

# Fairness in Personalauswahlverfahren – eine wissenschaftliche und praktische Perspektive

Julia Wagenschieber & Amelie Kleinmanns

Immer wieder stellen sich Organisationen, die Personalauswahl mit wissenschaftlich fundierten Auswahlinstrumenten wie z.B. dem kognitiven Leistungstest und / oder Assessment-Centers betreiben, die Frage, ob der Auswahlprozess für alle Bewerber\*innen gleichermaßen fair ist oder ob bestimmte Bewerber\*innen-Gruppen benachteiligt werden. Für die meisten Organisationen ist dieser Aspekt – neben der Vorhersagekraft der angewandten Auswahlinstrumente – zentral: Eine systematische Benachteiligung einzelner Bewerber\*innen kann und will sich niemand erlauben. Aufgrund aktueller Themen wie der Gleichbehandlung von Frauen und Männern im Arbeitskontext, dem Bedeutungszuwachs von Diversität in Organisationen und nicht zuletzt der Diskriminierung von „people of colour“ verdient das Thema „Fairness in Auswahlprozessen“ eine differenzierte Betrachtung. Wesentliche Fragen zur Fairness, die im Zusammenhang mit 1. Kognitiven Leistungstests und 2. Assessment-Centers auftreten können, werden aus wissenschaftlicher und praktischer Perspektive beantwortet.

Wir unterscheiden nach der einschlägigen Literatur zwischen zwei unterschiedlichen Fairness-Konzepten: Zum einen sprechen wir von einer Form der „objektiven Fairness“, die beschreibt, inwiefern ein Auswahlinstrument wie ein Leistungstest oder ein Assessment-Center (AC) für alle Bewerber\*innen mit gleichen Erfolgchancen und unter gleichen Voraussetzungen zu bearbeiten ist. Insbesondere Rahmenbedingungen für die Durchführung eines Auswahlverfahrens spielen hier eine wichtige Rolle. So spricht man vor diesem Hintergrund u.a. von der prozeduralen Fairness (Gilliland, 1993). Diese bezieht sich auf Aspekte, die die Chance auf ein erfolgreiches Abschneiden im Auswahlverfahren für alle Bewerber\*innen erhöhen oder ein erfolgreiches Abschneiden zumindest nicht systematisch behindern. Ein vergleichbares Vorgehen bei der Aufgabeninstruktion – um diesen Treiber der prozeduralen Fairness aufzugreifen – bildet sowohl während eines Leistungstests wie auch in einem Assessment-Center einen wichtigen Grundstein für die „objektive Fairness“ des Verfahrens. Man stelle sich vor, eine/r von zwei Bewerber\*innen würde eine ausführlichere Erklärung zu einer Aufgabenstellung bekommen als die/der andere. Zum Beispiel, weil sie/er eine vertiefende Frage stellt. In diesem Fall ist während der Durchführung eine Ungleichbehandlung zu verzeichnen, die sich in ihrem Ausmaß kaum bestimmen und daher in der Leistungsbewertung auch nicht berücksichtigen lässt. Die Fairness des Tests oder Assessment-Centers ist in diesem Fall durch eine mangelnde Durchführungsobjektivität eingeschränkt. Auch die DIN 33430 fordert, dass die Anwendung von Auswahlinstrumenten zur Eignungsbeurteilung nicht zu einer Benachteiligung einzelner Kandidat\*innen oder Gruppen führen darf. Insbesondere ist darauf zu achten, dass irrelevante Eigenschaften, Kenntnisse, Fertigkeiten oder Fä-

higkeiten keinen Einfluss auf das Ergebnis haben. Vor allem wenn sie zugleich bei der Zielgruppe des Verfahrens unterschiedlich ausgeprägt sein können (DIN 33430:2016-07, 2016).

Fairness in Auswahlprozessen ist ein multidimensionales Konstrukt (Arvey, Strickland, Drauden & Martin, 1990; Gilliland, 1993; Kersting, 2008): Neben den Tatsachen der Chancengleichheit und gleichen Rahmenbedingungen für alle Bewerber\*innen definiert sich Fairness auch in der subjektiven Wahrnehmung der Bewerber\*innen. Diese entscheidet, ob a) erfolgreiche Bewerber\*innen ein anschließendes Jobangebot annehmen (Cable & Judge, 1996) und b) wie nicht erfolgreiche Bewerber\*innen reagieren, ob sie beispielsweise einen Arbeitgeber weiterempfehlen würden, sich wiederholt bewerben würden und – vor allem in Zeiten sozialer Medien und eines ungehinderten Nutzens von Arbeitgeberbewertungsplattformen – wie sie ihr Bild des Arbeitgebers nach außen tragen (Gilliland, Groth, Baker, Dew, Polly & Langon, 2001; Hausknecht, Day & Thomas, 2004). Die wahrgenommene Fairness eines Auswahlinstruments bzw. -prozesses wird neben den Prozessvariablen auch von persönlichen Dispositionen der Individuen beeinflusst, die die Auswahlverfahren absolvieren. So hat die Wahrnehmung, dass ein Test oder ein Assessment-Center fair ist, damit zu tun, inwiefern ein/e Bewerber\*in glaubt, ihre/ seine Fähigkeiten in solchen Verfahren grundsätzlich unter Beweis stellen zu können, d.h. wie hoch die eigene Selbstwirksamkeitserwartung

**ist „die subjektive Gewissheit, neue oder schwierige Anforderungssituationen aufgrund eigener Kompetenz bewältigen zu können.“**

Schwarzer & Jerusalem, 1994, S.35

Solche Aufgaben, die i.d.R. durch Erfahrungen (z.B. mit Prüfungen, Klausuren etc.) gekennzeichnet sind, formen diese Selbstwirksamkeitserwartung (Sitzmann & Yeo, 2013). Ist sich jemand seiner Fähigkeiten sicher und hat vielleicht auch den einen oder anderen Leistungstest schon erfolgreich absolviert, wird er einen Auswahltest tendenziell als fairer bewerten. Jemand, der glaubt, ohnehin in solchen Testverfahren meistens schlecht abzuschneiden, tendiert dazu, einen solchen Bewerbungstest als weniger faire Möglichkeit zu sehen, sich zu beweisen.

## 1. Zur Fairness in Leistungstests

### Frage 1: Wann ist ein Leistungstest „objektiv fair“?

Die Fairness ist ein wesentlicher Teilaspekt der Gültigkeit (Validität) eines Leistungstests und der diagnostischen Entscheidung, die aufgrund seiner Ergebnisse getätigt wird (Standards for Educational and Psychological Testing, 2014). Die Definition der Fairness von Leistungstests ähnelt sehr stark den Anforderungen, die die DIN 33430 an alle Auswahlinstrumente stellt. Ein Leistungstest gilt als objektiv fair, wenn bestimmte Personen aufgrund ihrer Testergebnisse nicht systematisch benachteiligt werden (Eid & Schmidt, 2014). Diese Benachteiligungen können durch die Zugehörigkeit einer Person zu einer bestimmten ethnischen, soziokulturellen oder geschlechtsspezifischen Gruppe entstehen. Soll z.B. eine Aussage über die Fähigkeit von Bewerber\*innen getroffen werden, Textrechenaufgaben zu lösen, sollten die Aufgaben nur diese Fähigkeit überprüfen und nicht zusätzliches anderes irrelevantes Fachwissen zur Aufgabenbewältigung nötig sein, über das verschiedene Gruppen in einem unterschiedlichen Ausmaß verfügen. Es wäre zu vermuten, dass Männer bei ei-

ner Textrechenaufgabe, deren Hintergrund stark physiklastig wäre und deren Lösung eine Formel benötigt, die nicht im Text angegeben ist, im Durchschnitt bessere Leistungen erzielen. In Deutschland gibt es immer noch einen starken Interessensunterschied zwischen Frauen und Männern in Bezug auf Naturwissenschaften, wie Physik. Daher wäre vielen Männern die Formel vielleicht geläufig und das Fachvokabular eher präsent als Frauen. Die bessere Testleistung wäre in diesem Fall aber nicht auf die bessere Textrechenleistung zurückzuführen, sondern auf das fachspezifische Wissen.

