



Wie gehen Recruiter_innen mit der Herausforderung kulturübergreifender Bewerber_innenpools um?

Gamal Eid und Cornelius J. König 

Lehrstuhl für Arbeits- und Organisationspsychologie, Universität des Saarlandes, Saarbrücken, Deutschland

Zusammenfassung: Bisherige Forschungen zu interkulturellen Job-Interviews fokussieren vorwiegend die Sicht der Bewerbenden und implizieren eine Benachteiligung kulturferner Bewerber_innen, möglich wäre jedoch auch, dass diese Bewerber_innen Kriterien nicht erfüllen. In der vorliegenden Studie wurden Kurzgutachten und numerischen Beurteilungen von $N = 287$ Bewerbungsinterviews regionaler Flugbegleiter_innen qualitativ und quantitativ untersucht. Eine Besonderheit lag darin, dass sich die Bewerber_innen in ihrem jeweiligen Heimatland (China, Indien, Deutschland) bei einer deutschen Fluggesellschaft auf die gleiche Stelle beworben haben. Die Ergebnisse zeigen, dass Recruiter_innen zwar meistens in der Lage sind, ihre spontanen Reaktionen zu reflektieren, dass bei ihnen aber in einigen Fällen Stereotype wirksam werden. Dies und fehlende kulturspezifische Verhaltensanker können für die deutlich höhere Ablehnungsquote chinesischer Bewerber_innen mit verantwortlich sein.

Schlüsselwörter: Job Interview, kulturübergreifend, Personalauswahl

How Do Recruiters Manage the Challenge of Cross-Cultural Applicant Pools?

Abstract: To date, research regarding intercultural job interviews has mainly focused on the applicant's perspective, suggesting a disadvantage for applicants from more distanced cultural backgrounds. However, it is also possible that these applicants fail to meet certain recruiting criteria. This study examines short candidate reports and numerical evaluations of $N = 287$ recruiting interviews of regional flight attendants qualitatively and quantitatively. A peculiarity of this set-up was that applicants applied from their home country (China, India, and Germany) to the same position of this German airline. The results indicate that recruiters are largely able to reflect on their spontaneous reactions but likely also harbor some stereotypes. This and the absence of more culture-specific behavior anchors might be responsible for the higher rejection rate of Chinese applicants.

Keywords: job interview, crosscultural, personnel selection

Mit der internationalen Rekrutierung von Mitarbeitenden (Kholin & Blickle, 2015; Lipsmeyer & Zhu, 2011) sind etliche Herausforderungen verbundenen – z.B. müssen kulturell bedingte non-verbale Signale von Recruiter_innen korrekt wahrgenommen, interpretiert und beurteilt werden, ohne von Vorurteilen beeinflusst zu werden (Birkner, 2004; Imada & Hakel, 1977). Während die bisherigen Studien vor allem die Bewerber_innen fokussiert haben, werden die Herausforderungen für Recruiter_innen wenig diskutiert. Insbesondere gibt es kaum Forschung zu der Frage, wie Recruiter_innen versuchen, mit kulturellen Unterschieden bei Bewerber_innen und der daraus entstehenden interkulturellen Dynamik umzugehen (Manroop et al., 2013). Diese Studie soll deswegen quantitativ und qualitativ über-

prüfen, inwiefern sich in den Urteilen von Recruiter_innen die Herkunft der Bewerber_innen spiegeln. Dazu wurde eine Untersuchung von Jobinterviews für die Stelle einer Flugbegleiterin oder eines Flugbegleiters bei einer großen deutschen Fluggesellschaft durchgeführt. Das Besondere daran ist, dass für diesen Job auch im Ausland rekrutiert wird – das gleiche Suchprofil mit den gleichen Kriterien und Rahmenbedingungen wird für die gleiche Interviewsituation verwendet, was erlaubt, die Beurteilung der ausländischen Bewerber_innen (aus Indien und China) unmittelbar mit der Beurteilung deutscher Bewerber_innen zu vergleichen.

Hintergrund und Forschungsfragen

Einflussfaktoren auf die prognostische Validität von Einstellungsinterviews

Einstellungsinterviews sollen Informationen erheben, die dazu dienen, anforderungsbezogenes Verhalten möglichst zutreffend vorherzusagen (Westhoff & Strobel, 2011). Obwohl die Kriterien bezogene Validität von strukturierten Interviews nach einer neuen metaanalytischen Auswertungen bei $\rho = .42$ liegt (Sackett et al., 2021), gibt es Faktoren, die das Interviewergebnis unberechtigterweise beeinflussen können. Dazu zählt unter anderem der erste Eindruck, nonverbales Verhalten und Impression Management.

Erster Eindruck. Bei nahezu allen Jobinterviews, auch bei hoch strukturierten, gibt es zu Beginn des Gesprächs eine Begrüßungs- und Kennenlernphase (Schuler, 1992). Sie dient in erster Linie der Herstellung einer angenehmen Gesprächsatmosphäre und zum Aufbau einer ersten Beziehung zwischen der Recruiter_innen und der Bewerberin oder dem Bewerber. Diese Phase der ersten Eindrucksbildung kann unter Umständen die Urteilsbildung der Recruiterin und des Recruiters beeinflussen (Levashina et al., 2014; Swider et al., 2016).

Nonverbales Verhalten. Auch im weiteren Verlauf des Gesprächs werden nonverbale Signale registriert und bewertet, z. B. visuelle Signale wie Augenkontakt (Howard & Ferris, 1996). Nicht nur diese visuellen Signale, sondern auch statische wie Kleidung führen zu einer positiveren Bewertung der Bewerber_innen. Neben den visuellen haben auch die akustischen, paralinguistischen Signale (z. B. Sprechgeschwindigkeit) einen Einfluss auf die Urteilsbildung (DeGroot & Gooty, 2009). Recruiter_innen verwenden diese Fülle an Informationen, um sich ein Bild über die Persönlichkeitsmerkmale der Bewerber_innen zu machen. Wichtig ist in diesem Zusammenhang, dass diese Informationen nicht in jedem Fall die Fairness des Auswahlverfahrens beeinträchtigen: Solange die entsprechenden Persönlichkeitsmerkmale anforderungsrelevant sind und sie durch nonverbale Signale korrekt bewertet werden können, tragen sie durchaus zur prognostischen Validität des Verfahrens bei. Zu einer Benachteiligung der Bewerber_innen kann es kommen, wenn diese vermuteten Persönlichkeitsmerkmale lediglich mit der eigenen stereotypen Vorstellung der zu besetzenden Position abgeglichen werden und nicht mit dem Anforderungsprofil der zu besetzenden Stelle (DeGroot & Gooty, 2009). Wie zutreffend solche Beurteilungen und Rückschlüsse auf Grund kleiner Verhaltensausschnitte, wie sie im Einstellungsinterview gegeben sind, tatsächlich sind, ist jedoch umstritten. Einerseits gibt es Hinweise darauf, dass auch

nur kleine Verhaltensausschnitte relativ akkurat beurteilt werden können (Back & Nestler, 2016); andererseits darauf, dass die Akkuratheit der Beurteilung auch von der Anzahl der Informationen abhängt (Funder, 2012; Krzyzaniak et al., 2019). Diese sind im Einstellungsinterview auf Grund der kurzen Zeit und der für Bewerber_innen wahrgenommenen Ausnahmesituation, in der sie eher maximales als normales Verhalten zeigen (Klehe & Latham, 2006), eher begrenzt.

Impression Management. Der Selbstdarstellung der Bewerber_innen in Interviewsituationen kommt ein besonderer Stellenwert zu. Manche Bewerber_innen bereiten sich intensiv auf das Vorstellungsgespräch vor, um einen möglichst guten Eindruck zu hinterlassen (Levashina & Campion, 2006). Es gibt Hinweise darauf, dass das Bild, das Bewerber_innen vermitteln wollen, Einfluss auf ihre Beurteilung und auf ihre Chance, angestellt zu werden, hat (Barrick et al., 2009; Horverak et al., 2012). Das Eigeninterview kann als Situation verstanden werden, in der Erwartungen zu angemessenen Verhalten transportiert wird und Interviewte ihre Selbstpräsentation strategisch planen, um die Situationsanforderungen zu erfüllen (Jansen et al., 2012).

Kultur als Einflussfaktor

Viele dieser Verhaltensweisen sind auch kulturell bedingt. Beispielsweise sind in Kulturen mit große Machtdistanz Interaktionen zwischen Personen mit sehr unterschiedlicher gesellschaftlicher Stellung und Einfluss eher unerwünscht; folglich würden Bewerber_innen den Augenkontakt mit den Recruiter_innen als Zeichen des Respekts vermeiden (Neuliep, 2014). Ein anderes Beispiel ist, dass in Korea das Anlächeln von Fremden oder Autoritäten unerwünscht ist (Manroop et al., 2013). In allen Kulturen gibt es bestimmte Erwartungen zu nonverbialem Verhalten, und es kommt zu einer negativen Bewertung des Verhaltens und der Person, die die Erwartungen nicht erfüllt (Manroop et al., 2013). So werden Bewerber_innen gegebenenfalls negativ bewertet, wenn Handschlag und Lächeln fehlen. Recruiter_innen werden auf unerwartetes Verhalten der Bewerber_innen vermutlich eher mit einer negativen Beurteilung reagieren (Manroop et al., 2013; Reeder et al., 2004).

