktionale Aspekte können Grund sein, die kulturelle Diversität der Belegschaft zu erhöhen: Kulturelle Diversität hat für Organisationen ganz offensichtliche Vorteile, wenn sie in einem multikulturellen Kontext agieren und eine multikulturelle Bevölkerung mit Produkten und Dienstleistungen erreichen wollen. Muslimische Bürger werden möglicherweise die Angebote einer nur mit christlichen Mitarbeiterinnen besetzten Beratungsstelle nur zögerlich annehmen (Kersting & Ott, 2016, S. 691). Und nicht zuletzt können auch demographische Entwicklungen Grund sein, Diversitätsziele zu verfolgen. Für den Öffentlichen Dienst in Deutschland zeichnet sich zum Beispiel ab, dass über 50% der aktuell dort Beschäftigten bis zum Jahr 2036 in den Ruhestand treten werden. Eine Studie der Friedrich Ebert Stiftung empfiehlt daher aus arbeitsmarktpolitischen Gründen, bei den notwendigen Neueinstellungen möglichst alle Bevölkerungsgruppen anzusprechen, also die „migrationsbedingte Vielfalt“ der Gesellschaft unbedingt zu berücksichtigen (Baumann et al., 2019, S. 1).

Grundsätzlich gilt, dass sich zunehmende kulturelle Vielfalt für eine Organisation als Entwicklungschance erweisen kann, dass aber auch problematische Aspekte der Diversität beherrscht werden müssen, die vor allem in einem geringeren Ausmaß an Kohäsion bestehen und in der Interaktion mit Fehlkommunikation und Stress einhergehen können (Podsiadlowski 2007, S. 579). Zunehmende Diversität ist also immer beides: Chance und Herausforderung, die in der schrittweisen Herausbildung einer interkulturellen Personal- und Organisationsentwicklung besteht (Leenen et al., 2014). Investitionen in eine entsprechende Personalrekrutierung und in kulturfairer Auswahlprozesse könnte man als Teil dieser Bewältigungsaufgabe verstehen.

Neben zentralen Gütekriterien wie Validität, Reliabilität und Objektivität (Schuler, 2014, S. 45ff. und 367ff.), die jedes Personalauswahlverfahren erfüllen soll, kommt von daher dem Kriterium der *Fairness* zunehmende Bedeutung zu. Fairnessgesichtspunkte resultieren aus dem Prinzip der Chancengleichheit und dem Gleichbehandlungsgrundsatz bzw. dem Benachteiligungsverbot. Görllich und Schuler (2014) unterscheiden bei Personalauswahlentscheidungen folgende Aspekte von Fairness:

- *Statistische Fairness*: Verwendung valider Personalauswahlverfahren, die berufliche Leistung prognostizieren und eine unbegründete Schlechterstellung von Angehörigen spezifischer Subgruppen möglichst vermeiden sollen.
- *Biografische Fairness*: Verzicht auf nicht selbst zu verantwortende biografische Indikatoren, also z.B. keine Berücksichtigung des Berufs der Eltern.
- *Anforderungsbezogene Fairness*: Berücksichtigung nur derjenigen Merkmale bei der Auswahl, die begründet mit dem beruflichen Erfolg in einer Zielposition in Zusammenhang stehen.
- *Prozedurale Fairness*: Aufrichtige Information über die Zielposition, Transparenz des Verfahrens, realistische Handlungs- und Einflussmöglichkeiten für die Kandidat*innen in der Auswahl-situation sowie wahrhaftige und konstruktive Rückmeldung der Ergebnisse.
- *Gesellschaftliche Fairness*: Zur Stellenbesetzung dienen ausschließlich Leistungskriterien als Auswahlgesichtspunkte und nicht soziale Beziehungen oder Nepotismus.

Statistische Fairness ist insbesondere in den USA ein wichtiges Thema der Personalauswahl; dabei geht es vor allem darum, eine Benachteiligung von ethnischen Minderheiten zu vermeiden. „*Kulturfairness*“ ist im Grunde lediglich eine Ausprägung der statistischen Fairness. Der Begriff fokussiert aber weniger auf den Minderheitenstatus als vielmehr auf das Problem, dass in Auswahlprozessen bestimmte kulturelle Selbstverständlichkeiten vorausgesetzt werden können, die nicht in allen sozialen Gruppen „normal“ und selbstverständlich sind. So sind Personen, die mehrsprachig aufgewachsen sind, häufig bestimmte Ausdrücke und altmodische Redewendungen der Mehrheitsgesellschaft nicht vertraut. Testverfahren können mit Denkmustern operieren, die in Akademikerhaushalten, aber nicht in Handwerkerfamilien üblich sind. Interviewende setzen in Auswahlgesprächen Motive und Einstellungen voraus, die ihnen aus ihrer eigenen Biographie plausibel erscheinen, aber nicht in allen in Deutschland lebenden Fa-

milien selbstverständlich sind (Stumpf, Leenen, & Scheitza, 2017). Der Anforderung des Auswahlverfahrens für den Polizeiberuf, ein Schwimmzeugnis vorzulegen, können Personen, die ihre Kindheit in Deutschland verbracht haben, sehr viel leichter nachkommen als Menschen, die zum Beispiel in der Türkei aufgewachsen sind. „Kulturfairness“ bezieht sich also auf Unterschiede, die aus der kulturellen Sozialisation erwachsen, für den Berufserfolg aber irrelevant sind. Ein Auswahlverfahren ist kulturfair, wenn Personen mit unterschiedlicher kultureller Sozialisation im Verfahren gleiche Chancen haben (ausführlich dazu: Leenen & Stumpf, 2021). In der Forschung wird ein solches Problem der Schlechterstellung bestimmter Gruppen durch ein Test- bzw. Auswahlverfahren, das zwar formal dem Gleichbehandlungsgebot folgt, aber real einen nicht intendierten Selektionseffekt hat, unter dem Stichwort „Adverse Impact“ diskutiert. Forschungsergebnisse liegen hierzu bezeichnenderweise vor allem aus Ländern mit stärker entwickeltem Bewusstsein für Diversitätsprobleme vor (wie USA, Kanada oder Niederlande). Der Begriff „Adverse Impact“ stammt aus dem US-amerikanischen Arbeitsrecht und meint, dass ein den Diversitätszielen entgegenstehender Effekt feststellbar ist (Morris & Dunleavy, 2016). Von einem Adverse Impact wird gesprochen, wenn eine oder mehrere soziale Gruppen im Auswahlverfahren erheblich schlechter abschneiden als andere soziale Gruppen. Als Indikator für „erheblich“ („substantial“) wird in der US-amerikanischen Rechtsprechung eine einfache 80%-Regel verwendet, nach der ein Adverse Impact dann vorliegt, wenn die Erfolgchancen der Mitglieder einer Minorität weniger als 4/5 der Erfolgchancen der Mitglieder der Majorität betragen (Zedeck, 2010). Die Feststellung eines Adverse Impact ist als Warnsignal zu sehen und als Anlass zu nehmen, ein Auswahlverfahren kritisch auf versteckte Benachteiligungseffekte zu überprüfen.

Im Folgenden wird aufgezeigt, wie sich die Forschungslage zu den Themen Kulturfairness und Adverse Impact in Auswahlverfahren darstellt. Daraufhin werden Leitlinien entwickelt, die dabei helfen sollen, ein Auswahlverfahren kulturfair zu gestalten.