Damit jede\*r Bewerber\*in eine faire Chance bei der Teilnahme an einem Leistungstest eingeräumt wird, sind drei Gestaltungsaspekte auf ihre Fairness zu überprüfen und bei der Erstellung des Auswahlverfahrens und dem Einsatz eines Leistungstests besonders zu beachten:

- Die Bedingungen, unter denen ein Test durchgeführt wird
- Die Entscheidungen, die aufgrund der Testergebnisse getroffen werden
- Die psychometrischen Eigenschaften des Tests selber

### Frage 2: Wann sind die Bedingungen, unter denen ein Test durchgeführt wird, fair?

Bei der fairen Gestaltung der Bedingungen zur Durchführung eines Leistungstests spielt die Durchführungsobjektivität eine zentrale Rolle. Um die bestmögliche Durchführungsobjektivität zu gewährleisten, sollte die Durchführung des Leistungstests standardisiert sein. D.h. vorab sollten alle Regeln, die die Durchführung betreffen, festgelegt werden. Das reicht von organisatorischen Fragen (z. B. Wann werden

die Bewerber\*innen in welchem Raum getestet?) über Verhaltensregeln für den/die Testanweiser\*in (wie wird z.B. auf Nachfragen reagiert) bis hin zu Standardtexten zur Erklärung der Aufgabengruppen und den gleichen Zeitvorgaben für die Bearbeitung der Aufgaben. Wird von den standardisierten Vorgehensweisen abgewichen, führt dies zu schwer einschätzbaren Vor- bzw. Nachteilen gegenüber bestimmten Bewerber\*innen. Zulässig ist ein Abweichen von der Standardisierung für Personen, die z.B. aufgrund einer Behinderung einen Nachteil dadurch erfahren würden, wie das Testmaterial präsentiert wird. Solche Benachteiligungen, die nicht eignungsrelevant sind, gilt es zu verhindern. Ziel der Standardisierung ist die Minimierung des Einflusses der Testsituation und der/des Testanweiser\*in.

### Frage 3: Wann sind die Entscheidungen, die aufgrund der Testergebnisse getroffen werden, fair?

Ob die aufgrund der Testergebnisse getroffenen Entscheidungen fair sind, hängt entscheidend davon ab, ob nur relevante Eignungskriterien erfasst werden und in die Entscheidung einfließen. Dieser Anforderungsbezug kann durch eine sorgfältige Erstellung eines Anforderungsprofils gewährleistet werden (Ristel & Jobmann, 2017). Wird hier bereits bei dem Anforderungsprofil eine Fähigkeit aufgenommen, die für den späteren Berufserfolg irrelevant ist und die in Untergruppen der Bewerber\*innen unterschiedlich stark ausgeprägt ist, führt dies unweigerlich zur Benachteiligung innerhalb der Untergruppen, obwohl der Leistungstest fair durchgeführt wurde. Daher sollte immer überprüft werden, ob die im Leistungstest abgefragten Fähigkeiten wirklich relevant sind. Metaanalytisch lässt sich zeigen, dass kognitive Leistungen einen substantiellen Zusammenhang mit Berufserfolg in den verschiedensten Berufen

aufweisen (Kramer, 2009; Schmidt & Hunter, 1998). Des Weiteren muss sichergestellt sein, dass eine standardisierte Auswertung des Leistungstests durchgeführt wird (Auswertungsobjektivität). Abschließend muss eine eindeutige Interpretationsregelung der Ergebnisse vor der Durchführung des Auswahlverfahren festgelegt sein (Standards for Educational and Psychological Testing, 2014).

### Frage 4: Wann sind die psychometrischen Eigenschaften eines Leistungstests fair?

Damit die psychometrischen Eigenschaften eines Tests als fair eingestuft werden können, wird die Gültigkeit von statistisch-mathematischen Modellen überprüft, durch die getestet wird, ob für verschiedene Gruppen die gleiche Eigenschaft gemessen wird (Jobmann, 2017). Wenn für alle Bewerbergruppen die gleiche Eigenschaft gemessen wird, dann spricht man von Messinvarianz. Wenn diese vorliegt, ist ein Leistungstest in dem Sinne fair, dass für alle Gruppen das gleiche Konstrukt (Fähigkeit) gemessen wird. In diesem Fall können Unterschiede in der Testleistung als Unterschiede in der Fähigkeit interpretiert werden (Jobmann, 2017). Diese statistisch-mathematischen Modelle werden sowohl auf der Ebene der einzelnen Fragen durchgeführt als auch auf der Ebene des gesamten Tests.

### Frage 5: Gibt es in objektiv fairen Leistungstests in Auswahlverfahren keine Gruppenunterschiede zwischen Subgruppen?

Oft wird im allgemeinen Verständnis von Fairness davon ausgegangen, dass es keine Gruppenunterschiede geben darf, damit ein Leistungstest fair ist. Allerdings erlauben Fairnesskonzepte durchaus Unterschiede in unterschiedlichen Bewerbergruppen,

wenn die Unterschiede auf tatsächliche Unterschiede in der Fähigkeit zurückzuführen sind und diese Fähigkeit relevant für den Berufserfolg ist (Jobmann, 2017; Bühner, Ziegler & Kersting, 2018).

Die internationale Wissenschaft beschäftigt sich schon seit mehreren Jahrzehnten mit Gruppenunterschieden bei Kriterien, die für die Personalauswahl relevant sind. Man spricht hier im Englischen von Adverse Impact (negative Auswirkungen). Dabei geht es um die Frage, ob Personengruppen (Geschlechtergruppen, Altersgruppen usw.) bevor- oder benachteiligt werden, weil sie im Durchschnitt in diagnostischen Kriterien besser oder schlechter abschneiden. Hough, Oswald, & Ployhart (2001) geben eine Übersicht über Gruppenunterschiede in Leistungstests. Ein Auszug von Mittelwertsunterschieden soll hier kurz vorgestellt werden:

- Es werden keine Unterschiede der Werte für allgemeine Intelligenz zwischen Männern und Frauen festgestellt. Es wird darauf hingewiesen, dass solche Tests aber auch mit dem Ziel, keine systematischen Gruppenunterschiede zu produzieren, konzipiert werden. Auffällig ist, dass Männer eine höhere Standardabweichung aufweisen als Frauen. Das bedeutet, dass es sowohl mehr Männer im ganz unteren Bereich als auch im ganz oberen Bereich der allgemeinen Intelligenzverteilung gibt. Jüngere Personen erzielten im Mittel höhere Werte als ältere Personen ( $d = .4$ ).
- Bei spezifischeren Fähigkeitsmessungen treten allerdings meist kleinere Mittelwertunterschiede zwischen den Geschlechtern auf. So schneiden Frauen besser bei verbalen Fähigkeiten ( $d = .1$ ), Männer besser bei numerischen Fähigkeiten ( $d = .2$ ) und räumlichen Fähigkeiten ( $d = .4$ ) ab.

→ Zusätzlich zu den Ergebnissen von Hough, Oswald, & Ployhart (2001) zeigt sich, dass Männer deutlich bessere Leistungen in allgemeinen Wissenstests zeigen (Engelberg, 2015). Es zeigt sich auch, dass ältere Personen über mehr Wissen verfügen als jüngere (Nyberg, et al., 2003).

Wie diese Unterschiede entstehen, ist wissenschaftlich nicht abschließend zu beantworten. Sie bedeuteten allerdings nicht, dass Leistungstests objektiv unfair sind. Man darf nicht außer Acht lassen, dass die Bewerber\*innengruppe, die einen Personalauswahlprozess durchläuft, selbstselektiert ist und nicht repräsentativ für die Bevölkerung steht. Es kann daher zu einer Ungleichverteilung der Fähigkeiten zwischen den Untergruppen in einer Bewerber\*innengruppe – bezogen auf die Auswahlkriterien – kommen.