Allerdings ist eine methodische Schwäche der bisherigen Studien zu kulturell bedingten Schwierigkeiten in Interviewsituationen der Bezug auf Jobinterviews innerhalb eines Kulturraumes. Das heißt, die Recruiter_innen hatten es im gleichen Setting mit Bewerber_innen ohne Migrationshintergrund und solchen mit Migrationshintergrund zu tun. Untersuchungen zur kulturellen Benachteiligung werden deswegen möglicherweise durch bereits erfolgte

Anpassung der Bewerber_innen einer Minorität an die jeweilige Majoritätskultur verzerrt: Bewerber_innen, die schon länger in einem für sie fremden Kulturraum leben, lernen im Laufe der Zeit die spezifischen Erwartungen und Regeln der Majoritätskultur kennen, und passen ihr Verhalten im Alltag und dann eventuell auch in Interviewsituationen so an, wie sie vermuten, dass die Majoritätskultur es von ihnen erwartet (Bye et al., 2014; Horverak et al., 2013).

Schemata und Stereotype von Recruiter_innen. Recruiter_innen unterliegen, wie alle Menschen, einer kulturellen Sozialisierung (Pettigrew, 2016; Tajfel & Turner, 1986) und können sich nicht ohne weiteres Einflüssen durch kulturell bedingtes, ihnen fremdes Verhalten entziehen (Parsons & Liden, 1984; Roulin et al., 2014). Zudem lässt sich nicht ausschließen, dass Recruiter_innen voreingenommen oder vorurteilsbelastet sind (Byrne, 1997; Derous, 2017). Auch bestimmte Stereotype können Vorurteile aktivieren (Phills et al., 2020). Schemata, Stereotype und Vorurteile können dazu führen, dass sie bestimmte Verhaltensweisen der Bewerber_innen mit ungünstigen Persönlichkeitsmerkmalen verknüpfen (DeGroot & Gooty, 2009; Reeder et al., 2004) und damit Bewerber_innen zu Unrecht ablehnen. Allerdings ist der Begriff des Stereotyps nicht unproblematisch. Seit seiner Einführung (Lippmann, 1922) ist er durch weiche Formulierungen und vielen unterschiedlichen Sichtweisen geprägt (Stroebe & Insko, 1989). Während früher auf Merkmale wie beispielsweise „unkorrekte Übergeneralisierung“ als Bestandteil des Konzepts Stereotyp Bezug genommen und damit unweigerlich auch bewertet wurde, geht die Tendenz heute in eine abstraktere Richtung: Ein Stereotyp ist demnach ein subjektives Wahrscheinlichkeitsurteil über das Bestehen einer Verbindung zwischen einem Objekt und einem Attribut (Stroebe & Insko, 1989). Im weiteren Verlauf möchten wir Stereotypen als Wahrscheinlichkeitsurteile über die Merkmale, Eigenschaften oder Attribute von Personen definieren, die bestimmten Gruppen zugeordnet werden.

Adverse impact. Dieses aus den USA stammende Konzept beschäftigt sich seit den 1970er Jahren mit der Fairness und Chancengleichheit von Personalentscheidungen. Von einem *adverse impact* wird gesprochen, wenn es für unterschiedliche Bewerber_innen Gruppen erhebliche Unterschiede in der Erfolgsquote eines Auswahlverfahrens gibt. Als Definition für die erheblichen Unterschiede hat sich die 4/5-Regel etabliert (Outtz & Newman, 2010). Dies bedeutet, dass ein *adverse impact* immer dann vorliegt, wenn die Chancen für Minoritätsgruppen weniger als 80 % der Erfolgchancen der Majoritätsgruppen betragen. Findet man bei einem Personalauswahlverfahren einen *adverse impact*, so bedeutet es nicht zwangsläufig, dass die Ursache in einer Benachteiligung einer be-

stimmten Gruppe liegt. Es kann auch zu einer unterschiedlichen Erfolgsquote kommen, wenn in einer Subgruppe jobrelevante Merkmale abweichend ausgeprägt sind – wie zum Beispiel Kraft bei Männern und Frauen. Unternehmen sind deswegen gut beraten, ihr Auswahlverfahren immer wieder auf einen möglichen *adverse impact* zu überprüfen.

Forschungsfragen dieser Studie

Während die Perspektive der Bewerber_innen oft und intensiv diskutiert wird (Derous, 2017; Fell & König, 2016), findet man wenig Arbeiten zur Perspektive der Recruiter_innen (Macan, 2009). Aus diesem Grund soll der Fokus dieser Arbeit auf den Prozessen liegen, die sich bei den Recruiter_innen abspielen, also die Frage danach, wie Recruiter_innen mit der Herausforderung kulturübergreifender Bewerber_innenpools umgehen.

Gerade unter dem unmittelbaren Einfluss der Eindrücke und Informationen aus dem Jobinterview könnten die unter Umständen vorhandenen Schemata und Stereotype besonders wirksam sein. Vorausgesetzt, Recruiter_innen fixieren diese Eindrücke und Informationen schriftlich spontan und zeitnah, ohne Formulierungen zu überdenken, sollten sich in den entsprechenden Texten Hinweise auf Stereotype finden lassen (Bhatia, 2017). Daraus folgt:

Forschungsfrage 1: Finden sich in den schriftlichen Gutachten zur Begründung der Beurteilung von deutschen, indischen und chinesischen Bewerber_innen Hinweise auf Stereotype der deutschen Recruiter_innen?

Essentiell für die Urteilsbildung durch Interviewfragen ist die wahrgenommene Passung zwischen Bewerber_innen und den Herausforderungen der zu besetzenden Position (Bye et al., 2014; Derous, 2017). Die Wahrnehmung der Recruiter_innen unterliegt zwar den oben genannten Einflüssen, aber möglicherweise werden kulturelle Unterschiede schon von den Recruiter_innen reflektiert und berücksichtigt. Ablehnungen von Bewerber_innen würden sich vielmehr durch die Nichterfüllung jobrelevanter Kriterien ergeben. Vorausgesetzt, man hat mehrere vergleichbare Bewerber_innenpools aus unterschiedlichen Kulturkreisen für die gleiche Stelle, sollte ein Vergleich zwischen Jobinterviews von Bewerber_innen mit gleichem kulturellem Hintergrund wie der der Recruiter_innen und Jobinterviews kulturferner Bewerber_innen innerhalb der einzelnen relevanten Kriterien keinen signifikanten Unterschied ergeben. Falls die Recruiter_innen sich jedoch dem Einfluss eigener Stereotype nicht entziehen können, sollte es sowohl einen Unterschied in der Beurteilung der Bewerber_innen als auch einen Zusammenhang zwischen den Stereotypen und den Beurteilungen im Interview geben. Daraus folgt:

Forschungsfrage 2a: Unterscheiden sich die Beurteilungen von indischen und chinesischen Bewerber_innen durch erfahrene deutsche Recruiter_innen signifikant von den Beurteilungen deutscher Bewerber_innen?

Forschungsfrage 2b: Gibt es einen Zusammenhang zwischen Stereotypen gegenüber kulturfernen Bewerber_innen und der Beurteilung im Interview?

Methode

Kontext

Untersucht wurden die Kurzgutachten und numerischen Beurteilungen zu Bewerber_innen auf die Stelle einer Flugbegleiterin oder eines Flugbegleiters bei einer deutschen Fluggesellschaft. Verglichen wurden die entsprechenden Daten der Einstellungsinterviews in Deutschland mit denen in Indien und China, wo sich Bewerber_innen um eine Stelle als regionale Flugbegleiter_innen beworben haben.

Die Fluggesellschaft stellt einen besonderen Kontext, denn sie agiert international mit multikulturellen Kund_innen und Mitarbeitenden. Viele Arbeitsplätze, auch im Ausland, werden nach einheitlichen Auswahlkriterien besetzt. Dies galt auch für die sogenannten regionalen Flugbegleiter_innen, die für die Fluggesellschaft immer die Strecke zu oder von ihrem Heimatland als zusätzliches Crewmitglied fliegen. Sie sind nicht nur für die entsprechenden regionalen Fluggäste zuständig, sondern für alle, also auch für deutsche Fluggäste. Deutsche Recruiter_innen führten jeweils vor Ort in dem entsprechenden Kulturkreis (Deutschland, China und Indien) Interviews zu den identischen jobrelevanten Kriterien. Zum Zeitpunkt der Auswahlverfahren sowie der Datenerhebung waren alle Recruiter_innen ausgebildete Psycholog_innen, die als Freelancer_innen für die Fluggesellschaft arbeiten. Bevor die Psycholog_innen selbst Interviews führten, wurden sie durch Kolleg_innen intensiv mit dem Kompetenzmodell vertraut gemacht und die Verhaltensanker erläutert. Anschließend folgte eine Phase, in der neue Psycholog_innen gemeinsam mit erfahrenen Kolleg_innen Interviews führen und die Beurteilungen des gezeigten Verhaltens der Bewerber_innen gemeinsam besprechen. Bewerber_innen mussten die Kriterien grundsätzlich erfüllen, weil das Kompetenzmodell für alle Mitarbeiter_innen unabhängig von ihrer Herkunft bindend war.