2. Überblick zum Stand der Forschung

Zum Adverse Impact-Forschungsfeld liegen international zahlreiche Meta-Analysen (z.B. Roth et al. 2001) und Überblicksartikel (z.B. Ployhart & Holtz, 2008; de Soete et al. 2012; Lindsey et al. 2013) vor. Die wesentlichen Ergebnisse sollen kurz skizziert werden.

2.1 Befundlage zu ethnischen Subgruppenunterschieden in Auswahlverfahren

Ethnische Subgruppenunterschiede werden im Forschungskontext üblicherweise anhand des Effektstärkenmaßes Cohen's d beschrieben (Cohen, 1988), das Mittelwertdifferenzen in Standardabweichungseinheiten beschreibt. Ein d von 1,0 bezeichnet eine Mittelwertdifferenz, die genau der Merkmalsstreuung innerhalb der Gruppen und damit einer Standardabweichungseinheit entspricht. Das Vorzeichen wird dabei üblicherweise so verwendet, dass ein positives d für ein besseres Abschneiden der Majoritätsgruppe steht, ein negatives dagegen für das bessere Abschneiden der Minoritätsgruppe. Die Verwendung von Cohen's d als Indikator für Subgruppenunterschiede hat gegenüber Unterschieden in Erfolgs- oder Durchfallquoten den Vorteil, dass Cohen's d unabhängig von willkürlich festsetzbaren Selektionsquoten ist (Kehoe, 2010). Gemäß der üblichen Konvention gilt ein d ab 0,20 als kleine Effektstärke, ab 0,50 als mittlere und ab 0,80 als große Effektstärke (Bortz & Döring, 1995, S. 568). Basierend auf den Übersichten von Bobko und Roth (2013) sowie Roth et al. (2017) werden in Abbildung 1 die primär US-amerikanischen Befunde zum Abschneiden von drei Minoritätengruppen (Black, Asian, Hispanic) im Vergleich zur Majorität (White) für sieben Verfahrensklassen wiedergegeben.

Es zeigt sich, dass die höchsten Unterschiede zwischen Majoritäts- und Minoritätsgruppen bei kognitiven Fähigkeitstests auftreten. Hier wird von Bobko und Roth (2013) etwas konservativer als bei früheren Analysen beim Black-White-Vergleich ein großer Effekt von $d=0,72$ für moderat komplexe Berufe angegeben. Ähnlich fallen die Ergebnisse bei hispanisch-amerikanischen Bewerbern aus ($d=0,72$); asiatisch-stämmige Personen erzielen hingegen bessere Werte als die weiße Mehrheit ab ($d=-0,16$). Die d -Werte zu allgemeinen kognitiven Fähigkeiten liegen oftmals höher als für (Sub-) Tests, die spezifische Intelligenzfacetten messen (Roth et al., 2001; Outtz & Newman, 2010, S. 65). Die Detailanalyse zu Roth et al. (2017) zeigt auch, dass der gefundene Vorteil der asiatisch-stämmigen Personen bei dem untersuchten Studieneingangstest im Wesentlichen auf mathematische Fähigkeiten zurückzuführen ist ($d=-0,45$), während sie im verbalen Bereich schlechter abschneiden ($d=0,16$).

Bei den übrigen Auswahlverfahren fallen die Subgruppenunterschiede geringer aus. Trend ist aber immer, dass die Minoritäten „Black“ und „Hispanic“ schlechter als die „White“-Majorität abschneiden. Bei der „Asiatic“-Gruppe sind die Ergebnisse heterogen. Aufschlussreich ist die Detailanalyse von Bobko und Roth (2013) zum

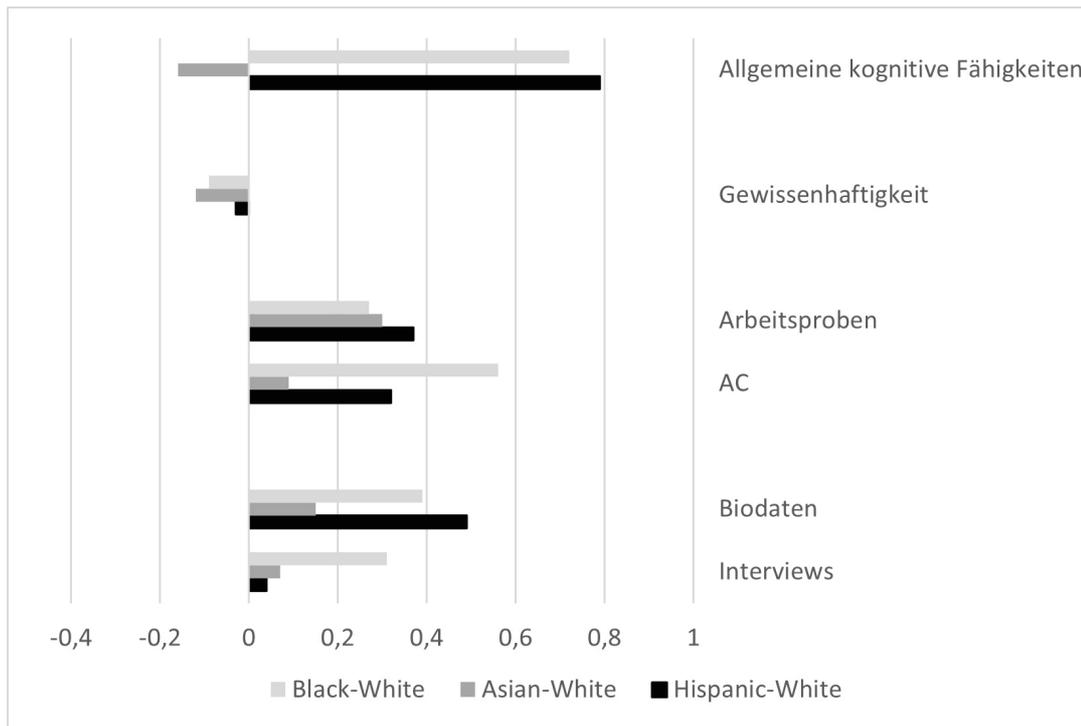


Abbildung 1

Mittelwertunterschiede (d-Werte) im Abschneiden verschiedener ethnischer Gruppen in Abhängigkeit vom Auswahlverfahren

Anmerkungen:
Ausgewählte Ergebnisse
Black-White: Bobko & Roth, 2013, Asian-White und Hispanic-White: Roth et al., 2017)

Positive d-Werte bedeuten, dass die Majorität (White) höhere Mittelwerte aufweist

„Black-White“-Vergleich: Sie weisen auf, dass der d-Wert bei Arbeitsproben (work samples) dramatisch steigt, wenn mehr verbale Inhalte bearbeitet werden müssen. Gleichzeitig ergeben sich im Assessment Center drastisch höhere d-Werte, wenn intelligenznahe Anforderungen erfasst werden.