#### Frage 6: Wovon hängt ab, ob Bewerber\*innen einen Leistungstest als fair einstufen?

Generell sollte auch bei der Durchführung eines Leistungstests natürlich darauf geachtet werden, dass man mit den Bewerber\*innen wertschätzend, freundlich und respektvoll umgeht.

Darüber hinaus ist der wichtigste Aspekt für eine positive Fairnessbewertung, dass der Bezug zwischen dem eingesetzten Leistungstest und der Stelle, auf die sich der/die Bewerber\*in bewirbt, erkennbar wird. Dies ist dadurch zu erreichen, dass die Auswahl des Leistungstests mit Hinblick auf die Anforderungsanalyse erfolgt und somit Fähigkeitsdimensionen enthält, die für den Berufserfolg der zu besetzenden Position relevant sind. Sollte dieser Bezug nicht direkt aus dem Aufgabenmaterial oder der Benennung hervorgehen, dann ist es hilfreich diesen den Bewerber\*innen zu erläutern (Kersting, 2018).

Generell wirken sich Erklärungen positiv auf die Fairnesswahrnehmung von Bewerber\*innen aus (Truxillo, Bodner, Bertolino, Bauer, & Yonce, 2009). Daher sollte auch der Einsatz von Leistungstests so transparent wie möglich gestaltet werden. Es bietet sich daher an, Informationen zu Ablauf, grobem Inhalt und Bewertung bereits vor der Testung zur Verfügung zu stellen. Je mehr Informationen zur Verfügung stehen, desto besser können sich Bewerber\*innen auf den Leistungstest vorbereiten. Es steigt die Selbstwirksamkeit bei den Bewerber\*innen und sie nehmen den Leistungstest eher als eine Möglichkeit zur Präsentation ihrer Fähigkeiten wahr. Dies ist, wie anfangs im Artikel bereits erwähnt, neben der Gestaltung des Auswahlprozesses von großer Bedeutung für die Fairnesswahrnehmung der Bewerber\*innen.

Leistungstest werden von Bewerber\*innen in der Regel eher positiv wahrgenommen, wenn Sie die oben genannten Kriterien erfüllen. Eine Metaanalyse zeigt: Sie schneiden besonders gut in den Bereichen wahrgenommene Wissenschaftlichkeit und Wahrung der Privatsphäre ab. In den Bereichen Augenscheinvalidität und vermutete Aussagekraft liegen Leistungstests nur knapp hinter ACs und Interviews, allerdings ist die Beurteilung der „zwischenmenschlichen Wärme“ aufgrund der Standardisierung eher gering (Anderson, Salgado, & Hülsheger, 2010). Es zeigt sich außerdem, dass es einen positiven Zusammenhang zwischen der tatsächlichen Aussagekraft eines Personalauswahlinstrumentes gibt und der positiven Wahrnehmung durch die Bewerber\*innen. D.h. objektiv fair gestaltete Auswahlinstrumente werden auch fairer wahrgenommen. Die Ziele, einen validen Leistungstest zu diagnostischen Zwecken einzusetzen und eine positive Wahrnehmung der Bewerber\*innen gegenüber dem Auswahlverfahren zu erreichen, stehen somit nicht im Gegensatz.

## 2. Zur Fairness in Assessment-Centers

### Frage 1: Sind Assessment-Center fairer als Leistungstests?

Ein Assessment-Center versteht sich als ein eignungsdiagnostisches Instrument, das durch ein Zusammenspiel unterschiedlicher Aufgaben, die unterschiedliche Anforderungen messen können, eine umfassende Prognose der späteren Leistungen auf der zu besetzenden Position abgeben kann (z.B. Lievens & Klimoski, 2001). Ebenso wie bei einem kognitiven Leistungstest basiert diese Prognose auf Wahrscheinlichkeiten; die Arbeitsleistung einer/eines einzelnen Bewerber\*in mit Sicherheit vorherzusagen, das vermag derzeit kein Auswahlinstrument.

Ein kognitiver Test konzentriert sich darauf, die kognitiven Fähigkeiten der Bewerber\*innen vorherzusagen, während das Assessment-Center eher dafür bekannt ist, die sozialen und kommunikativen Anforderungsbereiche sichtbar zu machen (wobei auch kognitive Anforderungen hier involviert sein können). Kernmerkmal des Assessment-Centers ist daher die Multidimensionalität – mit einem Messwert (Gesamtwert) erfasst man die Eignung auf der Basis mehrere Anforderungskriterien (Lievens & Klimoski, 2001). Die Frage, ob Assessment-Center fairer sind als kognitive Leistungstests, könnte man daher eventuell voreilig mit „Ja“ beantworten, da die Eignungsprognose auf mehreren Kriterien, die für den beruflichen Erfolg relevant sind, getroffen wird, und das Urteil daher umfassender zu sein scheint als das eines kognitiven Leistungstests.

Allerdings lässt die Multidimensionalität eines Assessment-Centers das Ergebnis sehr komplex werden. Die Gesamtbewer-

tung für jede/jeden Bewerber\*in, die am Ende des Assessment-Centers steht und in eine Rangreihenfolge überführt wird, suggeriert vollständige Vergleichbarkeit. Tatsächlich ist diese – im Vergleich zu einem kognitiven Leistungstest, der nach festen Regeln durchgeführt und standardisiert ausgewertet wird – eingeschränkt. Im Assessment-Center werden Leistungen und Verhalten der Bewerber\*innen meistens durch mehrere Vertreter\*innen der Organisation bewertet, durch Menschen, die einen eigenen Erfahrungshintergrund mitbringen, was unterbewusst dazu führen kann, dass die Beurteilungen der Bewerber\*innen subjektiv verzerrt werden (Koch, D’Mello & Sackett, 2015). Da die Durchführung und Auswertung eines Assessment-Centers nie so standardisiert und „objektiv“ erfolgen kann wie die eines kognitiven Leistungstests, ist die Gefahr für Fairness-Einschränkungen hier größer, auch wenn es auf den ersten Blick so scheint, dass Bewerber\*innen in einem Assessment-Center weitaus größere Chancen haben, ihre Eignung unter Beweis zu stellen, als in einem kognitiven Leistungstest. Assessment-Center erfreuen sich oft einer höheren Augenscheinvalidität, d. h. sie suggerieren durch ihre Praxisnähe, eher die relevanten Anforderungskriterien zu erfassen, als abstrakte kognitive Leistungstests dies zu tun scheinen. Tatsächlich ist der Erfolg eines Assessment-Centers an eine Reihe von Bedingungen geknüpft, die unter Frage 2 näher erläutert werden.

Auch die Wissenschaft hat sich bereits umfassend mit der Frage beschäftigt, ob Assessment-Center das fairere Auswahlinstrument sind, wenn man sie mit einem kognitiven Test vergleicht. Dazu lassen sich immer wieder Studien finden, die zeigen, dass Assessment-Center Subgruppen weniger diskriminieren als kognitive Leistungstests. Stumpf et al. (2016) zeigten, dass Bewerber\*innen mit Migrationshintergrund

im Assessment-Center besser abschnitten als im kognitiven Leistungstest. Die Quote derjenigen, die das Assessment-Center erfolgreich bestanden, fiel höher aus als beim Leistungstest. Stumpf et al. (2016) fügen ihrem Befund jedoch die folgende einschränkende Erklärung hinzu, die zur Vorsicht bei der Interpretation der Ergebnisse aufruft: Häufig ist die Einstellungsquote der Subgruppen auch durch die starke Selektion in vorherigen Auswahlritten (z.B. kognitiven Tests) begrenzt; diese selektieren auch in den Subgruppen diejenigen aus, die die Anforderungen an die kognitiven Fähigkeiten nicht erfüllen. Folglich sind im Assessment-Center, das i.d.R. erst spät im Auswahlprozess eingesetzt wird, weil es sehr aufwendig ist, weniger Personen der Subgruppe vertreten. Somit ist es möglich, dass der Eindruck, das Assessment-Center diskriminiere weniger, verzerrt ist, da kognitive Leistungstests und Assessment-Center zu unterschiedlichen Zeitpunkten im Auswahlprozess eingesetzt werden. Das Fairnessmaß von kognitiven Tests und AC kann somit nicht unmittelbar miteinander verglichen werden, da häufig Grundgesamtheiten mit unterschiedlichen Eigenschaften zugrunde liegen. Im Assessment-Center hat man jedoch eher die Möglichkeit, mehr und verschiedenartige Auswahl-Kriterien heranzuziehen, die eine Diskriminierung unwahrscheinlicher machen. Dennoch lässt sich auf Basis der aktuellen wissenschaftlichen Erkenntnisse nicht pauschal sagen, das eine Instrument sei fairer als das andere. Die Fairness hängt jedoch maßgeblich an der Vorhersagekraft des Auswahlinstruments, und zwar an dem tatsächlich quantifizierbaren Zusammenhang zwischen Erfolg im Auswahlverfahren und im späteren Beruf, nicht an der vermeintlich sichtbaren Augenscheinvalidität.