Das Auswahlverfahren

Zum Zeitpunkt der Datenerhebung wurden viele neue Mitarbeiter_innen gesucht. Sie wurden gruppenweise eingeladen und nach einer Vorauswahl (Leistungstest und

Englisch-Sprachtest in Form eines Konversationsinterviews) zu einem abschließenden psychologischen Interview geführt. Jeweils eine Psychologin oder ein Psychologe interviewte eine Bewerberin oder einen Bewerber. Diese ca. 30–45 Minuten dauernden Interviews waren teilstrukturiert, das heißt die abzuarbeitenden Kriterien waren vorgegeben, die Reihenfolge und die Auswahl der Fragen waren jedoch dem jeweiligen Interviewer überlassen. Wichtige Kriterien waren Kontaktverhalten, Motivation und Berufsvorstellung, Serviceorientierung, Konfliktverhalten, Belastbarkeit, Teamorientierung und interkulturelle Kompetenz. Bei Gesprächen im Ausland wurden zusätzlich die Leistungsbereitschaft und das Erscheinungsbild beurteilt. Als Arbeitsmittel interviewten die Recruiter_innen einen kurzen schriftlichen Interviewleitfaden mit Raum für Beobachtungen und Bewertungen sowie mit einer Tabelle, auf der sie die jeweiligen Bewerber_innen für jedes Kriterium auf einer Skala von 1 bis 3 (1 = „Kriterium nicht erfüllt“, 2 = „Kriterium teilweise erfüllt“, 3 = „Kriterium voll und ganz erfüllt“) nach dem Gespräch beurteilen. Auch wenn es auf diesem Interviewleitfaden keine Verhaltensanker gab, die angemessenes und unangemessenes Verhalten beschreiben, waren diese Verhaltensanker jedoch in einem internen Dokument, das den Recruiter_innen zur Verfügung stand, ausführlich dokumentiert. Durch die oben beschriebene Einarbeitung waren die entsprechenden Operationalisierungen bekannt, auch wenn es keine speziellen oder auch kulturspezifischen Operationalisierungen für die Interviews mit den Bewerber_innen aus Indien oder China gab, weder als separates Dokument noch in den entsprechenden Interviewdokumentationen.

Eine Besonderheit dieser Interviews war ein kurzes, etwa fünf-minütiges Rollenspiel, in dem die Recruiter_innen die Rolle von Passagier_innen einnahmen. Sie übernahmen dabei die Rolle eines Gastes, der sich beschwert. Beobachtet wurde dabei insbesondere die Fähigkeit, die Konflikthaftigkeit zu erkennen und darauf angemessen zu reagieren. Je nach simulierter Situation wurde auch eine angemessene Durchsetzungsfähigkeit erwartet. Obwohl das Vorgehen, Recruiter_innen gleichzeitig als Rollenspieler_innen einzusetzen, nicht Vorgaben in der Literatur und der DIN 33430 (Westhoff et al., 2010) entspricht, wurde es zum Zeitpunkt der Datenerhebung auf Grund des hohen Informationsgehalts und der hohen Anzahl an Bewerber_innen so durchgeführt. Am Ende des Bogens wurde noch eine Gesamtbeurteilung abgegeben und eine Empfehlung, ob die Bewerberin oder der Bewerber für die Stelle geeignet ist oder nicht. An dieser Stelle gab es auch noch Raum für spezielle Stärken und Schwächen der Bewerber_innen. Dieses Feld wurde oft genutzt, um spezielle Entwicklungsfelder oder Einschränkungen zu beschreiben. Die Gesamtbeurteilung lag zu einem großen Teil im

Ermessen der jeweiligen Recruiter_innen. Es gab keinen festen Algorithmus zur Berechnung des Gesamturteils aus den Einzelkriterien; ein ausführliches Gutachten wurde nicht erstellt. Im Alltag wurde oft stichwortartig das gezeigte Verhalten durch adjektivische Beschreibungen (z. B. „nimmt die Bedürfnisse des Gastes wahr“, „bietet sofort Alternativen an“, „sehr leise um keinen Konflikt zu provozieren“, „gutes Umgangsgeschick in schwierigen Situationen“) oder auch die jeweilige Antwort (z. B. „always observe their needs“) notiert. Diese Notizen sind für eine qualitative Auswertung wertvoll, weil sie spontan, unter dem aktuellen Eindruck der Bewerberin oder des Bewerbers, angefertigt wurden.

Daten und Operationalisierung

Falls es Stereotype der Recruiter_innen geben sollte, müssten sich diese in den Notizen widerspiegeln. Die Operationalisierung erfolgte in Anlehnung an das Katz/Braly-Verfahren (Katz & Braly, 1933) die zur Operationalisierung von Stereotypen ihren Probanden eine Liste mit Adjektiven vorlegten, aus denen sie diejenigen auswählen sollten, die besonders typisch für bestimmte ethnische Gruppen seien. Diese häufig zur Messung von Stereotypen verwendete Methode (z. B. Schneider & Bos, 2014) ist zwar nicht unumstritten (Linville et al., 1986), jedoch in der Praxis bewährt, weil es vergleichsweise einfach in der Anwendung ist.

Zunächst wurden alle Gutachten digitalisiert und anschließend in SPSS (Version 25) sowie Atlas Ti 8 (Friese, 2018) überführt. Qualitativ wurden die adjektivischen Beschreibungen aus den Notizen extrahiert und beurteilt. In einem ersten Schritt wurde überprüft, ob die verwendeten Adjektive möglicherweise typisch für eine bestimmte ethnische Gruppe sind, und in einem zweiten Schritt, ob sie sich durch eine *pejorative* (implizit abwertende) Beschreibungen des gezeigten Verhaltens charakterisieren lassen oder durch eine *kulturneutrale* Beschreibung. Unter kulturneutral werden solche Beschreibungen verstanden, die neutral, also ohne Hinweise auf einen kulturellen Zusammenhang oder Missverständnis, formuliert wurden. Dagegen werden unter pejorativ solche Formulierungen verstanden, bei denen es möglicherweise eine bewusste oder unbewusste Fehlinterpretation der Recruiterin oder des Recruiters bezüglich des, zumindest zum Teil als fremdartig wahrgenommenen, Verhaltens gegeben haben könnte. Dies wäre ein Indiz für mögliche kulturelle Stereotype gegenüber chinesischen und indischen Bewerber_innen.

So sind Formulierungen wie „zeigt wenig Empathie“, „sehr oberflächlich, reflektiert nicht“, „kann Gast nicht beruhigen“, „unterbricht mehrfach“, „sehr bestimmt,

geht nicht auf Kompromiss ein“, „wenig Durchsetzungsvermögen“ und „bemüht, aber überfordert“ zwar durchaus negative Verhaltensbeschreibungen, tangieren aber die kulturelle Fairness nicht. Formulierungen wie zum Beispiel „wirkt unterwürfig und unsicher“, „zu überschwängliche, unangemessene Kontaktaufnahme“, „etwas überdreht – Authentizität?“ „ignoriert regionalen Kollegen“, „devot, wirkt sehr unsicher“, „undurchsichtig, glatt“, „sehr laut, viel Gestik“ „steif, sehr höflich“, „devot, bemüht, sehr chinesisch“ „puppenhaft“ und „extrem höflich, fast anstrengend“ enthalten Adjektive, die sich durchaus einer kulturellen Gruppe zuordnen lassen und zumindest das Wirksam-Werden kultureller Stereotypen bei den Recruiter_innen nahelegen. Um dieses subjektive Entscheidungskriterium zu untermauern, wurde diese Liste auch jeweils einer erfahrenen Recruiterin und einem erfahrenen Recruiter, die selbst kulturübergreifende Eignungsinterviews führen, als Subject Matter Experts vorgelegt. Solch ein Vorgehen wird empfohlen, um eine vielseitige Perspektive auf die vorhandenen Daten sicher zu stellen (Corbin & Strauss, 2008; Wilhelmy et al., 2016), und entspricht auch dem Vorgehen nach dem Katz/Braly-Verfahren, in dem der Grad der übereinstimmenden Nennungen ein Hinweis auf Stereotype sind (Katz & Braly, 1933). Nach einer intensiven Einarbeitung in die Fragestellung wurden sie gebeten, ebenfalls eine Kategorisierung der Zitate vorzunehmen.