Die berichteten Befunde basieren fast ausschließlich auf US-Befunden. In Europa stellt sich die Untersuchungssituation anders dar: Bei den untersuchten Minderheiten handelt es sich sehr häufig um Bewerbende mit Migrationshintergrund, d.h. Personen die selbst zugewandert sind oder aus zugewanderten Familien stammen. In einer niederländischen Studie von De Meijer et al. (2006) mit Polizeibewerbern erzielten Migranten der ersten Generation generell deutlich schlechtere Ergebnisse als Personen, die der zweiten Generation zuzuordnen sind. In der Studie von Melchers und Annen (2010) zu Unterschieden von deutsch-/französisch/italienisch-sprachigen Bewerbern im Assessment Center des Schweizer Militärs ergaben sich keine einheitlichen Unterschiede, wobei die Teilnehmenden allerdings die Sprache, in der das Verfahren zu absolvieren ist, aussuchen konnten. Eine größere deutsche Studie zum Thema wurde von Stumpf et al. (2017) vorgelegt. Sie analysierten ethnische Subgruppenunterschiede im Personalauswahlprozess für den gehobenen polizeilichen Dienst in Nordrhein-Westfalen. Die Ergebnisse dieser Studie replizieren bei den Verfahrensunterschieden im Wesentlichen die internationalen Befunde: große Effekte bei Testverfahren zur Intelligenz und zur Rechtschreibung, mittlere Effekte – nach dem Ausscheiden von Kandidaten und Kandidatinnen mit nicht ausrei-

chenden Ergebnissen in den Tests – im sich anschließenden verhaltensorientierten Assessment Center. Die Höhe der Subgruppenunterschiede wird deutlich moderiert durch den Migrationsgrad: Die höchsten Abweichungen zur Majorität liegen für ausländische Staatsbürger vor, während die niedrigsten Abweichungen bei Bewerbenden bestehen, die von Geburt an die deutsche Staatsbürgerschaft besitzen und bei denen meist nur ein Elternteil die ausländische Staatsbürgerschaft hat. Allerdings konnten in der Studie keine Daten zum sozioökonomischen Status der Bewerber, zu ihrer Migrationsgeschichte oder zur Mehrsprachigkeit bzw. zum Lebensalter des Zweitspracherwerbs erhoben werden. Subgruppendifferenzen können aber nicht nur durch Kulturunterschiede bedingt, sondern auch durch spezifische sozioökonomische Lebensbedingungen verursacht sein.

2.2 Mögliche Ursachen für ethnische Subgruppenunterschiede

Die Theorieentwicklung zur Erklärung der gefundenen Adverse Impact-Effekte befindet sich noch in den Anfängen, wobei Outtz und Newman (2010) einen ersten Modellentwurf zur Erklärung ethnischer Subgruppendifferenzen in Auswahlprozessen vorgelegt haben. Dieses Modell berücksichtigt insbesondere so genannte Umweltvariablen (sozioökonomischer Status, familiäre Entwicklungsumgebung, gesellschaftliche Bildungschancen etc.) sowie kulturelle und identitätsbildende Faktoren (Werte, soziale Selbst- und Fremdkategorisie-

rung etc.) und bietet sich als Anleitung zu weiterer kausalorientierter Forschung an. Allerdings ist dieses Modell auf die gesellschaftlichen Rahmenbedingungen der USA ausgerichtet und stellt Subgruppeneffekte beim Vergleich zwischen Schwarzen und Weißen im US-amerikanischen Kontext in den Vordergrund. So bleibt z.B. die Beherrschung der für das Auswahlverfahren maßgeblichen Sprache in dem Modell bisher unberücksichtigt. Gerade dies sollte aber in Europa ein wichtiger Faktor sein. So zeigte sich bei De Meijer et al. (2006), dass die Beherrschung der holländischen Sprache mehr Varianz im Abschneiden beim Auswahlverfahren erklärt als Bildungsunterschiede und Ethnizität.

2.3 Subgruppenunterschiede in Auswahlverfahren: Wann sind sie unfair?

Der Nachweis von Mittelwertunterschieden zwischen ethnischen Gruppen bei eignungsdiagnostischen Auswahlverfahren gibt zunächst einen wichtigen Hinweis auf mögliche Unfairness, ist alleine aber noch kein schlüssiger Beweis dafür. Cleary (1968) verdeutlicht dieses anhand von drei Modellkonstellationen, die in Abbildung 2 dargestellt sind.

In den drei Teilabbildungen wird jeweils der mögliche Zusammenhang zwischen dem Ergebnis im Auswahlverfahren und der späteren Leistung am Arbeitsplatz dargestellt. Vereinfachend wird dabei angenommen, dass Auswahlverfahren und berufliche Leistung jeweils eindimensional über einen Punktwert (Auswahlverfahren: Aggregiertes Gesamturteil, berufliche Leistung: unverfälschter Gesamtwert aller spezifischen Leistungsaspekte) abgebildet werden können. Das Ergebnis jeder Person kann damit als Koordinate in dem zweidimensionalen Diagramm verortet werden. Die Ellipsen (Punktwolken von Personengruppen) deuten die Ergebnisse von jeweils zwei Subgruppen (Kategorie A: z.B. Personen ohne Migrationshintergrund, Kategorie B: z.B. Personen mit Migrationshintergrund). Der generelle Zusammenhang zwischen Auswahlverfahren und beruflicher Leistung wird über die Regressionsgerade gekennzeichnet.

- In Abbildung 2a wird zunächst ein fairer Test dargestellt: Unterschiede im Auswahlverfahren korrespondieren mit entsprechenden beruflichen Leistungsunterschieden. Das Auswahlverfahren bildet also echte Leistungsunterschiede ab und ist damit valide.
- In Abbildung 2b wird eine erste Form eines unfairen Tests dargestellt. Hier korrespondieren Unterschiede im Auswahlverfahren nicht mit Leistungsunterschieden am Arbeitsplatz. Angehörige der Kategorie B erzielen deutlich

schlechtere Verfahrensergebnisse als Angehörige der Kategorie A, obwohl sie die gleiche berufliche Leistung erzielen bzw. erzielen würden. Wenn für das Auswahlverfahren beispielsweise eine mittlere Mindestbestehensgrenze festgelegt werden würde, wären Personen der Kategorie B systematisch benachteiligt.

- Abbildung 2c gibt schließlich eine andere Form der Ungleichheit wieder. Hier erzielen Personen der Kategorie A und B ungefähr gleich verteilte Ergebnisse im Auswahlverfahren. Bei der beruflichen Leistung erzielen Angehörige der Kategorie B aber im Trend höhere Werte und der Zusammenhang zwischen Auswahlverfahren und Leistung ist enger. Somit ist das Verfahren für diese Gruppe valider als für die Personen der Kategorie B.

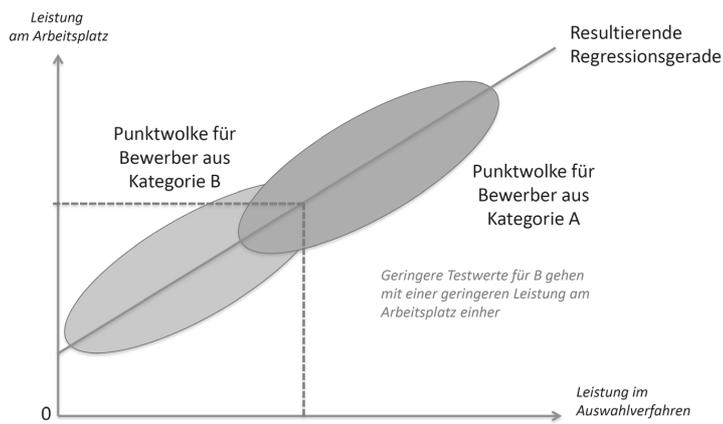
Für die endgültige Bewertung von Adverse Impact als fair oder unfair ist also der Zusammenhang mit der beruflichen Leistung wichtig. Wie bei der Polizeistudie von Stumpf und Kollegen (2017) liegen solche kriteriumsbezogenen Validierungsbefunde aber häufig nicht vor. Ergänzend ist dem Cleary-Modell hinzuzufügen, dass Unfairness nicht nur bei der Wahl der Auswahlverfahren und Prädiktoren bestehen kann, sondern auch bei der Festlegung des Kriteriums für berufliche Leistung: Dies ist der Fall, wenn z.B. beruflicher Aufstieg als Leistungsindikator in Studien verwendet wird und Personen mit Migrationshintergrund aber systematisch bei Beförderungen umgangen werden.