## Frage 2: Was sind mögliche Ursachen für unfaire Assessment-Center?

Damit man ein faires Assessment-Center durchführt, sollte man sich die wesentlichen Stellschrauben vergegenwärtigen, die für Fairness in Auswahlverfahren verantwortlich sind. Ähnlich wie für die Fairness-Definition in kognitiven Leistungstests ist auch im AC zu argumentieren: Fairness wird im Wesentlichen dadurch bestimmt, dass anforderungsbezogene Eigenschaften und Kriterien von Bewerber\*innen mit aussagekräftigen Auswahlinstrumenten erfasst werden. Nur dann ist eine zuverlässige und valide Eignungsaussage auf einer annehmbaren Wahrscheinlichkeitsbasis möglich. Ein Assessment-Center zählt – wenn es sauber konstruiert und vorbereitet wurde – zu den vorhersagestärksten Auswahlinstrumenten. Metaanalytische Ergebnisse (Schmidt & Hunter, 1998; Salgado, Anderson, Moscoso, Bertua, & De Fruyt, 2003; Kramer, 2009) konnten dies immer wieder bestätigen. Dabei sind psychometrische Gütekriterien, wie sie für kognitive Leistungstest erhoben werden und eindeutige Indikatoren für die Qualität des Tests darstellen, schwieriger zu erfassen. Zum Teil gelten dabei auch andere Maßstäbe: Für die Qualität eines Assessment-Centers ist es zum Beispiel maßgeblich, dass Eigenschaften, die als Anforderungen definiert sind, zu einem gewissen Grad über verschiedene Aufgaben des Assessment-Centers in konstanter Ausprägung zu beobachten sind. Bewerber\*innen, die in einem Führungsszenario empathisch reagieren, treten mit hoher Wahrscheinlichkeit auch empathisch in einem Gespräch mit einem Gesprächspartner auf gleicher Ebene auf. Dies sollte durch die Stationen des Assessment-Centers auch abgebildet werden, wenn „Empathie“ ein Anforderungskriterium ist. Man geht also davon aus, mit dem Assessment-Center situationsunabhängige stabi-

le Verhaltenskriterien messen zu können. Diese sind im Anforderungsprofil für eine zu besetzende Position enthalten, weil sie für mehrere Tätigkeiten und Situationen der Zielposition relevant sind. Auch wenn man die Anforderungsbereiche im Assessment-Center situationsgebunden (z.B. in einem simulierten Mitarbeitergespräch) misst, macht man situationsunabhängige Eigenschaftskriterien sichtbar.

Gleichzeitig wird von einem Assessment-Center erwartet, dass es eine gewisse Anforderungsbreite abdeckt, d.h. eben nicht ausschließlich kognitive, sondern auch eine Bandbreite unterschiedlicher sozialer Kompetenzen erfassen kann. Dies bedeutet auch, dass unterschiedliche Aspekte der sozialen Kompetenz, beispielsweise „Empathie“ und „Konfliktfähigkeit“, annäherungsweise trennscharf zu messen sind. So sollten einzelne Aufgabenstellungen im Assessment-Center die zu messende Anforderung ansprechen, so dass Bewerber\*innen die Chance bekommen, genau die Fähigkeit, die von Interesse ist, zeigen zu können, diese dabei jedoch auch trennscharf und beständig erfassen. Es liegt auf der Hand, dass der Vorbereitung und sauberen Konzeption des Assessment-Centers eine hohe Bedeutung zukommt, wenn es um die Sicherstellung von Fairness geht.

Ob ein Assessment-Center fair ist, hängt davon ab, ob die Anforderungen, die mit dem Assessment-Center erfasst werden, maßgeblich für den beruflichen Erfolg auf der zu besetzenden Position sind (vgl. auch DIN33430). Werden Anforderungskriterien definiert, die nichts damit zu tun haben, ob jemand seine Tätigkeiten erfolgreich oder nicht erfolgreich erledigt, kann das Assessment-Center nicht fair sein. Beispielsweise wird manchmal Kreativität ins Anforderungsprofil für Sachbearbeiter\*innen in der öffentlichen Verwaltung aufgenommen.



Kreativität bezieht sich oftmals darauf, dass jemand originelle, neuartige Ideen generieren kann, die einen tatsächlichen Mehrwert bieten. Nur wenn beides erfüllt ist, spricht man tatsächlich von Kreativität (Runco & Jaeger, 2012). In den meisten Fällen jedoch stellt Kreativität im dargestellten Sinne keine Voraussetzung für den Erfolg von Sachbearbeiter\*innen im Verwaltungsbereich dar. Viel eher ist gemeint, dass jemand Probleme lösen kann und dafür auch einmal auf Möglichkeiten zurückgreift, die vom Standard abweichen. Dass solche Lösungsstrategien immer komplett neuartig sein müssen, ist selten Voraussetzung. Kreativität im Auswahlverfahren für Sachbearbeitungspositionen als Anforderungskriterien heran zu ziehen, könnte daher in vielen Fällen in einem weniger fairen Assessment-Center münden.

Der Assessment-Center-Vorbereitung, der Erstellung des Anforderungsprofils sowie der Konstruktion der Stationen gilt daher auch in Bezug auf die Anforderungsanalyse für die zu besetzende Position besondere Aufmerksamkeit, wenn man ein faires AC erstellen will. Fairness ist im Wesentlichen von der Validität des Assessment-Centers abhängig, d.h. das AC ist so konstruiert, dass es die erfolgskritischen Anforderungen misst. Alle so entstehenden Unterschiede in der Leistung, die eventuell dafür sorgen, dass bestimmte Bewerber\*innengruppen seltener erfolgreich abschneiden als andere, sind auf dieser Argumentationsbasis für eine erfolgreiche Stellenbesetzung relevante Unterschiede, wenn man dem Auswahlinstrument vertraut, was bei einem sauber konstruierten AC gerechtfertigt wäre.

### Frage 3: Wie kann man im Assessment-Center Fairness herstellen?

Wie aus den Ausführungen zu Frage 2 hervorgeht, erfolgt die unfaire Behandlung be-

stimmter Subgruppen im Assessment-Center häufig durch die Definition der Anforderungen, nicht durch das Messinstrument „Assessment-Center“ an sich. Werden Kriterien für die zu besetzende Position ins Anforderungsprofil aufgenommen, die eine bestimmte Gruppe an Bewerber\*innen per sé seltener mitbringt, wird im Vorfeld bereits festgelegt, dass Bewerber\*innen der entsprechenden Gruppe es schwerer haben. Definiert man beispielsweise eine autoritäre Wirkung oder ein dominantes Auftreten als Voraussetzung im Anforderungsprofil für eine Führungskraft, kann es sein, dass weibliche Bewerber\*innen dadurch bereits benachteiligt werden, da sie oft allein aufgrund ihrer durchschnittlichen Körpergröße weniger autoritär und dominant wirken als männliche Bewerber\*innen, da Körpergröße häufig (unterbewusst) mit einem dominanten Auftreten in Verbindung gebracht wird (Blaker, Rompa, Dessing, Vried, Herschberg & Van Vugt, 2013). Hier kann es helfen, sich zu vergegenwärtigen, warum eine autoritäre Wirkung und ein dominantes Auftreten für die zu besetzende Position wichtig sind. Geht es möglicherweise um Durchsetzungsstärke? So gelingt es eher, Kriterien zu definieren, die bestimmte Gruppen nicht der Chancen berauben, im Assessment-Center erfolgreich abzuschneiden. Bei der Anforderungsanalyse bzw. der Erstellung des Anforderungsprofils ist daher im ersten Schritt höchste Vorsicht geboten.