Mittels der zwei Kategorien „pejorativ“ und „kulturneutral“ wurden in einem ersten Schritt in den 287 Gutachten $N = 1035$ Zitate kodiert. Durch mehrfach verwendete Beschreibungen konnte die Liste der verwendeten adjektivischen Formulierungen auf $N = 762$ gekürzt werden ($n = 453$ negative und $n = 309$ positive Beschreibungen). Bei den positiven adjektivischen Beschreibungen wurde von keinem der drei Rater_innen eine pejorative Formulierung zugewiesen. Die Raterübereinstimmung bei den verbleibenden $n = 453$ negativen Beschreibungen wurde mit Krippendorffs $\alpha = .885$ berechnet. Die erzielte Reliabilität ist als ausgezeichnet einzuschätzen (Wirtz & Caspar, 2007). Über die Formulierungen, für die es keine Übereinstimmung gab, wurde diskutiert, bis man sich auf eine Kategorisierung einigen konnte (wobei sich in den vielen Gesprächen die lange Erfahrung als sehr wertvoll erwies).

Im weiteren Verlauf der Auswertung der Daten wurden nur die Kriterien betrachtet, die einheitlich in allen drei betrachteten Herkunftsländern beurteilt wurden. Die Kriterien „Erscheinungsbild“ und „Leistungsbereitschaft“ sowie pejorative Beschreibungen im „Stärken/Schwächen“-Feld wurden nicht mit in die Auswertung mit einbezogen, da dort nur in wenigen Fällen überhaupt Eintragungen vorgenommen wurden.

Tabelle 1. Häufigkeit der pejorativen Beschreibungen in den Kategorien

Dimension	Land			$\chi^2(2)$	Cohens w	p
	China	Indien	Deutschland			
Kontaktverhalten	34	19	1	31.63	.33	< .001
Motivation	0	0	0	0	0	
Berufsvorstellung	0	2	0	2.68	.10	.26
Serviceorientierung	1	3	0	2.03	.09	.36
Konfliktverhalten	10	6	0	8.40	.17	.02
Belastbarkeit	0	0	0	0	0	
Teamorientierung	0	2	0	2.68	.10	.26
Interkulturelle Kompetenz	1	0	0	2.00	.08	.37

Stichprobe

Untersucht wurden $N = 324$ zufällig ausgewählte Kurzgutachten zu geführten Jobinterviews lokaler Flugbegleiter. Zu Beginn der Auswertung fiel die hohe Anzahl an fehlenden Eintragungen auf. In der Praxis wurde zum Zeitpunkt der Datenerhebung auf Teile des Kriterienkatalogs verzichtet, wenn schon zu Beginn oder im Verlauf des Interviews klar wurde, dass Bewerber_innen zum Beispiel auf Grund mangelnder Sprachkenntnisse für das Unternehmen nicht in Frage kommen oder sich durch das gezeigte Verhalten bei einem Kriterium das explizite Eingehen auf ein anderes Kriterium erübrigt. Toleriert für die Aufnahme in die Auswertung wurden ca. 20% fehlende Werte. Aufgrund zu vieler fehlender Eintragungen wurden 37 Gutachten von der Auswertung ausgeschlossen, so dass letztlich $N = 287$ Gutachten untersucht wurden. Davon wurden $n = 68$ (24%) in Deutschland, $n = 96$ (33%) in Indien und $n = 123$ (43%) in China geführt; 134 Bewerber_innen waren weiblich (47%) 149 waren männlich (53%), bei 4 Bewerber_innen ließ sich das Geschlecht aus den Gutachten nicht mehr feststellen. Das Alter der Bewerber_innen lag zwischen 18 und 54 Jahren ($M = 27.1$, $SD = 4.8$), wobei 82% der Bewerber_innen 30 Jahre und jünger waren. Insgesamt waren 16 Recruiter_innen an der Auswahl beteiligt, von denen fünf (2 Frauen und 3 Männer) auf Grund ihrer Einsätze im Ausland etwa 83% der Gutachten erstellt haben. Eine Zusage und damit ein Jobangebot bekamen 47% der Bewerber_innen aus Indien, 44% der deutschen Bewerber_innen, jedoch nur 28% der chinesischen Bewerber_innen. Damit betrug der *adverse impact* Faktor bei chinesischen Bewerber_innen nur .59 verglichen mit den indischen Bewerber_innen als Gruppe mit der höchsten Angebotsquote (und bei deutschen Bewerber_innen .94).

Kommunikative Validierung

Nach der Datenauswertung wurden noch einmal Einzelgespräche mit fünf Recruiter_innen geführt. Dieses auch als kommunikative Validierung bekannte Verfahren dient zum einen der Überprüfung der Ergebnisse und der Zuordnung der Kategorien, zum anderen können damit möglicherweise Erklärungsansätze für Befunde generiert werden (Bluhm et al., 2011). Vor den Gesprächen wurde ein schriftlicher Leitfaden erarbeitet, in dem alle relevanten Punkte, die zu besprechen waren, festgehalten wurden (Kategorisierung, Befunde zu den Einzelkriterien, mögliche Unterschiede in der Interviewsituation, mögliche Ursachen für die gefundenen Ergebnisse). Die Gespräche wurden telefonisch geführt und handschriftlich, stichpunktartig protokolliert. Die Auswertung der Protokolle erfolgte digital. Anschließend wurden Aussagen zu den Einzelkriterien im Interview, Unterschiede im Interview-Setting sowie mögliche Ursachen für die gefundenen Ergebnisse gebündelt und nach Häufigkeit ihrer Nennung sortiert. Für Nachfragen wurde vereinzelt noch einmal telefonisch Kontakt zu den jeweiligen Recruiter_innen aufgenommen.

Ergebnisse

Qualitative Analysen

Die qualitative Auswertung der Gutachten ergab ein überraschendes Bild. Einerseits waren in 73.9% der Gutachten alle Verhaltensbeschreibungen kulturneutral, das heißt, es gab keinerlei Hinweise auf negative kulturelle Stereotype (s. Tabelle 1). Andererseits fanden sich in 20.9% ($n = 58$) der Gutachten eine, in 4.9% ($n = 14$) zwei und in einem Fall sogar drei pejorative Verhaltensbeschreibungen. Betrachtet man die Kriterien genauer, bei denen laut Literatur die häufigsten kulturell bedingten

Tabelle 2. Kriterienspezifische Unterschiede nach Herkunftsländern

Dimension	Indien		China		Deutschland		Kruskal-Wallis H			Effektstärke
	<i>n</i>	MR	<i>n</i>	MR	<i>n</i>	MR	χ^2	<i>df</i>	<i>p</i>	η^2
Kontaktverhalten	94	133.1	97	124.6	67	131.5	.77	2	.68	.005
Motivation	94	133.0	98	118.7	66	140.5	4.20	2	.12	.009
Berufsvorstellung	89	116.1	72	112.6	67	114.4	.13	2	.94	.008
Serviceorientierung	93	124.7	86	130.4	68	115.0	2.07	2	.36	< .001
Konfliktverhalten	93	139.1	109	134.8	66	127.5	.99	2	.61	.004
Belastbarkeit	89	127.4	75	101.7	65	113.4	7.26	2	.03*	.023
Teamorientierung	93	112.0	64	118.9	64	101.6	2.82	2	.24	.004
IKK	92	109.1	72	116.5	67	125.0	2.58	2	.27	.003
Gesamtbewertung	96	159.0	120	123.1	67	151.4	15.14	2	<.001**	.047

Anmerkungen: MR = Mittlerer Rang, IKK = Interkulturelle Kompetenz. * $p < .05$, ** $p < .001$.

Missverständnisse zu vermuten sind, nämlich Konflikt- und Kontaktverhalten (Macan, 2009; Manroop et al., 2013), so zeigt sich dort eine Anhäufung pejorativer Verhaltensbeschreibungen. Die Beschreibungen des Konfliktverhaltens enthalten $n = 16$ pejorative Formulierungen, $\chi^2(2) = 8.40$, $p = .02$, Cohens $w = .17$, die des Kontaktverhaltens sogar $n = 54$, $\chi^2(2) = 31.63$, $p < .001$, Cohens $w = .33$. Eine differenzierte Betrachtung nach Herkunftsländern der Bewerber_innen zeigt, dass bei chinesischen Bewerber_innen 42.5% aller das Kontaktverhalten beschreibenden Kommentare als pejorativ einzuordnen sind, bei indischen Bewerber_innen sind es 43.2%. Zum Vergleich findet sich nur in einem Gutachten deutscher Bewerber_innen eine pejorative Beschreibung des Kontaktverhaltens. Die Forschungsfrage F1, ob sich in den schriftlichen Gutachten zur Begründung der Beurteilung von deutschen indischen und chinesischen Bewerber_innen Hinweise auf Stereotype der Recruiter_innen finden, muss dementsprechend bejaht werden, wenn gleich die Effekte nach Cohen klein sind. Dies bestätigt die Annahme, dass auch bei erfahrenen Recruiter_innen gegenüber fremden Kulturen Stereotype wirksam werden: Sie scheinen spontan auf ihnen fremde Verhaltensmuster zu reagieren (Ashburn-Nardo et al., 2001) und dies auch sprachlich zum Ausdruck zu bringen.