2.4 Ansatzpunkte zur Reduktion von Adverse Impact: Sprache und Kulturabhängigkeit

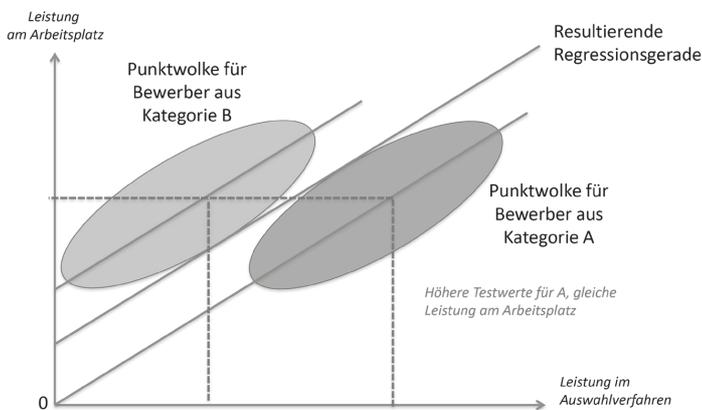
Zwei wesentliche Ansatzpunkte für die Reduktion von Adverse Impact ergeben sich aus folgenden Überlegungen (vgl. Leenen & Stumpf, 2021; de Soete et al, 2012):

- Ein Schlüsselmerkmal für bestehende Adverse Impact-Effekte bei ethnienbedingten Subgruppenunterschieden dürfte in der starken *Sprachgebundenheit* („verbal load“) vieler Auswahlverfahren liegen. Dies ist bei Tests zu verbalen Fähigkeiten (z.B. Wortanalogie-Verfahren oder Verfahren zu komplexem Textverständnis) direkt in der Konstruktmessung begründet. Andere Verfahren sind beispielsweise durch detaillierte Arbeitsanweisungen und Instruktionen verbal eingekleidet.
- Eine zweite Grundlage von Adverse Impact ist in der *Kulturgebundenheit von eignungsdiagnostischen Bewertungen* („cultural load“) zu verorten. Speziell bei interaktiven Verfahren wie Interviews, Gruppendiskussionen und Simulationen

a) Fairer Test



b) Unfairer Test Typ1



c) Unfairer Test Typ2

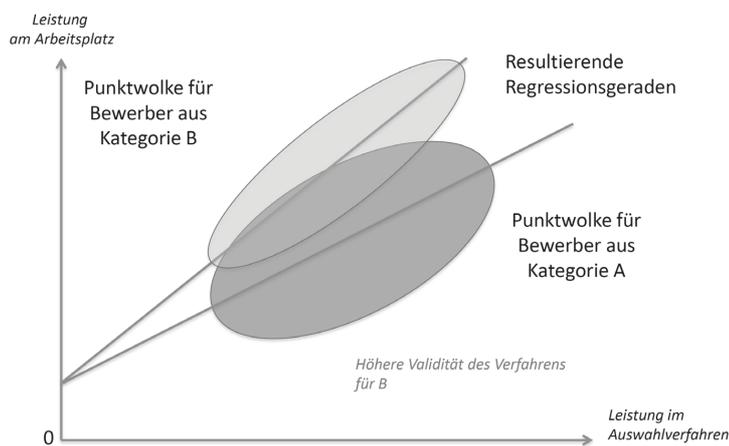


Abbildung 2

Clery-Modell zur Veranschaulichung von (un-)fairen Auswahlverfahren (Darstellung folgt Russell, 2000)

kommen Personen mit einem Beurteilungsauftrag zum Einsatz, die kognitiven Verzerrungen, kulturellen Fehlattributionen, Stereotypisierungstendenzen und Vorurteilen unterliegen können.

Die im Weiteren beschriebenen Maßnahmen zielen zum Teil darauf ab mit veränderlichen Anteilen eine Reduktion von „verbal load“ und „cultural

load“ im Auswahlverfahren zu bewirken. Weitere Empfehlungen greifen das interaktive Moment auf, das z.B. bei Rollenspielen oder im Auswahlinterview gegeben ist, und wirken Verzerrungen entgegen, die in Form kultureller Fehlinterpretationen, Stereotypisierungen und Vorurteile seitens der Beobachterinnen und Beobachter auftreten können.

3. Leitlinien für eine kulturfaire Personalauswahl

Im Folgenden werden am Beispiel der Personalauswahl mit Assessment Centern Leitlinien für eine kulturfaire Personalauswahl formuliert. Das Assessment-Center-Verfahren (AC) bietet sich hierfür an, da es ein komplexes Personalauswahlinstrument ist, das verschiedene Einzelverfahren wie Rollenspiele und Simulationsübungen, ein eignungsdiagnostisches Interview oder psychologische Testverfahren beinhalten kann. Die Darstellung der Leitlinien erfolgt in Orientierung an den Standards der Assessment-Center-Technik des Arbeitskreises Assessment Center / Forum Assessment e.V. (2016), wobei hier der Logik der Standards folgend die Prozessschritte „Anforderungsanalyse“, „Methodenauswahl und Gestaltung des Auswahlverfahrens“, „Beobachtung und Bewertung“, „Entscheidung und Rückmeldung“ sowie „Evaluation und Qualitätssicherung“ unterschieden werden. Für jeden Prozessschritt werden Handlungsempfehlungen formuliert, die darauf abzielen, dass das Auswahlverfahren nicht nur aussagekräftig und damit valide wird, sondern sich auch durch Kulturfairness auszeichnet, nicht begründbare Benachteiligungen bestimmter Gruppen von Bewerbenden vermeidet und die Diversity-Ziele einer Organisation unterstützt. Die Berücksichtigung von Fairnessgesichtspunkten und Diversity-Zielen hat zudem Berührungspunkte mit den grundlegenden AC-Qualitätsstandards der Auftragsklärung und Einbettung in die Organisation, da in diesem vorgängigen Schritt diese Gesichtspunkte bereits angesprochen und als anstrebenswert ausgewiesen werden sollten.

3.1 Anforderungsanalyse

Jede Auswahlentscheidung setzt ein klares Anforderungsprofil voraus, das beschreibt, welche Fähigkeiten, Fertigkeiten, Wissens Elemente oder Orientierungen jemand benötigt, um eine bestimmte Zielposition in einer Organisation erfolgreich ausfüllen zu können (Höft & Kersting, 2017). Eine fundierte Anforderungsanalyse und ein explizites, differenziertes Anforderungsprofil sind eine gute Absicherung gegen die Auswirkungen von impliziten, also nicht bewusstseinspflichtigen Persönlichkeitstheorien im Auswahlprozess. Stereotype und insbesondere Vorurteile (im Sinne verallgemeinerter negativ gefärbter Annahmen über Charakteristika von Personen mit einer bestimmten Gruppenzugehörigkeit) haben weniger Chancen, auf die Beurteilungen durchzuschlagen. Außerdem schützt ein explizites Anforderungsprofil davor, implizite und nicht in der Logik der Tätigkeit verankerte Verhaltenserwartungen in den