Folgende Maßnahmen können konkret auf die Vorbereitung und Durchführung eines diskriminierungsfreien Assessment-Centers hinwirken.

→ Es wird empfohlen, Anforderungen im Vorfeld so zu definieren, dass sie nicht zwischen einzelnen Bewerber\*innen-Gruppen diskriminieren (Stumpf et al., 2016). Ist dies geschehen, sind Unterschiede in der Leistung oder im Ver-

halten möglicherweise auf tatsächlich vorhandene Unterschiede in den Fähigkeiten oder dem Verhalten bestimmter Bewerber\*innen-Gruppen zurückzuführen. In diesem Fall – wenn das AC sauber konstruiert ist – soll es als „gutes“ Messinstrument diese Unterschiede für die Eignungsprognose aufdecken. Wenn dann einzelne Subgruppen schlechter abschneiden als andere, liegt dies nicht an der Qualität des Auswahlinstruments, sondern daran, dass tatsächlich Unterschiede in der Bewerber\*innen-Grundgesamtheit bestehen.

→ Alle Beurteiler\*innen sollten durch umfassende Beobachtungsschulungen dafür sensibilisiert werden, wie das Anforderungsprofil definiert ist und wie sie in der Beurteilung einer möglichen Diskriminierung entgegenwirken (vgl. auch Stumpf et al., 2016). In den wissenschaftlichen Studien zur Diskriminierung in Auswahlverfahren finden sich zahlreiche Studien zum Adverse Impact (s.o.). Dieser wird wissenschaftlich vor allem im Hinblick auf die Diskriminierung von schwarzen im Vergleich zu weißen Bewerber\*innen untersucht und ist vorrangig in der US-amerikanischen Auswahlpraxis seit Jahrzehnten Untersuchungsgegenstand. Aber auch andere Gruppen können davon betroffen sein. Adverse Impact könnte verringert werden, wenn in Beurteilungsgremien die Sensibilität für interkulturelle Anforderungen erhöht würde. Zum einen bedeutet dies, dass Anforderungen sich weniger an den Normen einer speziellen Kultur orientieren sollten. Definiert man den Führungsstil jedoch im Anforderungsprofil in Anlehnung an hierarchische Werte, ist es für Bewerber\*innen, die einen eher demokratischen Führungsstil kennen und leben, schwieriger zu überzeugen, obwohl es in der Realität

einer Organisation kontraproduktiv wäre, hierarchieorientiert zu führen. Zum anderen bedeutet die Sensibilisierung für interkulturelle Anforderungen auch, dass ein Beurteilungsgremium erkennt, wo und wann sich kulturelle Unterschiede im Verhalten in einem Assessment-Center zeigen können, und dass diese – nur, weil sie vom eigenen Verhaltensspektrum abweichen – nicht zwangsläufig für eine niedrigere Wahrscheinlichkeit für den beruflichen Erfolg sprechen. „Anders“ bedeutet zunächst nur „anders“, nicht aber „schlechter“. Die Absicht, Diskriminierung – auch unbewusste – im Assessment-Center zu reduzieren, geht mit dem Anspruch an ein reflektiertes und kulturell-offenes Beurteilungsgremium einher, das kulturell bedingte Unterschiede im Verhalten entsprechend erkennen, einordnen und anforderungsrelevante von zu vernachlässigenden Unterschieden differenzieren kann.

→ Festzuhalten gilt es auch für das Assessment-Center, dass das Ausmaß an Standardisierung bei der Durchführung, Auswertung und Ergebnisinterpretation in einem Assessment-Center stark darüber entscheidet, wie fair es ist (Gilliland, 1993). Dies spricht für das aufwendige Vorgehen in der Vorbereitung: Instruktionen und Aufgabenstellungen sollten für alle Bewerber\*innen gleich lauten, die äußeren Rahmenbedingungen (z.B. Ort und Ablauf) sollten für alle identisch sein, das Auswahlgremium sollte bei der Beurteilung aller Bewerber\*innen gleichartig vorgehen. Ebenso sollen eventuelle Rollenspieler\*innen für gleichartiges Verhalten geschult sein. Für die Ergebnisauswertung und -interpretation sollte sich im Vorfeld auf Cut-off- und Verrechnungsregeln geeinigt werden, die für alle Bewerber\*innen gleichermaßen

angewandt werden. Je stärker von dieser hohen Standardisierung, die in der Natur eines qualitativ guten Assessment-Centers liegt, abgewichen wird, desto anfälliger ist es dafür, ein diskriminierendes Instrument zu sein und das Gütekriterium der Fairness zu missachten.

→ Zu raten ist auch zu einem weitreichend anonymisierten Auswahlprozess, will man auf Fairness besonderen Wert legen. Erwiesenermaßen wirken sich Vorkenntnisse über Bewerber\*innen (z.B. Lebensläufe etc.) stark auf die Auswahlentscheidung aus (z.B. Kanning, 2015). Das ist qualitätseinschränkend, da dieses Vorgehen auch die Erkenntnisse aus dem standardisierendsten und einwandfreiesten vorbereiteten Assessment-Center aushebelt. Das Assessment-Center lässt ein eignungsdiagnostisches Urteil zu gezeigtem Verhalten zu; dieses mit den Urteilen Dritter (z.B. aus Arbeitszeugnissen) oder subjektiv interpretierten und manchmal modifizierten Informationen aus dem Lebenslauf zu vermengen, entspricht einem Vergleich von „Äpfeln und Birnen“. Zudem sind diese Informationen oft nicht sinnvoll in das Anforderungsprofil zu integrieren, sondern wirken sich häufig auf die Bewertung aller Anforderungsdimensionen aus, indem sie das Ergebnis „overall“ beschönigen oder verschlechtern. So wird die kontrollierte Subjektivität, die mit viel Mühe herzustellen versucht wurde, wieder zunichte gemacht. Es wird daher empfohlen, erst zuletzt in einem Auswahlprozess, wenn die Bewertung des Verhaltens im Assessment-Center abgeschlossen wurde, Informationen über die Bewerber\*innen offenzulegen und – unabhängig von den Erkenntnissen aus anderen Auswahlritten – diese in die Entscheidung über eine Einstellung einfließen zu lassen.

#### Frage 4: Welche Subgruppen werden häufig benachteiligt, wenn ein Assessment-Center unfair ist?

Das Allgemeine Gleichbehandlungsgesetz (AGG) bezieht sich auf die Diskriminierung von Menschen mit verschiedenen Merkmalen. Obwohl es das AGG bereits seit einigen Jahren gibt, sind systematische Unterschiede bei der Besetzung von Stellen immer wieder offensichtlich: Das klassische Beispiel ist hierfür die Unterrepräsentanz von Frauen in Führungspositionen (Kaup, 2015).

Dass auch Assessment-Center diskriminieren können – und Maßnahmen zur Verhinderung von Diskriminierung daher von besonderer Bedeutung sind – legen die folgenden Studienergebnisse (stellvertretend für viele andere) dar.