Quantitative Analysen

Da die oben beschriebene Skala in den untersuchten Gutachten (1 = „Kriterium nicht erfüllt“, 2 = „Kriterium teilweise erfüllt“, 3 = „Kriterium voll und ganz erfüllt“) nicht die Voraussetzungen einer Intervallskalierung erfüllt, weil die mittlere Stufe viel Spielraum für die individuelle Beurteilung der Einschränkungen lässt, sind die

Daten als ordinalskaliert zu betrachten, weswegen der Kruskal-Wallis H-Test verwendet wurde.

Die quantitative Analyse des Datensatzes zeigte ebenfalls ein unerwartetes Bild (s. Tabelle 2). Betrachtet man die einzelnen Kriterien, findet man kaum Unterschiede in der Beurteilung der Bewerber_innen aus den drei Herkunftsländern Indien, China und Deutschland. Bei nahezu allen Kriterien sind die Unterschiede im mittleren Rang nicht signifikant. Allerdings weist das Kriterium Belastbarkeit deutliche Unterschiede auf, $\chi^2(2) = 7.26$, $p = .026$. Außerdem erzielten chinesische Bewerber_innen im Vergleich mit indischen und deutschen Bewerber_innen bei dem Kriterium Motivation deutlich schlechtere Werte ($MR = 118.7$). Die Gesamtbeurteilung unterscheidet sich signifikant zwischen den drei Gruppen, $\chi^2(2) = 15.14$, $p < .001$. Bei den Kriterien, bei denen laut Literatur kulturell bedingten Missverständnisse am häufigsten zu vermuten sind, nämlich Konflikt- und Kontaktverhalten (Macan, 2009; Manroop et al., 2013), findet man die geringsten Unterschiede in der Beurteilung. So ergab sich für Kontaktverhalten ein $\chi^2(2) = 0.77$, $p = .68$, und für Konfliktverhalten ein $\chi^2(2) = 0.99$, $p = .61$. Allerdings zeigt eine Betrachtung der Einzelkriterien aufgeteilt nach Herkunftsländern (Tabelle 3), dass chinesische Bewerber_innen bei vielen Kriterien signifikant häufiger das Urteil 2 „erfüllt Kriterium teilweise“ bekamen als die Bewerber_innen aus Indien und China. So bekamen mehr als 50% der chinesische Bewerber_innen bei nahezu allen Kriterien das Urteil 2, und die Unterschiede waren bis auf die Kriterien Berufsvorstellung und Teamorientierung signifikant. Interessant ist jedoch, dass bei dem Gesamturteil 73.3% der chinesischen Bewerber_innen das Urteil 1 (erfüllt die Anforderungen nicht) bekamen, $\chi^2(2) = 15.44$, $p = .004$, wodurch eine *adverse impact ratio* von .59 resultierte.

Insgesamt lässt sich bei Betrachtung der Tabellen 2 und 3 die Forschungsfrage F2a, ob sich die Beurteilungen von

Tabelle 3. Verteilung der Beurteilungen über die Länder

Dimension	Urteil	Anteil in %			$\chi^2(2)$	Cohens w	p
		Indien	China	Deutschland			
Kontaktverhalten	1	28.7	26.08	35.8	11.84	.21	.02
	2	39.4	52.6	26.9			
	3	31.9	20.6	37.3			
Motivation	1	40.4	38.8	30.3	15.66	.25	.004
	2	27.7	49.0	40.9			
	3	31.9	12.2	28.8			
Berufsvorstellung	1	21.4	12.5	19.4	6.74	.17	.15
	2	43.8	63.9	49.3			
	3	34.8	23.6	34.3			
Serviceorientierung	1	22.6	10.5	29.4	11.19	.21	.02
	2	37.6	53.5	35.3			
	3	39.8	36.0	35.3			
Konfliktverhalten	1	35.5	29.4	40.9	9.35	.19	.05
	2	33.3	51.4	34.8			
	3	31.2	19.2	24.3			
Belastbarkeit	1	14.6	18.7	16.9	10.07	.21	.04
	2	29.2	49.3	40.0			
	3	56.2	32.0	43.1			
Teamorientierung	1	19.3	10.9	21.9	4.29	.14	.37
	2	39.8	46.9	46.9			
	3	40.9	42.2	31.2			
Interkulturelle Kompetenz	1	35.9	18.0	25.4	16.38	.27	.003
	2	38.0	65.3	40.3			
	3	26.1	16.7	34.3			
Gesamturteil	1	50.0	73.3	55.2	15.44	.23	.004
	2	28.1	19.2	25.4			
	3	21.9	7.5	19.4			

Anmerkungen: Urteil 1 = erfüllt Anforderung nicht, 2 = erfüllt Anforderung mit Einschränkungen, 3 = erfüllt Anforderungen.

indischen und chinesischen Bewerber_innen durch deutsche Recruiter_innen signifikant von den Beurteilungen deutscher Bewerber_innen unterscheiden, wie folgt beantwortet: Die mittlere Beurteilung der Bewerber_innen aus China, Indien und Deutschland unterscheidet sich zwar in den mittleren Rängen nicht, mit Ausnahme des Kriteriums Belastbarkeit, trotzdem werden signifikant mehr Bewerber_innen aus China und Indien abgelehnt und bekommen eine niedrigere Gesamtbeurteilung als Bewerber_innen aus Deutschland. Auffällig ist der hohe Anteil chinesischer Bewerber_innen mit der Beurteilung 2 (erfüllt Anforderung teilweise) und die signifikant hohe Ablehnungsquote chinesischer Bewerber_innen, während es bei indischen Bewerber_innen kaum Unterschiede in der Beurteilung gibt.

Zur Beantwortung der Forschungsfrage F2b, ob es einen Zusammenhang zwischen möglichen Stereotypen gegenüber chinesischen und indischen Bewerber_innen

und der Beurteilung im Interview gibt, wurden Korrelationen zwischen der Anzahl pejorativer Zitate und der mittleren Beurteilung in den einzelnen Kriterien berechnet (s. Tabelle 4). Die Übersicht macht deutlich, dass die Bewerber_innen, bei denen sich mindestens eine pejorative Verhaltensbeschreibung im Gutachten fand, deutlich weniger Zusagen bekamen (14.6 %) als die ohne entsprechende Formulierungen (85.4 %). Keinen Einfluss scheint die Verwendung pejorativer Formulierungen auf die wahrgenommene Belastbarkeit der Bewerber_innen zu haben. Die Frage F2b „Gibt es einen Zusammenhang zwischen möglichen Stereotypen gegenüber kulturfernen Bewerber_innen und der Beurteilung im Interview?“ lässt sich also aufgrund dieser Daten nicht klar beantworten. Zwar gibt es einen in der Höhe niedrigen, aber statistisch signifikanten negativen Zusammenhang zwischen den Kriterien Bewertungen und der Anzahl pejorativer Beschreibungen, aber es bleibt unklar, warum diese Vorbe-

halte nur bei dem Kriterium Belastbarkeit wirksam werden sollten, denn bei den anderen Kriterien lässt sich kein Zusammenhang feststellen. An dieser Stelle könnte man auch vermuten, dass lediglich fremdartig erlebtes Verhalten zu tendenziell schlechteren Beurteilungen führt. Die anschließenden Gespräche mit den betroffenen Recruiter_innen haben dazu mehr Klarheit gebracht.

Tabelle 4. Korrelationen der Anzahl pejorativer Zitate mit Beurteilungen über ausgewählte Kriterien

Dimension	Beurteilung	Prozentualer Anteil	Rangkorrelation (Kendalls τ_b)
Entscheidung	Zusage	14.6 %	-.21**
	Absage	85.4 %	
Belastbarkeit			.00
Kontaktverhalten			-.17**
Motivation			-.14*
Serviceorientierung			-.04
Konfliktverhalten			-.13*
Teamorientierung			.02
Interkulturelle Kompetenz			-.18**
Gesamtbewertung			-.20**

Anmerkungen: Anteil der Gutachten mit mindestens einer pejorativen Beschreibung über alle Kriterien in %. * $p < .05$, ** $p < .001$.

Kommunikative Validierung

Im Rahmen der kommunikativen Validierung wurde deutlich, dass es für die Recruiter_innen durchaus nicht einfach ist, dem unter Umständen ungewohnten Verhalten der Bewerber_innen neutral zu begegnen. Im Folgenden werden die wichtigsten Ergebnisse dieser Gespräche zusammengefasst.