Auswahlprozess einfließen zu lassen, die kulturell bedingt sind: Dies betrifft z.B. Erwartungen, inwieweit man in einer Auswahl-situation eigene Vorzüge ins rechte Licht rücken sollte, was zum Beispiel in sog. kollektivistischen Kulturen eher mit Scham behaftet ist (zum Schamkonzept in kollektivistischen Kulturen siehe z.B. Hofstede, 1997, S. 80)

Anforderungsprofile müssen aber nicht zwangsläufig kulturell neutral oder offen sein. Entscheidend ist, worauf es in einer konkreten Zielposition ankommt, was von Tätigkeit zu Tätigkeit, von Organisation zu Organisation und über verschiedene kulturelle Kontexte hinweg variieren kann. Wenn z.B. eine deutsche Auslandsentsandte französische Mitarbeiter in einer Niederlassung in Frankreich führen soll, dann ist ihr Verhalten auch danach zu beurteilen, inwieweit es in diesem französischen Handlungskontext zielführend und effektiv ist (Harris & Lievens, 2005). Effektive Verhaltensweisen werden aufgrund ihrer Einbettung in eine bestimmte Organisationskultur oftmals eine kulturelle Tönung haben. Dies bedeutet aber auch, dass man kulturell getönte Verhaltensweisen nicht in das Anforderungsprofil aufnimmt, wenn kein Zusammenhang zum beruflichen Erfolg gegeben ist.

Aufbauend auf diesen Überlegungen werden folgende Handlungsempfehlungen gegeben:

- Die notwendige Anforderungsanalyse sollte Personen in der Organisation einbeziehen, die das Berufsfeld sehr gut kennen. Personen mit Zuwanderungsgeschichte sollten berücksichtigt werden, da sie möglicherweise einen spezifischen Blick auf diese Anforderungen haben.
- Im Rahmen der Anforderungsanalyse sollte auch geprüft werden, ob es spezifische Ressourcen von Angehörigen von Minderheiten oder Personen mit anderem kulturellen Hintergrund gibt, die für die Zielposition erfolgsrelevant sind (z.B. Mehrsprachigkeit) und in das Anforderungsprofil aufgenommen werden sollten.
- Da Bewerbende aus Minoritäten möglicherweise in ihren Deutschkenntnissen gegenüber der Majorität Nachteile haben, sollte überprüft werden, welche Sprachfähigkeiten in mündlicher Kommunikation und in Schriftsprache tatsächlich vorliegen müssen, um die berufliche Tätigkeit erfolgreich auszuüben.
- Die Präzision und Konkretetheit der Beschreibung erfolgskritischer Verhaltensweisen sollte nicht zu Gunsten einer vermeintlich kulturoffenen und vieldeutigen abstrakten Beschreibung zur Disposition gestellt werden.
- Merkmale wie Teamfähigkeit oder Zielorientierung können je nach Kulturhintergrund anders gedeutet und verstanden werden. Anforderungsmerkmale sollten deshalb gegenüber den Bewerbenden transparent gemacht werden. Dies beginnt mit deren Nennung in einer Stellenausschreibung und reicht hin bis zu ihrer differenzierten Erläuterung zu Beginn des Verfahrens.

3.2 Von der Methodenauswahl zur Ausgestaltung von Auswahlverfahren

Im Rahmen der Vorauswahl können Benachteiligungen für Angehörige bestimmter Kulturgruppen entstehen, wenn die Vorauswahlkriterien nicht kultursensibel festgelegt werden. Außerdem können bereits bei der Bewertung der Bewerbungsunterlagen Urteilsverzerrungen auftreten. Leistungsunterschiede fallen offenbar bei simulationsorientierten Verfahren (z.B. Gesprächssimulationen, Arbeitsproben) schwächer aus als bei kognitiven Fähigkeitstests. Leistungsverzerrungen können auch durch einen Stereotyp-Threat auftreten: Personen können sich demnach in Auswahl-situationen von der Möglichkeit einer stereotypen Wahrnehmung und Bewertung bedroht fühlen, was Ressourcen binden und zu Leistungseinbußen führen kann (Schmader & Johns, 2003).

Handlungsempfehlungen zur Förderung der Kulturfairness in diesem Prozessschritt sind:

- Bereits die Vorauswahl sollte anhand objektiver Auswahlkriterien und kultursensibel ausgestaltet werden.
- Bei der Analyse der Bewerbungsunterlagen sollten anonymisierte Unterlagen genutzt werden (Krause et al, 2012), die ohne kritische Schlüsselreize wie Namen oder Foto auskommen, um Verzerrungen durch Stereotype und Vorurteile entgegenzuwirken.
- Generell sollte im Vorfeld Transparenz zur Art und Form der eingesetzten Verfahren hergestellt werden. Möglichkeiten zur Vorerprobung haben sich speziell für Minoritätengruppen als günstig für eine Leistungsverbesserung erwiesen (Campion et al., 2019).
- Anstelle oder in Ergänzung zu kognitiven Fähigkeitstests oder abstrakten sprachlichen Orthographie- und Grammatiktests sollten verstärkt simulationsorientierte Verfahren eingesetzt werden. So empfehlen Stumpf et al. (2017) für das Auswahlverfahren der Polizei, die schriftliche Kommunikationsfähigkeit nicht mit einem abstrakten Rechtschreibtest, sondern über eine Arbeitsprobe (Vernehmungsprotokoll) zu erfassen.
- Alle Verfahren sollten der Arbeitsrealität möglichst nahekommen: höhere Realitätsnähe („Fidelity“) führt zu geringeren Subgruppendifferenzen (vgl. de Soete, 2012). Dies gilt sowohl für die Realitätsnähe der Aufgaben (Stimulus Fidelity) als auch für die Realitätsnähe des Kandidatenverhaltens (Response Fidelity). Verfahren sollten komplexe Verhaltensreaktionen ermöglichen und nicht in einem einfachen Ankreuzen von Antwortalternativen bestehen.
- Für die berufliche Leistung irrelevante Varianzanteile im Auswahlverfahren sollten vermieden werden. Hierzu gehören neben der sensiblen

Handhabung der „kulturellen Ladung“ des Verfahrens (z.B. über vorausgesetztes kulturelles Wissen in Bildern, Cartoons, Karikaturen etc.) auch die Verringerung irrelevanter sprachlicher Anforderungen („verbal load“). Zu empfehlen ist z.B. für die Intelligenzmessung die Nutzung sprachfreier Verfahren (Matrizentest, Zahlenreihen ...). Eher spezifische als allgemeine kognitive Fähigkeiten zu erheben, hat sich als eine Möglichkeit erwiesen, den Adverse Impact bei gesicherter Validität zu reduzieren (Wee, Newman & Joseph, 2014).

- In allen Materialien und im gesamten Auswahlverfahren sollte eine einfache und klare Sprache gewählt werden und es sollten genügend lange Vorbereitungszeiten auf Aufgaben gegeben werden, um sprachlich bedingte Verständnisprobleme zu entschärfen.
- Sprachcoachs sollten eingesetzt werden, um unterstützend bei sprachlichen Problemen zur Seite zu stehen und Verständnisprobleme klären zu können.
- Als Maßnahme gegen einen Stereotype-Threat sollten die Durchführungsfunktionen (Interviewende, AC-Beobachterinnen, Sprachcoachs ...) kulturell divers besetzt werden, um den Bewerbenden zu signalisieren, dass kulturelle Vielfalt willkommen ist (Eschenbach et al., 2014).