→ In einer Studie von Anderson, Lievens, vanDam und Born (2006) zeigte sich, dass Frauen in Assessment-Centers hinsichtlich der Indikatoren für den „interpersonally oriented leadership style“ besser bewertet wurden als männliche Bewerber. Dieser Rückschluss auf den Führungsstil wurde in der Art der mündlichen Kommunikation der Interaktionsform mit anderen messbar gemacht, die durch interaktive Assessment-Center-Stationen (z.B. Gruppenübungen) abgebildet wurden. Dieses Ergebnis reiht sich ein in verschiedene Befunde, die die Stärke weiblicher Bewerber\*innen bei Aufgaben zu zwischenmenschlichen Führungsaufgaben darlegen (Huffcut, Conway, Roth & Stone, 2001). Immer wieder zeigt sich in der Forschung, dass weibliche Bewerberinnen hinsichtlich dieser Faktoren besser abschneiden. Ist ein AC, welches diese sozialen Aspekte überprüft, und in dem Frauen häufig besser abschneiden als Männer, nun unfair? Die Frage ist mit „Nein“ zu beant-

worten, wenn diese zwischenmenschlichen Fähigkeiten zwischen erfolgreichen und weniger erfolgreichen Stelleninhaber\*innen trennen und das Kriterium damit tatsächlich relevant ist.

- Bewerber\*innen mit Migrationshintergrund schneiden in der Untersuchung von Stumpf et al. (2016) durchschnittlich schlechter im AC ab als Bewerber\*innen ohne Migrationshintergrund. Dieser Unterschied wird deutlicher, je präsenter (oder erkennbarer) dieser Migrationshintergrund zutage tritt. So werden Bewerber\*innen, die in Deutschland geboren wurden und aufgewachsen sind, deren (Groß-) Eltern einen Migrationshintergrund haben und die entsprechend in Deutschland von der Geburt an sozialisiert sind, besser bewertet als Bewerber\*innen, die eine ausländische Staatsangehörigkeit haben und auch eine Sozialisation einer anderen Kultur erlebt haben. Für diese Unterschiede ist in besonderem Maße zu sensibilisieren, da alle Beurteilenden hier Gefahr laufen, unbewusst Verzerrungstendenzen in ihrer Beurteilung zum Opfer zu fallen. Eine differenzierte Schulung – wie unter Frage 3 beschrieben – kann hier Abhilfe leisten.
- In einer Untersuchung von Truxillo, McCune, Bertolini und Fraccaroli (2012) wurde nachgewiesen, dass Arbeitnehmer\*innen unterschiedlichen Alters unterschiedliche Persönlichkeitseigenschaften zugeschrieben werden: So bewerteten Proband\*innen verschiedener Altersgruppen ältere Arbeitnehmer\*innen (55 bis 65 Jahre) durchschnittlich als gewissenhafter und zum Teil auch verträglicher, während jüngere Arbeitnehmer\*innen (24 bis 34 Jahre) von den Proband\*innen als durchschnittlich offener für Neues und proaktiver in ihrer Persönlichkeit bewertet wurden. Die Proband\*innen hatten in der Untersuchung über die Altersspanne hinaus keine weiteren Informationen über die zu bewertende Gruppe. Die Autor\*innen halten fest, dass allein diese Unterschiede in der Wahrnehmung der Persönlichkeitseigenschaften das Vorgehen in einem Auswahlprozess beeinflussen können. Diese These wird noch davon gestützt, dass die Proband\*innen die zu Bewertenden auch hinsichtlich einer Facette der Arbeitsleistung unterschiedlich bewerteten – obwohl ihnen keine Informationen über die tatsächliche Leistung gegeben wurden. Die älteren Arbeitnehmer\*innen wurden deutlich besser hinsichtlich des „Organisational Citizenship Behavior directed towards the organisation“ bewertet. Allein aufgrund des Alters schrieben die Proband\*innen den Arbeitnehmer\*innen also ein bestimmtes Verhalten zu, welches dem Arbeitgeber zuträglich ist und in seiner Zielerreichung unterstützt. Somit wird deutlich, dass die Altersstereotypisierung sehr schnell Effekte auf die Wahrnehmung der Eignung von Bewerber\*innen haben kann. Diese – vermutlich unterbewussten – Wahrnehmungen werden selten geprüft, in der Auswahlpraxis oft eher zu bestätigen versucht und münden nicht selten in einem nicht intendierten, aber dennoch maßgeblich diskriminierenden Vorgehen in der Personalauswahl.
- Dass auch Personen mit Behinderung von Diskriminierung in Auswahlverfahren betroffen sein können, zeigen Gouvier, Sytsma-Jordan und Mayville (2003): Sie machen deutlich, dass bei vorliegender Information über die Art der Behinderung (physisch oder mental) die Bewertung der sogenannten „employability“ oder Beschäftigungsfähigkeit bestimmt wird. Bei einer mentalen Behinderung fallen die employability-Bewertungen

schlechter aus als bei einer physischen Behinderung, es sei denn, es soll die employability für eine Hausmeister-Tätigkeit bewertet werden, die mit erhöhten physischen Anforderungen verbunden ist. Dieses Bewertungsmuster tritt auch auf, ohne dass den Befragten weitere Informationen über die Eignung der Bewerber\*innen zur Verfügung stehen. Auch hier werden voreilig stereotype Schlüsse gezogen, die in der Auswahlpraxis mit diskriminierender Personalauswahl einhergehen können.

Die aufgeführten Beispiele zeigen, dass eine Sensibilität für die häufig von unfairen Behandlung betroffenen Personengruppen vonnöten ist. An dieser Stelle sei dennoch noch einmal betont: Wenn Diskriminierung als Folge des Bezugs auf Anforderungen auftritt, handelt es sich nicht um Diskriminierung, sondern entweder um ein suboptimal definiertes Anforderungsprofil, welches die unfaire Behandlung hervorrufen kann, oder um Urteile, die verzerrt sind, beispielsweise, weil die Beurteilung nicht kontrolliert genug verlief, um subjektive Verzerrungstendenzen auszuschalten. Ist beides auszuschließen, so ist es möglich, dass das eignungsdiagnostische Instrument tatsächlich bestehende Unterschiede in der vorselektierten Bewerber\*innen-Gruppe abbildet.

Ohne in eine detaillierte Begründung für die Diskriminierung in der Auswahlpraxis einzusteigen, lässt sich in Bezug auf die oben genannten Studienergebnisse festhalten, dass neben strukturellen Ursachen, wie zum Beispiel der Bezugnahme auf nicht diskriminierungsfreie Anforderungsprofile, unterbewusste Einstellungen und Lernerfahrungen bei der Beurteilung eine Rolle spielen. Hier werden häufig Stereotype aktiviert, deren Gültigkeit jedweder empirischen Grundlage entbehrt. Diese Denkmuster zu verändern, ist eine langfristige

Aufgabe für unsere Gesellschaft. Sie kann momentan noch nicht vollständig erledigt, sondern nur durch eine gewissenhafte Orientierung an Maßnahmen, die Diskriminierung und Ungerechtigkeit in Auswahlverfahren entgegenwirken, angestoßen werden. Dass es diese Maßnahmen gibt, befreit jedes einzelne an einer Auswahlentscheidung beteiligte Individuum nicht davon, selbstreflektiert und im vollen Bewusstsein über diskriminierende Denkmechanismen zu urteilen und zu entscheiden.

#### Frage 5: Warum ist es relevant, ob die Bewerber\*innen ein Assessment-Center als fair betrachten?

Fairness-Wahrnehmung bestimmt mit, ob Bewerber\*innen einen Job nach erfolgreichem Auswahlverfahren annehmen, wie sie den Arbeitgeber wahrnehmen und ob sie den Arbeitgeber weiterempfehlen (Cable & Judge, 1996; Gilliland, Groth, Baker, Dew, Polly & Langon, 2001; Hausknecht, et al., 2004). Ein als fair bezeichneter Auswahlprozess spricht für einen Arbeitgeber, in der Hinsicht, dass dieser sich auf die tatsächliche Eignungsdiagnose konzentrieren möchte, wenn es um die Auswahl von neuem Personal geht. Wenn ein Auswahlverfahren als fair wahrgenommen wird, trägt dies zu einer positiven Wahrnehmung der Organisation in der Gesellschaft bei; es sichert dadurch auch ihr Fortbestehen.