Kontakt-Verhalten. Besonders schwierig sei die Beurteilung des Kontaktverhaltens. Dieses sei nur beobachtbar und es gäbe in dem teilstrukturierten Interviewleitfaden keine Hinweise oder Verhaltensanker, die gewünschtes von ungewünschtem Verhalten trennen würden. Das Verhalten der Bewerber_innen erscheine den Recruiter_innen teilweise sehr fremd und andersartig. Die Recruiter_innen betonten in den Gesprächen die Wichtigkeit, Verhaltensauffälligkeiten zu dokumentieren und in der anschließenden Beurteilung zwischen akzeptablen und nicht akzeptablen Auffälligkeiten zu trennen. So könne es zum Beispiel vorkommen, dass eine Bewerberin als *puppenhaft* beschrieben werde, weil es der Recruiterin oder dem Recruiter besonders auffalle, sich diese_r aber im Nachgang bei der Beurteilung sehr wohl darüber im Klaren sei, dass solch eine Wahrnehmung kein Grund für eine Abwertung sei. Anders verhalte es sich bei kulturellen

Auffälligkeiten im Kontakt, die nicht zur Organisation passen würden: Wenn etwa ein Bewerber die am Gespräch teilnehmende weibliche Führungskraft völlig ignoriere, führe das nicht nur zu einer negativen Beschreibung, sondern auch in der Beurteilung zu einer Abwertung.

Konflikt-Verhalten. Eine Besonderheit der Eignungsinterviews für Flugbegleiter_innen sei die Erfassung des Kriteriums Konfliktverhalten. Gerade im Rollenspiel seien kulturelle Unterschiede deutlich. So würden asiatische Bewerber_innen vermehrt den Weg der Konfliktvermeidung wählen und seien darauf bedacht, dass der Gast sein Gesicht nicht verliere. Dieses Verhalten könne als fehlendes Durchsetzungsvermögen und damit als nicht den Anforderungen entsprechendes Konfliktverhalten fehlinterpretiert werden.

Belastbarkeit. Alle befragten Recruiter_innen machten deutlich, dass chinesische Bewerber_innen sehr oft nur unzureichende Stressbewältigungsstrategien formulieren könnten und deutlich nervöser aufgetreten seien als indische und deutsche Bewerber_innen. Eine Besonderheit sei allerdings die Interviewsituation, die trotz aller Bemühungen zur Standardisierung nicht in allen Aspekten vergleichbar sei. Zum einen sei noch eine erfahrene chinesische Flugbegleiterin bei dem Einstellungsinterview anwesend gewesen, was den Druck auf die Bewerber_innen erhöht haben könne. Zum anderen seien die örtlichen Rahmenbedingungen sehr viel ungünstiger gewesen. Der Wartebereich für die Bewerber_innen sei sehr unruhig und wenig separiert von anderen Personen, die sich dort aufgehalten haben. Trotzdem halte sich bei den befragten Recruiter_innen der Eindruck, dass chinesische Bewerber_innen tatsächlich weniger belastbar seien als indische oder deutsche Bewerber_innen.

Motivation. Vor allem bei chinesischen Bewerber_innen sei die Motivation oft unklar. Während in Deutschland der Beruf Flugbegleiter_in nicht besonders hoch angesehen sei, genieße er in China und Indien sehr hohes Ansehen. Außerdem würde die Beschäftigung bei einem deutschen Unternehmen die Möglichkeit bieten, die Familie zu versorgen, und eventuell sogar, in Deutschland zu leben. Tatsächlich sei von chinesischen Bewerber_innen oft auch die Chance, eine_n Partner_in zu finden, als Motivation benannt worden. Diese Antwort führe zu einer Abwertung im Kriterium Motivation und oft auch zu einer Ablehnung der Bewerber_innen, da der entsprechende Verhaltensanker vorwiegend intrinsische Motive (z. B. Freude am Service, Freude am Umgang mit Menschen) vorgibt.

Diskussion

Die Analyse der Gutachten ergibt ein unerwartetes Bild zu den Schwierigkeiten und Herausforderungen für Recruiter_innen im Umgang und in der Beurteilung eines kulturübergreifenden Bewerber_innenpools. Die Datenanalysen und die im Nachgang geführten Gespräche mit den Psychologen_innen deuten darauf hin, dass sich die Recruiter_innen bemühen, kulturferne Bewerber_innen möglichst fehlerarm und Kriterien orientiert zu beurteilen, auch wenn Recruiter_innen sich den kulturellen Besonderheiten sowohl im Kontakt- und Konfliktverhalten, als auch bei der Motivation der Bewerber_innen nicht entziehen können und mit ihren pejorativen Beschreibungen eine spontane Reaktion zeigen. Es spricht aber für ein kulturelles Bewusstsein, dass ihre numerischen Beurteilungen im Mittel nicht schlechter ausfallen als bei deutschen Bewerber_innen. Nur bei genauerer Betrachtung der Urteile zu den einzelnen Kriterien fällt der hohe Anteil an mittleren Bewertungen bei den chinesischen Bewerber_innen auf.

Zwar gibt es im Mittel keine statistisch nachweisbaren Unterschiede in der Beurteilung der Bewerber_innen aus verschiedenen Kulturkreisen in den einzelnen Kriterien, trotzdem werden signifikante Unterschiede in dem Gesamturteil und damit in der Entscheidung, die Bewerberin und Bewerber anzunehmen, gemacht. Darüber hinaus fallen gerade bei chinesischen Bewerber_innen die hohe Anzahl an Einschränkungen auf und der signifikant hohe Anteil an Bewerber_innen mit dem Gesamturteil, dass sie den Anforderungen nicht genügen, was sich auch in der niedrigen *adverse impact ratio* ($< .8$) widerspiegelt. Damit entsprechen die Befunde den Vermutungen von (Huffcutt, 2011), dass es selbst dann, wenn nur jobrelevante Konstrukte betrachtet werden, kulturelle Unterschiede zwischen Interviewer und Interviewten die Auswahlentscheidung beeinflussen können – ein Hinweis auf eine Benachteiligung kulturferner Bewerber_innen (Kholin & Blicke, 2015).

Der Einfluss kultureller Besonderheiten macht sich besonders bei der Beurteilung des Kontaktverhaltens bemerkbar. Bei diesem Kriterium spielt vor allem das non-verbale Verhalten eine Rolle, das laut verschiedener Studien während des Interviews die Eindrucksbildung der Recruiter_innen beeinflusst (Levashina et al., 2014; Swider et al., 2016). Beim Kontaktverhalten finden sich in dieser Studie die meisten pejorativen Verhaltensbeschreibungen; trotzdem findet man keine Unterschiede in der numerischen Beurteilung der Bewerber_innen. Daraus könnte man schließen, dass sich die Recruiter_innen in diesem frühen Stadium des Interviews durchaus der „Andersartigkeit“ der Bewerber_innen bewusst sind (Manroop et al., 2013; Parsons & Liden, 1984). Tatsäch-

lich bestätigten die Recruiter_innen in den Gesprächen die Wichtigkeit, Verhaltensauffälligkeiten zu dokumentieren, jedoch in der anschließenden Beurteilung zwischen akzeptablen kulturell bedingten und nicht akzeptablen Auffälligkeiten zu trennen. Nicht akzeptabel seien in diesem Zusammenhang Eigenschaften, die nicht zu der Organisation passen wie das Ignorieren weiblicher Führungskräfte (Bye et al., 2011; Horverak et al., 2012). Allerdings sind diese nicht akzeptablen Eigenschaften in den Interviewleitfäden nicht klar definiert und die Verhaltensanker beziehen sich auf den Kontakt mit einheimischen Bewerber_innen. Dieser fehlende Bezug zum jeweiligen Kulturkreis lässt den Recruiter_innen Raum für eigene Werte, und die wahrgenommene Passung der Bewerber_innen zur Organisation kann so zwischen den einzelnen Recruiter_innen differieren.

Ähnliches gilt auch für das Konfliktverhalten, einem weiteren Kriterium der Fluggesellschaft. Auch bei diesem Kriterium finden sich viele pejorative Verhaltensbeschreibungen. Auffällig ist jedoch, dass es auch bei dem Konfliktverhalten keine numerischen Unterschiede zwischen den Kulturen gibt. Auch hier scheinen sich die Recruiter_innen der kulturellen Unterschiede bewusst zu sein, und sie lassen es anscheinend nicht in die Beurteilung mit einfließen. Jedoch fehlen auch bei diesem Kriterium kulturell angepasste Verhaltensanker, sodass die gleichen Schwierigkeiten bei der numerischen Beurteilung entstehen.

Schwierig hingegen ist die Beurteilung der Ergebnisse in Bezug auf die Belastbarkeit der Bewerber_innen: Hier finden sich deutliche Unterschiede in der numerischen Beurteilung, jedoch keine pejorativen Zitate. Zunächst ist es wenig nachvollziehbar, warum chinesische Bewerber_innen weniger belastbar sein sollten als indische oder deutsche Bewerber_innen (s. auch (Schmitt et al., 2007), da die meisten von ihnen schon bei anderen Fluggesellschaften geflogen sind und die spezifische Belastung vermutlich ausgehalten haben. Auch die im Nachgang geführten Gespräche mit den Recruiter_innen brachten nicht die erwünschte Klarheit: Bewerber_innen seien sowohl durch die im Interview gestellten Fragen nach Stressbewältigungsstrategien als auch durch ihren nervöseren Eindruck als weniger belastbar eingestuft worden. Auch bei diesem Kriterium scheinen Recruiter_innen durch die fehlenden kulturspezifischen Verhaltensanker zu viel Ermessensspielraum für ihre Beurteilungen zu haben.