3.3 Beobachtung und Bewertung im Auswahlverfahren

Der Beobachtungs- und Beurteilungsprozess ist fehleranfällig für verschiedene Urteilsverzerrungen wie z.B. den Halo-Effekt oder implizite Persönlichkeitstheorien; auch Stereotype und Vorurteile können eine Rolle spielen. Um solche Verzerrungen zu reduzieren, sind grundsätzlich zwei Wege aussichtsreich:

(1) *Standardisierung der Abläufe und Instrumente* (Kersting & Ott, 2016): Beobachtungen und Beurteilungen sollten standardisiert ablaufen. Es sollte eine Trennung zwischen Beobachtung und Beurteilung geben, wobei erstere mittels Beobachtungsbögen genau zu dokumentieren sind. Die Beurteilung findet ausschließlich nach der Beobachtung statt, wobei verhaltensverankerte Skalen eingesetzt werden, die aus Anforderungsmerkmalen abzuleiten sind.

(2) *Schulung der Durchführenden*: Die Durchführenden werden mit den Spezifika des Auswahlverfahrens vertraut gemacht und lernen einen methodischen kompetenten Umgang mit der Beobachtungs- und Beurteilungsaufgabe (Höft & Melchers, 2010).

Unter Diversitätsbedingungen ist dabei nicht nur auf methodische Fallstricke, sondern auch auf kulturelle Fehlattraktionen sowie auf Verzerrungen

durch kulturbezogene Stereotype und Vorurteile zu achten (zum Folgenden vgl. Leenen & Stumpf, 2021). Ein Bewerber, der zum Auswahlgespräch im weißen Jackett erscheint, wird möglicherweise klischeehafte Assoziationen auslösen, wenn die Beobachter*innen den kulturellen Kontext nicht berücksichtigen können, in dem das angemessen erscheint. Beobachter*innen stehen grundsätzlich in Gefahr, Verhalten unreflektiert aus dem Blickwinkel eigener Kulturerfahrungen zu interpretieren: Ein lascher Händedruck verweist dann auf fehlendes Selbstbewusstsein, das Vermeiden von Blickkontakt spricht für Unsicherheit, eine starke Körperspannung kann als ‚machohaftes Gehabe‘ interpretiert werden. In Trainings zur kultursensiblen Personalauswahl wird geübt, „isomorph“ zu attribuieren, d.h. Verhalten so zu interpretieren, wie es Mitglieder der Kultur tun würden, in der der Akteur sozialisiert wurde. Beobachter*innen sollten also einordnen können, dass das Vermeiden expliziter Kritik in bestimmten Kulturen bei Konflikten der Beziehungsstabilisierung dient. Oder dass der im Auswahlinterview geäußerte Wunsch nach einem Beruf, „auf den die Familie stolz sein kann“, nicht einfach als Ausdruck mangelnden Selbstbewusstseins und kindlicher Naivität des Kandidaten zu interpretieren ist. Mehr kulturelles Hintergrundwissen hilft, eine unreflektierte negative Bewertung des Kandidatenverhaltens zu vermeiden.

Im Einzelnen ergeben sich für diesen Prozessschritt folgende Handlungsempfehlungen:

- Um die Wirkung von Stereotypen und Vorurteilen zu reduzieren, sollten systematische und standardisierte Beobachtungs- und Beurteilungsverfahren mit verhaltensverankerten Beurteilungsskalen eingesetzt werden.
- Eine kulturelle Sensibilisierung aller Verantwortungsträger sollte stattfinden, um die Wirkung von kulturellen Fehlinterpretationen zu reduzieren. Dies schließt auch die Entwickler des Verfahrens ein, die die Beurteilungskriterien formulieren und z.B. verhaltensverankerte Beurteilungsskalen konstruieren.
- Die Durchführenden des Verfahrens sind im Hinblick auf das Verfahren zu schulen, damit sie das Verfahren gründlich kennenlernen, das notwendige Anforderungsverständnis aufbauen und kompetent beobachten und beurteilen können.
- Weiterhin sollte man die Durchführungsfunktionen (Interviewerinnen, AC-Beobachter ...) kulturell heterogen zusammensetzen, nicht nur um einem Stereotype-Threat entgegenzuwirken, sondern auch um in den Auswahlgremien mehrere kulturelle Blickwinkel zur Geltung zu bringen.

3.4 Entscheidungsfindung und Rückmeldung

Auch das Gesamturteil im Rahmen einer auf Konsensbildung ausgerichteten Bobachterkonferenz kann durch kulturelle Fehlattritionen und Stereotype beeinflusst werden; auch hier ist Standardisierung das wirksame Gegenmittel. Am stärksten ist die Verzerrungsgefahr, wenn die Diskussion bei der Entscheidungsfindung nicht kompetent moderiert wird.

Für den Prozessschritt der Entscheidungsfindung und Rückmeldung werden folgende Handlungsempfehlungen gegeben:

- Moderator*innen von Beobachterkonferenzen sollte moderationsstechnisch sehr gut ausgebildet und zudem auch interkulturell geschult sein. Sie müssen in kulturkritischen Diskussionssituationen intervenieren und immer wieder konkrete und auswahlrelevante Verhaltensbeobachtungen und Befunde als Entscheidungsbasis einfordern.
- Bei der Festlegung der Gewichte, mit der die einzelnen Auswahlwahlergebnisse in das Gesamturteil eingehen, z.B. Ergebnisse eines kognitiven Fähigkeitstests einerseits, der stärkere Subgruppenunterschiede produzieren kann, oder eines minderheitenfreundlichen Kriteriums wie „Mehrsprachigkeit“ andererseits, muss mit Blick auf Diversitätsziele entschieden werden ohne Validitätsaspekte auszuklammern. Zur Bestimmung des besten Trade-offs wurden unterschiedliche statistische Analysemethoden entwickelt und in ihrer Anwendung veranschaulicht (z.B. Rupp, Song & Strah, 2020).
- Um die Bedeutung kultureller Vielfalt für die Organisation deutlich zu machen, sollte auch die Personengruppe, die die Ergebnisse an die Kandidatinnen und Kandidaten zurückmeldet, kulturell heterogen zusammengesetzt sein.

3.5 Evaluation und Qualitätssicherung

Bei der Evaluation eines Auswahlverfahrens unter Diversitätsgesichtspunkten ist zunächst nach dem Vorliegen eines Adverse Impact zu fragen: Schneiden Personen mit Zuwanderungsgeschichte im Gesamtverfahren schlechter ab? Wie groß sind die Unterschiede zu Angehörigen der Majorität? Und in welchen Verfahrensteilen sind die Unterschiede besonders hoch, wo sind sie niedrig? In komplexen Auswahlverfahren können sich entsprechende Analysen auf Verfahrensteile bzw. Einzelverfahren beziehen; man kann im Rahmen eines Differential-Item-Functioning-Ansatzes sogar untersuchen, ob Angehörige unterschiedlicher kultureller Gruppen bei bestimmten Items (z.B. bei einem Intelligenztest) unterschiedliche Erfolgswahrscheinlichkeiten aufweisen (Schmitt & Quinn, 2010).

Solche Analysen setzen allerdings voraus, dass sozio-demographische Daten der Bewerbenden differenziert erhoben werden. Um darüber hinaus Aussagen über die Fairness zu gewinnen, benötigt man nach dem Cleary-Modell weitere Daten über die Bewährung in der Position, für die ausgewählt wurde. Man muss dazu Erfolgskriterien festlegen und diese messen (z.B. Vorgesetzten- oder Kollegenbeurteilung, Erfolg beim Abschneiden in Ausbildungsmaßnahmen ...). Dabei muss man Sorge tragen, dass die Kriterienmessung selbst wiederum nicht durch einen Bias verzerrt wird.