Aufwändige Personalauswahlprozesse sollen sich lohnen: Personalentscheidungen und deren Qualität haben häufig vor allem langfristig gesehen einen finanziellen Mehrwert: Mitarbeiter\*innen, die nicht geeignet sind, sind auf Dauer viel kost intensiver als Mitarbeiter\*innen, die geeignet und dadurch erfolgreich in ihrem Job sind. Dadurch lohnt sich die Investition für eine saubere Vorbereitung des Assessment-Centers. In Assessment-Centers, in

denen zusätzlich auf Transparenz und Fairness sowie Faktoren, die zu einer wahrgenommenen Fairness beitragen, Wert gelegt wird, sind oft sehr aufwändig. Daher ist es wünschenswert, dass in solchen vorhersagestarken und durch ihren Anforderungsbezug fair gestalteten Auswahlverfahren als geeignet bewerteten Bewerber\*innen auch vom Arbeitgeber überzeugt sind und in ihrer Entscheidung, dort in ein Arbeitsverhältnis einzutreten, bekräftigt werden. Wenn die geeigneten Bewerber\*innen am Ende nicht bei einem Arbeitgeber einsteigen, ist auch ein sorgfältig konzipiertes und vorhersagestarkes Assessment-Center kein Beitrag zur Effizienz.

Insofern ist die wahrgenommene Fairness eines Assessment-Centers oder ganzen Auswahlprozesses als Hebel für effiziente Personalauswahl zu betrachten. Die Berücksichtigung dieses Nebengütekriteriums erspart dem Arbeitgeber unter Umständen hohe Kosten für die Neuausschreibung oder „schlechte“ Personalentscheidungen (Hausknecht et al., 2004).

#### Frage 6: **Wie kann man ein Assessment-Center so gestalten, dass die Bewerber\*innen es als fair wahrnehmen?**

Naheliegender ist es hier zunächst, dass den Bewerber\*innen das Gefühl vermittelt wird, fair behandelt zu werden. Ein wertschätzender Umgang, der Verzicht auf künstliche Stresssituationen, die in einem Assessment-Center oft erzeugt werden, ohne dass diese durch das Anforderungsprofil gerechtfertigt wären (z.B. durch Stressinterviews), bilden die Basis für eine faire Wahrnehmung des Assessment-Centers. Dazu zählt auch, dass der zeitliche Rahmen für den kompletten Auswahl- und Entscheidungsprozess möglichst klar definiert wird, damit Bewerber\*innen sich darauf einstellen können und nicht „in der Luft hängen“.

Transparenz auch in dem Sinne, dass das Anforderungsprofil für die zu besetzende Position den Bewerber\*innen im Vorfeld bekannt ist und auch klar ist, wie bei der Bewertung vorgegangen wird, trägt über diese Basis hinaus dazu bei, dass Bewerber\*innen sich fair behandelt fühlen. Häufig ist es hier nicht notwendig, dass die exakte Verrechnung der Ergebnisse offengelegt wird, sondern die Kenntnis, wer an der Bewertung beteiligt ist und welche Leistungen in die Bewertung einfließen. Immer wieder gehen Bewerber\*innen davon aus, dass auch ihr Verhalten zwischen den einzelnen Stationen eines Assessment-Centers bewertet wird, und sehen diese Zeit daher ebenfalls als Prüfungssituation an. Auch hier muss das Vorgehen klar definiert sein, wobei die DIN33430 sich ausdrücklich dafür ausspricht, Situationen, in denen Verhalten beurteilt wird, auch als solche zu kennzeichnen.

Darüber hinaus sollten bei der Vorbereitung und Konstruktion zwei wesentliche Prinzipien beachtet werden, die laut der wissenschaftlichen Literatur zur Wahrnehmung von Auswahlverfahren durch die Bewerber\*innen einen relevanten Einfluss zu haben scheinen:

Die sogenannte Job-relatedness des AC stellt einen Tätigkeitsbezug der einzelnen Stationen und inhaltlichen Schwerpunkte des Assessment-Centers zu den tatsächlichen Aufgaben einer/eines Stelleninhaber\*in her. Immer wieder zeigt sich in wissenschaftlichen Untersuchungen, dass ein hoher Tätigkeitsbezug des Assessment-Centers die wahrgenommene Fairness des Assessment-Centers positiv vorhersagt. Dies haben Zibarras und Patterson (2015) beispielsweise in einer Untersuchung gezeigt, in der die wahrgenommene Fairness der Bewerber\*innen nach dem Feedback zu der Leistung im Assessment-Center erfragt wurde.

Den Bewerber\*innen wurde mitgeteilt, ob sie das Assessment-Center erfolgreich bestanden oder nicht bestanden haben. Unabhängig vom Ergebnis fiel die Bewertung des Assessment-Centers positiver aus, wenn ein hoher Tätigkeitsbezug erkennbar wurde (Zibarras & Patterson, 2015). Dieses Argument wird dadurch untermauert, dass Assessment-Center häufig augenscheinlicher sind als kognitive Leistungstest (Marcan, Avedon, Pease & Smith, 1994). Das heißt, es fällt leichter, den Tätigkeitsbezug der Aufgabenstellungen zu erkennen, und dadurch erfolgt schnell – oft voreilig – der Schluss, dass Assessment-Center sei das bessere Messinstrument. Davon zu differenzieren ist jedoch die Thematik der Validität, die von der tatsächlichen – messbaren – Vorhersagestärke eines Instruments spricht. Diese kann, muss aber nicht mit der Augenscheinvalidität zusammenhängen. Für eine positive Wahrnehmung des Assessment-Centers ist Augenscheinvalidität, die durch job-relatedness hergestellt werden kann, zuträglich. Ein faires Auswahlverfahren basiert aber auf einer hohen Validität (s.o.) und stellt damit die Grundlage für ein tatsächlich und nicht nur wahrgenommen faires Assessment-Center dar.

Ein anderes Kriterium, wenn es um die wahrgenommene Fairness eines Assessment-Centers durch die Bewerber\*innen geht, ist die sogenannte Opportunity-to-perform, also die Chance, in diesem Assessment-Center zu zeigen, was man kann. In der Praxis bewerten Bewerber\*innen Assessment-Center häufig negativ, wenn sie ihre persönlichen Stärken nicht zeigen konnten, z.B., weil die Aufgaben diese nicht in den Vordergrund stellen. Häufig hört man dann Aussagen wie: „Es wurden mir nicht die richtigen Fragen gestellt.“. Sind die Aufgabenstellungen so gewählt, dass sie die relevanten Anforderungen für

die vakante Position abbilden können, tragen sie natürlich „objektiv“ zu einem fairen Assessment-Center bei. Bewerber\*innen, die ihre Stärken nach eigener Ansicht nicht präsentieren können, erfüllen dieses Anforderungsprofil demnach nicht.

Geht es jedoch darum, die Wahrnehmung der Bewerber\*innen zu verbessern, ist es wichtig zu wissen, dass die Opportunity-to-perform und die wahrgenommene eigene Leistung in einem Assessment-Center in Kombination vorhersagen, ob nicht-erfolgreiche Bewerber\*innen sich wieder bewerben (LaHuis, MacLane & Schlessmann, 2007). In der zitierten Studie sagt diese Kombination das Verhalten der nicht-erfolgreichen Bewerber\*innen sogar eher voraus als die job-relatedness.

Zusammenfassend lässt sich festhalten, dass die Fairness-Wahrnehmung der Bewerber\*innen auch durch andere Aspekte bestimmt wird als die „objektive“ Fairness, die man einem Auswahlinstrument zuschreiben oder sogar messen kann. In der Regel stehen diese Aspekte jedoch nicht im Widerspruch zueinander. Es ist durchaus möglich, ein faires – die relevanten Anforderungskriterien erfassendes – Assessment-Center zu konzipieren, durchzuführen und parallel auf Kriterien zu achten, die auch die Fairness-Wahrnehmung der Bewerber\*innen steigern. Dadurch wird die eignungsdiagnostische Qualität nicht eingeschränkt, sondern beide Zielsetzungen können koexistieren.