Ein kritischer Punkt scheint auch die Beurteilung der „richtigen“ Motivation der Bewerber_innen zu sein. Es gibt Hinweise darauf, dass sich die Motivation für einen bestimmten Job durchaus zwischen den verschiedenen Kulturen unterscheidet (Monnot, 2018): So erwartet man in individualistischen Kulturkreisen eine hohe intrinsische Motivation, in kollektivistischen Kulturen eher ex-

trinsische. Das macht es schwer nachvollziehbar, warum die ehrliche Antwort einer chinesischen Bewerberin, dass sie hoffe, einen netten Mann kennenzulernen, oder eines indischen Bewerbers, der sich von dem neuen Job die Möglichkeit verspricht, seine Familie zu ernähren, zu einer Abwertung führt. Dies können legitime Motivationsgründe sein. Zum anderen ist es möglich, dass durch das Äußern extrinsischer Motive das Vorhandensein intrinsischer Motive angezweifelt wird (Derfler-Rozin & Pitesa, 2020). Dieser Effekt könnte im Zusammenhang mit den Interviews mit chinesischen Bewerber_innen eine Rolle gespielt haben: Durch das Fehlen angepasster Verhaltensanker, die zum Beispiel bestimmte extrinsische Motive erlauben, können bei der Beurteilung der Motivation chinesischer Bewerber_innen dann eher eigene Werte und Einstellungen der Recruiter_innen zum Tragen kommen (Derous & Ryan, 2012; Tajfel & Turner, 1986). Auffallend ist insbesondere, dass bei diesem Kriterium keine pejorative Beschreibung gefunden wurde (siehe Tabelle 1), es jedoch deskriptiv recht große numerische Unterschiede in der Beurteilung gab, auch wenn diese nicht signifikant wurden. Eine mangelnde Motivation der chinesischen und indischen Bewerber_innen ist für Außenstehende auch deswegen schwierig nachvollziehbar, weil der überwiegende Anteil schon Flugbegleiter_innen bei anderen Fluggesellschaften sind.

Bei der Frage, wie sich die trotz ähnlicher Beurteilung in den einzelnen Kriterien deutlich unterschiedliche Ablehnungsquote der chinesischen Bewerber_innen erklärt, waren sich die befragten Interviewer einig, dass das am Zusammenspiel von geringerer Belastbarkeit und teilweise fragwürdiger Motivation läge. Allerdings könnte man auch anders argumentieren: Die Ergebnisse – vor allem die hohe Quote an chinesischen Bewerber_innen mit vielen Einschränkungen in vielen Kriterien, die niedrige *adverse impact ratio*, die fehlenden, kulturell angepassten Verhaltensanker – weisen darauf hin, dass das angewendete Verfahren durchaus optimierungsfähig ist. Anscheinend fehlen klare Vorgaben zu angemessenem Verhalten, sodass sich die Recruiter_innen für ihre Beurteilungen auf ihre eigenen Wertvorstellungen berufen müssen, die jedoch geprägt von den Gesprächen mit den deutschen Bewerber_innen sind, die im Alltagsgeschäft zahlreicher sind. Eine andere, ungewohnte Motivation würde den Recruiter_innen den Spielraum geben, eine Ablehnung zu begründen (Bye et al., 2014; Derous & Ryan, 2012).

Implikationen

Für die Praxis der Personalauswahl durch Jobinterviews in kulturübergreifenden Zusammenhang bedeuten die Ergebnisse einerseits eine Bestätigung der bisherigen Be-

mühungen (wie die Trennung von Beobachtung und Beurteilung). Hier ist es wichtig, die spontane Reaktion auf fremdartiges Verhalten zunächst auch zuzulassen, um sie anschließend reflektierend auf ihre Relevanz zu überprüfen. Andererseits eröffnen die Ergebnisse auch neue Lernfelder. Die Ergebnisse zeigen, dass Personaler_innen Verhaltensanker, gerade im Kontakt- und Konfliktverhalten, klar definieren und dabei auch kulturelle Besonderheiten berücksichtigen sollten. Denn je eindeutiger die erwünschten Verhaltensweisen definiert sind, desto weniger werden von den Recruiter_innen eigene Wertvorstellungen als Maßstab angelegt. Dies gilt auch für das Kriterium Motivation: Die Erwartung an eine hohe intrinsische Motivation für die Bewerbung auf eine neue Stelle kann nicht immer erfüllt werden und muss es auch nicht (Derfler-Rozin & Pitesa, 2020; Monnot, 2018).

Eine weitere wichtige Implikation der Ergebnisse für die Praxis könnte auch das Hinterfragen der Politik einheitlicher Kriterien für alle Bewerber_innen eines Berufsbildes sein. In diesem Fall könnte die Fluggesellschaft, die einheitliche Kriterien für alle Crewmitglieder vorgibt, auch für diejenigen, die primär zu Betreuung kulturverwandter Fluggäste eingestellt werden, von dieser Regel abweichen. Zum Beispiel könnten chinesische Fluggäste eine andere Form der Betreuung erwarten als deutsche Fluggäste. Das Verhalten von chinesischen Bewerber_innen könnte in diesem Fall angemessener sein als das deutscher Bewerber_innen, die entsprechend zu gut bewertet würden. Insofern müssten nicht nur die Verhaltensanker, sondern auch die Bewertungsmaßstäbe den kulturellen Anforderungen angepasst werden. Es wäre es vermutlich auch sinnvoll, einheimische Psycholog_innen in das Interview einzubeziehen, die mehr Erfahrung in der Beurteilung spezifischer Verhaltensweisen haben. Neben der Verhaltensoperationalisierung würde auch eine größere Systematik durch einen festen Algorithmus zur Herleitung der Gesamtbeurteilung aus den einzelnen Kriterien zu mehr Verfahrensfairness führen. Unabhängig von diesen Vorschlägen wäre es wünschenswert, Recruiter_innen und auch Verantwortliche in Organisationen mehr für dieses Thema zu sensibilisieren, scheint es doch mit der unerwartet höheren Ablehnungsquote zusammenzuhängen.

Einschränkungen

Einige Einschränkungen der vorliegenden Studie sollten erwähnt werden. Zunächst ist die Zahl der untersuchten Gutachten relativ begrenzt und der Anteil fehlender Werte in den Gutachten relativ hoch, vor allem aber die Zahl der mitwirkenden Recruiter_innen recht gering. Auch die Qualifikation der Bewerber_innen aus den un-

terschiedlichen Ländern auf die vakanten Stellen ist nicht vergleichbar. Während sich in Deutschland der Bewerber_innenpool hauptsächlich aus jungen Menschen ohne jegliche Berufserfahrung und zu einem großen Teil auch ohne Berufsausbildung rekrutiert, sind die Bewerber_innen aus China und Indien oft schon erfahrene Flugbegleiter_innen, die nur zu einem neuen attraktiven Arbeitgeber wechseln möchten. Möglicherweise hätte es sonst mehr Ablehnungen gegeben.

Neben diesen die Qualität der Daten betreffenden Einschränkungen, gibt es noch eine methodische Einschränkung, die an dieser Stelle diskutiert werden soll. Die Operationalisierung des Stereotypenkonzepts in Anlehnung an die Eigenschaftenmethode von Katz und Braly (1933) hat Schwächen. Erstens werden im Vorgehen nach Katz und Braly die Eigenschaften von nur wenigen Personen zugeordnet und dadurch als Stereotyp festgelegt. Zweitens ist das Verfahren nicht unumstritten. Es wird zwar immer noch sehr häufig zur Operationalisierung von Stereotypen verwendet (vergl. Schneider & Bos, 2014), weil es einfach anzuwenden ist und auf Gruppenebene gute Ergebnisse liefert, gilt jedoch mittlerweile im Vergleich mit kognitionstheoretischen Ansätzen wie zum Beispiel der Lexical Decision Task Methode (z. B. Gaertner & McLaughlin, 1983) als zu ungenau. Die zur Verfügung gestellten Daten von Interviews aus der Praxis ließen jedoch keine Operationalisierung auf kognitionstheoretischer Basis zu. Deswegen wurde versucht, dieses Manko auszugleichen, indem nicht zufällig ausgewählte Personen die adjektivischen Beschreibungen kategorisieren, sondern Recruiter_innen, die selbst langjährige Erfahrung mit kulturübergreifenden Job Interviews haben.