Im Einzelnen ergeben sich für diesen Prozessschritt folgende Handlungsempfehlungen:

- Es sollten angemessene soziodemografische Daten über die Bewerbenden erhoben werden, die Aufschluss über die kulturelle Herkunft geben, also z.B. ob ein Migrationshintergrund vorliegt und wie dieser genau beschaffen ist.
- Die Akzeptanz des Auswahlverfahrens sollte überprüft werden (z.B. mit Hilfe des Akzeptanz-Instrumentarium von Kersting, 2010). Es sollte analysiert werden, wie die Akzeptanz bei Bewerbenden mit Zuwanderungsgeschichte, auch im Vergleich zu Bewerbenden aus der Majorität, hinsichtlich der Verfahrensausgestaltung beschaffen ist.
- Es sollte differenziert analysiert werden, wie die Bewerbenden mit anderem kulturellen Hintergrund im Verfahren im Vergleich zu den Bewerbenden aus der Majoritätsgruppe abschneiden: Gibt es Unterschiede im Hinblick auf das Abschneiden im Gesamtverfahren, wie groß sind diese und wird die 80%-Regel des Adverse Impact verletzt? Gibt es Unterschiede im Hinblick auf das Abschneiden bei einzelnen Verfahrenskomponenten? Wie könnte man das Verfahren so verändern, dass die Unterschiede kleiner werden, ohne die Validität des Verfahrens zu beeinträchtigen?
- Es sollten auch Daten zu den Bewährungskriterien in der späteren beruflichen Tätigkeit erhoben und in Relation zum Abschneiden im Auswahlverfahren gesetzt werden, um begründete Rückschlüsse auf die Fairness des Verfahrens ziehen zu können. Schneiden z.B. Personen mit Zuwanderungsgeschichte im Verfahren schlechter ab, ohne dass den Unterschieden auch Differenzen in der späteren beruflichen Bewährung entsprechen, so sollten Maßnahmen abgeleitet werden, die das Verfahren fairer machen können. Die vorlaufenden Handlungsempfehlungen geben hierzu Anregungen.

4. Fazit/Schlussbemerkung

Organisationen verfolgen heute nicht nur ökonomische oder effizienzorientierte, sondern zunehmend auch auf Diversität bezogene Ziele: Die Belegschaft der Organisation soll kulturell divers zusammengesetzt sein und Minderheiten sollen nicht benachteiligt werden. Die Motive, zunehmend auch Diversitätsziele zu verfolgen, können sehr unterschiedlich sein: Das Bemühen um eine gesellschaftliche Vorbildrolle, der Wunsch, als eine Organisation gesehen zu werden, die ihrer gesellschaftlichen Verantwortung nachkommt, oder demographische bzw. arbeitsmarktpolitische Überlegungen können die Beweggründe sein.

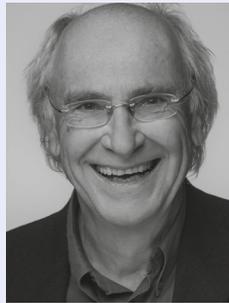
Die Personalauswahl dürfte die wichtigste Stellenschraube in einer Organisation sein, mit der Diversitätsziele verfolgt werden können. Kulturelle Diversität kann umso eher erreicht werden, je kulturfairer die Personalauswahl ausgestaltet ist. Kulturfairness bedeutet, dass es keine nicht mit den Zielpositionen begründbaren Benachteiligungen von Angehörigen unterschiedlicher kultureller Gruppierungen gibt.

Adverse-Impact-Effekte können aus einer starken Sprachgebundenheit („verbal load“) oder Kulturgebundenheit („cultural load“) des Auswahlverfahrens resultieren. Bei interaktiven Verfahren können auch Verzerrungen durch kulturelle Fehlinterpretationen, Stereotypisierungen und Vorurteile seitens der Beobachterinnen und Bewerber auftreten. Unsere Handlungsempfehlungen betonen, dass Gesichtspunkte von Kulturfairness bereits bei der Bestimmung der Anforderungsmerkmale für die zu besetzende Zielposition eine Rolle spielen. Bei der Entwicklung der Auswahlverfahren ist zu berücksichtigen, dass verschiedene Auswahlmethoden einen unterschiedlichen Adverse Impact produzieren. Unter Fairnessgesichtspunkten sind Simulationsmethoden, die nahe an der beruflichen Realität sind, abstrakteren Testverfahren (z.B. Intelligenz- oder Rechtschreibtests) vorzuziehen. Die Verfahrensentwickler, die mit Beobachtungs- und Bewertungsaufgaben Betrauten sowie die Moderator*innen der Entscheidungskonferenzen müssen in intensiven Schulungen kulturell sensibilisiert werden. Die Verfahren sollten möglichst standardisiert ablaufen, um zu verhindern, dass kulturelle Stereotype den Bewertungsprozess überlagern. Eine systematische Evaluierung kann einen Adverse Impact in einem Verfahren aufdecken und mit Blick auf Kulturfairness Verbesserungsmöglichkeiten aufzeigen.