## Fazit

Für die faire Gestaltung eines (mehrstufigen) Personalauswahlprozesses ist nicht nur die objektiv faire Gestaltung wichtig, sondern auch die Fairnesswahrnehmung der Bewerber\*innen. Für alle Auswahlinstrumente gilt, dass ein Bezug auf eine solide Anforderungsanalyse unbedingt gegeben sein muss. Neben den Fairnesskriterien, die spezifisch für bestimmte Auswahlinstrumente sind, stellen Transparenz, respektvoller Umgang und Anforderungsbezug die wichtigsten Faktoren für die Fairnesswahrnehmung der Bewerber\*innen dar. Es zeigt sich, dass es sowohl beim Einsatz und der Konstruktion von Leistungstests und Assessment-Centern spezifische Fallstricke gibt, die die Fairness bedrohen können. Diese gilt es durch eine sorgfältige Konstruktion auszuschließen. Es ergibt sich aus der daraus entstehenden positiven Bewerber\*innen-Wahrnehmung und deren bereits beschriebenen Wirkung ein deutlicher Mehrwert für Organisationen. Ein diagnostisch hochwertiges Auswahlverfahren steht daher in keinem Gegensatz zur positiven Fairnesswahrnehmung der Bewerber\*innen.



## Literatur:

- Anderson, N., Lievens, F., van Dam, K., & Born, M. (2006). A construct-driven investigation of gender differences in a leadership-role assessment center. *Journal of Applied Psychology, 91*(3), 555.
- Anderson, N., Salgado, J. F., & Hülshager, U. R. (2010). Applicant Reactions in Selection: Comprehensive meta-analysis into reaction generalization versus situational specificity. *International Journal of Selection and Assessment, 18*, 291-304.
- Arvey, R. D., Strickland, W., Drauden, G., & Martin, C. (1990). Motivational components of test taking. *Personnel Psychology, 43*(4), 695-716.
- Blaker, N. M., Rompa, I., Dessing, I. H., Vriend, A. F., Herschberg, C., & Van Vugt, M. (2013). The height leadership advantage in men and women: Testing evolutionary psychology predictions about the perceptions of tall leaders. *Group Processes & Intergroup Relations, 16*(1), 17-27.
- Bühner, M., Ziegler, M., & Kersting, M. (2018). Statistisch-methodische Grundlagen der Eignungsbeurteilung. In D. u. Testkuratorium, *Personalauswahl kompetent gestalten* (S. 155-187).
- Cable, D. M., & Judge, T. A. (1996). Person-organization fit, job choice decisions, and organizational entry. *Organizational behavior and human decision processes, 67*(3), 294-311.
- DIN 33430:2016-07. (2016). Anforderungen an berufsbezogene Eignungsdiagnostik.
- Eid, M., & Schmidt, K. (2014). *Testtheorie und Testkonstruktion*. Göttingen: Hogrefe.
- Engelberg, P. (2015). Ursachen für Geschlechterdifferenzen in Tests des Allgemeinen Wissens. Dissertation. Bergischen Universität Wuppertal.
- Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of management review, 18*(4), 694-734.
- Gilliland, S., Groth, M., Baker, R., Dew, A., & Polly and J. L. Langdon: 2001, Improving Applicants Reactions to Rejection Letters: An Application of Fairness Theory. *Personnel Psychology, 54*, 669-703.
- Gouvier, W. D., Sytsma-Jordan, S., & Mayville, S. (2003). Patterns of discrimination in hiring job applicants with disabilities: The role of disability type, job complexity, and public contact. *Rehabilitation psychology, 48*(3), 175.
- Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel psychology, 57*(3), 639-683.
- Hough, L., Oswald, F., & Ployhart, R. (2001). Determinants, Detection and Amelioration of Adverse Impact in Personnel Selection Procedures: Issues, Evidence and Lessons Learned. *International Journal of Selection and Assessment, 9*, 152-194.
- Huffcutt, A. I., Conway, J. M., Roth, P. L., & Stone, N. J. (2001). Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Journal of Applied Psychology, 86*(5), 897.
- Jobmann, A.-L. (2017). Fairness der dgp Testbatterie F1: Eine Untersuchung der Messinvarianz. *dgp Informationen*(67), S. 60-70.
- Kanning, U. P. (2015). Sichtung von Bewerbungsunterlagen. In *Personalauswahl in der Wissenschaft* (pp. 83-101). Springer, Berlin, Heidelberg.
- Kaup, J. (2015). Status quo: Zur aktuellen Situation von Frauen in Führung. In *Die Unterrepräsentanz von Frauen in Führungspositionen* (pp. 13-26). Springer Gabler, Wiesbaden.
- Kersting, M. (2008). Zur Akzeptanz von Intelligenz- und Leistungstests. *Report Psychologie, 33*(9), 420-433.
- Kersting, M. (2018). Was König Bewerber denkt. Zur Akzeptanz von Personalauswahlverfahren. *Personalmagazin, 5*, 26-29.
- Koch, A. J., D'Mello, S. D., & Sackett, P. R. (2015). A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *Journal of Applied Psychology, 100*(1), 128.
- Kramer, J. (2009). Allgemeine Intelligenz und beruflicher Erfolg in Deutschland: vertiefende und weiterführende Metaanalysen. *Psychologische Rundschau, 60*(2), 82-98.
- LaHuis, D. M., MacLane, C. N., & Schlessman, B. R. (2007). Do applicants' perceptions matter? Investigating reappraisal behavior using fairness theory. *International Journal of Selection and Assessment, 15*(4), 383-393.
- Lievens, F. & Klimoski, R. J. (2001). Understanding the assessment center process: Where are we now? *International Review of Industrial and Organizational Psychology, 16*, 245-286
- Macan, T. H., Avedon, M. J., Paese, M., & Smith, D. E. (1994). The effects of applicants' reactions to cognitive ability tests and an assessment center. *Personnel Psychology, 47*(4), 715-738.
- Nyberg, L., Maitland, S., Rönnlund, M., Bäckman, L., Dixon, R., Wahlin, Å., & Nilsson, L.-G. (2003). Selective Adult Age Differences in an Age-Invariant Multifactor Model of Declarative Memory. *Psychology and Aging, 18*, 149-160.
- Ristel, N., & Jobmann, A.-L. (2017). Gendergerechte Personalauswahl – rechtliche, wissenschaftliche und praktische Hinweise. *dgp Informationen*.
- Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity research journal, 24*(1), 92-96.
- Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., & De Fruyt, F. (2003). International validity generalization of GMA and cognitive abilities: A European Community meta-analysis. *Personnel Psychology, 56*(3), 573-605.
- Schwarzer, R., & Jerusalem, M. (2002). Das Konzept der Selbstwirksamkeit. *Zeitschrift für Pädagogik, 44*, 28-53.
- 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*(2), 262.
- Sitzmann, T., & Yeo, G. (2013). A meta-analytic investigation of the within person self-efficacy domain: Is self-efficacy a product of past performance or a driver of future performance? *Personnel Psychology, 66*(3), 531-568.
- Standards for Educational and Psychological Testing. (2014). American Educational Research Association; American Psychological Association; National Council on Measurement in Education.
- 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, 31*(1), 4-31.
- Truxillo, D. M., Bodner, T. E., Bertolino, M., Bauer, T. N., & Yonce, C. A. (2009). Effects of Explanations on Applicant Reactions: A meta-analytic review. *International Journal of Selection and Assessment, 17*, 346-361.
- Truxillo, D. M., McCune, E. A., Bertolino, M., & Fraccaroli, F. (2012). Perceptions of older versus younger workers in terms of big five facets, proactive personality, cognitive ability, and job performance. *Journal of Applied Social Psychology, 42*(11), 2607-2639.
- Zibarras, L. D., & Patterson, F. (2015). The Role of Job Relatedness and Self-efficacy in Applicant Perceptions of Fairness in a High-stakes Selection Setting. *International Journal of Selection and Assessment, 23*(4), 332-344.

**Kontakt:**

M. Sc. Psychologie Julia Wagenschieber  
wagenschieber@dgp.de

M. Sc. Psychologie Amelie Kleinmanns  
kleinmanns@dgp.de

Deutsche Gesellschaft für Personalwesen e.V.  
Hohenzollernstr. 11-13, 40211 Düsseldorf