Literaturverzeichnis

- Arbeitskreis Assessment Center / Forum Assessment e.V. (2016). *Standards der Assessment Center Methode*. Verfügbar am 10.10.2020 unter <https://www.forum-assessment.de/publikationen/standards>
- Baumann, A.-L., Feneberg, V., Kronenbitter, L., Naqshband, S., Nowicka, M. & Will, A.-K. (2019). *Ein Zeitfenster für Vielfalt. Chancen für die interkulturelle Öffnung der Verwaltung*. Herausgegeben von der Friedrich-Ebert-Stiftung. Bonn: Friedrich-Ebert-Stiftung.
- Bobko, P., & Roth, P. L. (2013). Reviewing, categorizing, and analyzing the literature on Black-White mean differences for predictors of job performance: Verifying some perceptions and updating/correcting others. *Personnel Psychology, 66*, 91–126.
- Bortz, J. & Döring, N. (1995). *Forschungsmethoden und Evaluation* (2. Auflage). Berlin: Springer.
- Campion, M. C., Campion, E. D. & Campion, M. A. (2019). Using practice employment tests to improve recruitment and personnel selection outcomes for organizations and job seekers. *Journal of Applied Psychology, 104*, 1089–1102.
- Cleary, T. A. (1968). Test bias: prediction of grades of negro and white students in integrated colleges. *Journal of Educational Measurement, 5*, 115–124.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York: Erlbaum.
- De Meijer, L. A. L., Born, M. P., Terlouw, G., & van der Molen, H. T. (2006). Applicant and method factors related to ethnic score differences in personnel selection: a study at the Dutch police. *Human Performance, 19*, 219–251.
- De Soete, B., Lievens, F. & Druart, C. (2012). An update on the diversity-validity dilemma in personnel selection: a review. *Psychological Topics, 21*(3), 399–424.
- Eschenbach, E., Virnoche, M., Cashman, E., Lord, S. & Camacho, M. (2014). *Proven Practices that can Reduce Stereotype Threat in Engineering Education: A Literature Review. Paper for Frontiers in Education Conference - October, 2014, Madrid, Spain*. Verfügbar am 02.11.2020 unter https://www.researchgate.net/publication/271967424_Proven_Practices_that_can_Reduce_Stereotype_Threat_in_Engineering_Education_A_Literature_Review
- Ette, A., Straub, S., Weinmann, M. & Schneider, N.F. (Hrsg.) (2021). *Kulturelle Vielfalt der öffentlichen Verwaltung. Repräsentation, Wahrnehmung und Konsequenzen von Diversität*. Opladen: Budrich.
- Görlich, Y. & Schuler, H. (2014). Personalentscheidung, Nutzen und Fairness. In H. Schuler & U. P. Kanning (Hrsg.), *Lehrbuch der Personalpsychologie* (3., überarbeitete und erweiterte Aufl., S. 1137–1199). Göttingen: Hogrefe.
- Harris, M. M. & Lievens, F. (2005). Selecting employees for global assignments: Can assessment centers measure cultural intelligence? In M. A. Rahim & R. Golembiewski (Hrsg.), *Current topics in management* (Vol. 10, S. 221–239). New Brunswick, New Jersey: Transaction Publishers.
- Höft, S. & Kersting, M. (2017). Die Anforderungsanalyse als Grundlage der beruflichen Eignungsdiagnostik. In Diagnostik- und Testkuratorium (Hrsg.), *Personalauswahl kompetent gestalten. Grundlagen und Praxis der Eignungsdiagnostik nach DIN 33430* (S. 97–111). Berlin: Springer.
- Höft, S. & Melchers, K. (2010). Training von AC-Beobachtern: Worauf kommt es an? *Wirtschaftspsychologie, 12*(2), 32–40.
- Hofstede, G. (1997). *Lokales Denken, globales Handeln. Kulturen, Zusammenarbeit und Management*. München: Beck
- Kehoe, J. F. (2010) Cut scores and adverse impact. In J. L. Outtz (Ed.) *Adverse Impact. Implications for Organizational Staffing and High Stakes Selection* (pp. 289–322). New York: Routledge.
- Kersting, M. (2010). Akzeptanz von Assessment Centern: Was kommt an und worauf kommt es an? *Wirtschaftspsychologie, 12*(2), 58–65. [Die Akzept!-Fragebögen sind online zu beziehen, verfügbar am 05.08.2019 unter: <http://kersting-internet.de/testentwicklungen/akzept-fragebogen/>]
- Kersting, M. & Ott, M. (2016). Diversity-gerechte Personalauswahl. In P. Genkova & T. Ringeisen (Hrsg.), *Handbuch Diversity Kompetenz Band I: Perspektiven und Anwendungsfelder*. (S. 679–692). Wiesbaden: Springer.
- Krause, A., Rinne, U., Zimmermann, K. F., Bösch, I. & Alt, R. (2012). *Pilotprojekt „Anonymisierte Bewerbungsverfahren“: Abschlussbericht*. IZA Research Report Nr. 44. Verfügbar am 20.10.2020 unter <https://www.iza.org/publications/r/146/pilotprojekt-anonymisierte-bewerbungsverfahren-abschlussbericht>
- Leenen, W.R., Scheitza, A. & Stumpf, S. (2014). Kulturfairness in der Personalauswahl. In H. Uske, A. Scheitza, S. Düring-Hesse & S. Fischer (Hrsg.), *Interkulturelle Öffnung der Verwaltung. Konzepte, Probleme, Beispiele* (S. 103–111). Duisburg u.a.: Eigenverlag.
- Leenen, W.R. & Stumpf, S. (2021). *Arbeitshilfe kultursensible Personalauswahl. Eine Arbeitshilfe für mit Personalgewinnung und Personalauswahl Beschäftigte in Bundesbehörden*. Brühl: Bundesakademie für Öffentliche Verwaltung (BAKöV) (im Druck).
- Lindsey, A., King, E., McCausland, T., Jones, K. & Dunleavy, E. (2013). What we know and don't: Eradicating employment discrimination 50 years after the Civil Rights Act. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 6*(4), 391–413.

- Melchers, K. G. & Annen, H. (2010). Officer selection for the swiss armed forces. An evaluation of validity and fairness issues. *Swiss Journal of Psychology*, 69, 105–115.
- Morris, S. B. & Dunleavy, E. M. (Eds.) (2016). *Adverse Impact Analysis: Understanding Data, Statistics, and Risk*. New York: Routledge.
- Ottz, J. L. & Newman, D. A. (2010). A theory of adverse impact. In J. L. Ottz (Hrsg.), *Adverse impact. Implications for organizational staffing and high stakes selection* (S. 53–94). New York: Routledge.
- Ployhart, R. E., & Holtz, B. C. (2008). The diversity-validity dilemma: strategies for reducing race-ethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, 61, 153-172.
- Podsiadlowski, A. (2007). Multinationale Teams. In J. Straub, A. Weidemann & D. Weidemann (Hrsg.), *Handbuch Interkulturelle Kommunikation und Kompetenz* (S. 576-586). Stuttgart/ Weimar: J.B. Metzler.
- Roth, P. L., Bevier, C. A., Bobko, P., Switzer, F. S. III & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: a meta-analysis. *Personnel Psychology*, 54, 297–330.
- Roth, P. L., Van Iddekinge, C. H., DeOrtentiis, P. S., Hackney, K. J., Zhang, L., & Buster, M. A. (2017). Hispanic and Asian performance on selection procedures: A narrative and meta-analytic review of 12 common predictors. *Journal of Applied Psychology*, 102, 1178–1202.
- Rupp, D. E., Song, Q. C. & Strah, N. (2020). Addressing the so-called validity–diversity trade-off: Exploring the practicalities and legal defensibility of Pareto-optimization for reducing adverse impact within personnel selection. *Industrial and Organizational Psychology*, 13, 246–271.
- Russell, C. (2000). *Cleary Model of Test Bias*. Verfügbar am 17.07.2019 unter http://www.ou.edu/russell/whitepapers/Cleary_model.pdf
- Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, 85, 440–452.
- Schmitt, N. & Quinn, A. (2010). Reductions in measured subgroup mean differences: What is possible? In J. L. Ottz (Ed.), *Adverse impact. Implications for organizational staffing and high stakes selection* (pp. 425-451). New York: Routledge.
- Schuler, H. (2014). *Psychologische Personalauswahl. Eignungsdiagnostik für Personalentscheidungen und Berufsberatung* (4. Auflage). Göttingen: Hogrefe.
- Stumpf, S., Leenen, W. R. & Scheitza, A. (2017). Adverse Impact in der Personalauswahl einer deutschen Behörde: Eine Analyse ethnischer Subgruppendifferenzen. *German Journal of Human Resource Management (Zeitschrift für Personalforschung)*, 31, 4–31.
- Wee, S., Newman, D. A. & Joseph, D. L. (2014). More than g: selection quality and adverse impact implications of considering second-stratum cognitive abilities. *Journal of Applied Psychology*, 99, 547–563.
- Zedeck, S. (2010). Adverse Impact: History and Evolution. In J. L. Ottz (Ed.), *Adverse impact. Implications for organizational staffing and high stakes selection* (pp. 3-27). New York: Routledge.



Prof. (emerit.) Dr. Rainer Leenen
Kölner Institut für Interkulturelle Kompetenz (KIiK)
An Groß St. Martin 6
50667 Köln
leenen@kiik.eu



Corresponding author

Prof. Dr. Siegfried Stumpf
Institute for Business Administration and Leadership
Technische Hochschule Köln,
Campus Gummersbach
Steinmüllerallee 1
51643 Gummersbach
siegfried.stumpf@th-koeln.de



Prof. Dr. Stefan Höft
Fachgruppe Psychologie
Hochschule der Bundesagentur für Arbeit (HdBA)
Seckenheimer Landstr.16
D-68163 Mannheim
stefan.hoeft@hdba.de