Fazit

Zusammenfassend kann diese Studie wichtige Erkenntnisse zur Personalauswahlforschung beitragen. Zum einen zeigt sie, dass Recruiter_innen zwar Menschen mit entsprechenden Schemata und Stereotypen sind, aber durchaus in der Lage sind, ihre spontanen Reaktionen zu reflektieren und sich nicht in der Bewertung von diesen Reaktionen zu stark leiten zu lassen. Das scheint ihnen, zumindest bei indischen Bewerber_innen, auch ganz gut zu gelingen. Zum anderen lässt sich aus dieser Studie ableiten, dass Bemühungen um ein kulturell angepasstes Auswahlverfahren mit entsprechenden Verhaltensankern und festen Algorithmen zur Berechnung der Gesamtbeurteilung weiter betrieben werden sollten, denn die hohe Ablehnungsquote chinesischer Bewerber_innen deutet auf ein suboptimales Verfahren hin. In Zukunft wird es immer wichtiger, solche Zusammenhänge zu verstehen,

um möglichst keine geeigneten Bewerber_innen abzulehnen, nur weil man deren Interviewverhalten nicht richtig deuten kann.

Literatur

- Ashburn-Nardo, L., Voils, C. I. & Monteith, M. J. (2001). Implicit associations as the seeds of intergroup bias: How easily do they take root? *Journal of Personality and Social Psychology*, 81, 789–799. <https://doi.org/10.1037/0022-3514.81.5.789>
- Back, M. D. & Nestler, S. (2016). Accuracy of judging personality. In J. A. Hall, M. Schmid Mast & T. V. West (Eds.), *The social psychology of perceiving others accurately* (pp. 98–124). Cambridge University Press. <https://doi.org/10.1017/CBO9781316181959.005>
- Barrick, M. R., Shaffer, J. A. & DeGrassi, S. W. (2009). What you see may not be what you get: Relationships among self-presentation tactics and ratings of interview and job performance. *Journal of Applied Psychology*, 94, 1394–1411. <https://doi.org/10.1037/a0016532>
- Bhatia, S. (2017). The semantic representation of prejudice and stereotypes. *Cognition*, 164, 46–60. <https://doi.org/10.1016/j.cognition.2017.03.016>
- Birkner, K. (2004). Hegemonic struggles or transfer of knowledge? *Journal of Language and Politics*, 3, 293–322. <https://doi.org/10.1075/jlp.3.2.08bir>
- Bluhm, D. J., Harman, W., Lee, T. W. & Mitchell, T. R. (2011). Qualitative research in management: A decade of progress. *Journal of Management Studies*, 48, 1866–1891. <https://doi.org/10.1111/j.1467-6486.2010.00972.x>
- Bye, H. H., Horverak, J. G., Sandal, G. M., Sam, D. L. & van de Vijver, F. J. R. (2014). Cultural fit and ethnic background in the job interview. *International Journal of Cross Cultural Management*, 14(1), 7–26. <https://doi.org/10.1177/1470595813491237>
- Bye, H. H., Sandal, G. M., van de Vijver, F. J. R., Sam, D. L., Cakar, N. D. & Franke, G. H. (2011). Personal values and intended self-presentation during job interviews: A cross-cultural comparison. *Applied Psychology: An International Review*, 60, 160–182. <https://doi.org/10.1111/j.1464-0597.2010.00432.x>
- Byrne, D. (1997). An overview (and underview) of research and theory within the attraction paradigm. *Journal of Social and Personal Relationships*, 14, 417–431. <https://doi.org/10.1177/0265407597143008>
- Corbin, J. & Strauss, A. L. (2008). *Basics of qualitative research* (3rd ed.). Thousand Oaks, CA: Sage. <https://doi.org/10.4135/9781452230153>
- DeGroot, T. & Gooty, J. (2009). Can nonverbal cues be used to make meaningful personality attributions in employment interviews? *Journal of Business and Psychology*, 24, 179–192. <https://doi.org/10.1007/s10869-009-9098-0>
- Derfler-Rozin, R. & Pitesa, M. (2020). Motivation purity bias: Expression of extrinsic motivation undermines perceived intrinsic motivation and engenders bias in selection decisions. *Academy of Management Journal*, 63, 1840–1864. <https://doi.org/10.5465/amj.2017.0617>
- Derous, E. (2017). Ethnic minorities' impression management in the interview: Helping or hindering? *Frontiers in Psychology*, 8, 86. <https://doi.org/10.3389/fpsyg.2017.00086>
- Derous, E. & Ryan, A. M. (2012). Documenting the adverse impact of résumé screening: Degree of ethnic identification matters. *International Journal of Selection and Assessment*, 20, 464–474. <https://doi.org/10.1111/ijsa.12009>

- Fell, C. B. & König, C. J. (2016). Cross-cultural differences in applicant faking on personality tests: A 43-nation study. *Applied Psychology: An International Review*, 65, 671–717. <https://doi.org/10.1111/apps.12078>
- Friese, S. (2018). *ATLAS.ti 8 Windows – User Manual – updated for program version 8.3*. Berlin: Scientific Software Development.
- Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science*, 21, 177–182. <https://doi.org/10.1177/0963721412445309>
- Gaertner, S. L. & McLaughlin, J. P. (1983). Racial stereotypes: Associations and ascriptions of positive and negative characteristics. *Social Psychology Quarterly*, 46(1), 23. <https://doi.org/10.2307/3033657>
- Horverak, J. G., Bye, H. H., Sandal, G. M. & Pallesen, S. (2012). Managers' evaluations of immigrant job applicants: The influence of acculturation strategy on perceived person-organization fit (P-O fit) and hiring outcome. *Journal of Cross-Cultural Psychology*, 44, 46–60. <https://doi.org/10.1177/002202211430256>
- Horverak, J. G., Sandal, G. M., Pallesen, S. & Timmerman, M. E. (2013). Hiring rankings of immigrant job applicants: Immigrants' acculturation strategies and managers' personality trait perception. *International Migration and Integration*, 14, 493–510. <https://doi.org/10.1007/s12134-012-0247-3>
- Howard, J. L. & Ferris, G. R. (1996). The employment interview context: Social and situational influences on interviewer decisions. *Journal of Applied Social Psychology*, 26, 112–136. <https://doi.org/10.1111/j.1559-1816.1996.tb01841.x>
- Huffcutt, A. I. (2011). An empirical review of the employment interview construct literature. *International Journal of Selection and Assessment*, 19, 62–81. <https://doi.org/10.1111/j.1468-2389.2010.00535.x>
- Imada, A. S. & Hakel, M. D. (1977). Influence of nonverbal communication and rater proximity on impressions and decisions in simulated employment interviews. *Journal of Applied Psychology*, 62, 295–300. <https://doi.org/10.1037/0021-9010.62.3.295>
- Jansen, A., König, C. J., Stadelmann, E. H. & Kleinmann, M. (2012). Applicants' self-presentational behavior: What do recruiters expect and what do they get? *Journal of Personnel Psychology*, 11, 77–85. <https://doi.org/10.1027/1866-5888/a000046>
- Katz, D. & Braly, K. (1933). Racial stereotypes of one hundred college students. *Journal of Abnormal and Social Psychology*, 28, 280–290. <https://doi.org/10.1037/h0074049>
- Kholin, M. & Blicke, G. (2015). Zum Verhältnis von Erwerbsarbeit, Arbeitswerten und Globalisierung. *Zeitschrift für Arbeits- und Organisationspsychologie*, 59, 16–29. <https://doi.org/10.1026/0932-4089/a000169>
- Klehe, U.-C. & Latham, G. (2006). What would you go – really or ideally? Constructs under-lying the behavior description interview and the situational interview in predicting typical versus maximum performance. *Human Performance*, 19, 357–382. https://doi.org/10.1207/s15327043hup1904_3
- Krzyzaniak, S. L., Colman, D. E., Letzring, T. D., McDonald, J. S. & Biesanz, J. C. (2019). The effect of information quantity on distinctive accuracy and normativity of personality trait judgments. *European Journal of Personality*, 33, 197–213. <https://doi.org/10.1002/per.2196>
- Levashina, J. & Campion, M. A. (2006). A model of faking likelihood in the employment interview. *International Journal of Selection and Assessment*, 14, 299–316. <https://doi.org/10.1111/j.1468-2389.2006.00353.x>
- Levashina, J., Hartwell, C. J., Morgeson, F. P. & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology*, 67, 241–293. <https://doi.org/10.1111/peps.12052>
- Linville, P. W., Salovey, P. & Fischer, G. W. (1986). Stereotyping and perceived distributions of social characteristics: An application to ingroup-outgroup perception. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 165–208). Orlando, FL: Academic Press.
- Lippmann, W. (1922). *Public opinion*. London: Allen & Unwin.
- Lipsmeyer, C. S. & Zhu, L. (2011). Immigration, globalization, and unemployment benefits in developed EU states. *American Journal of Political Science*, 55, 647–664. <https://doi.org/10.1111/j.1540-5907.2011.00509.x>
- Macan, T. (2009). The employment interview: A review of current studies and directions for future research. *Human Resource Management Review*, 19, 203–218. <https://doi.org/10.1016/j.hrmr.2009.03.006>
- Manroop, L., Boekhorst, J. A. & Harrison, J. A. (2013). The influence of cross-cultural differences on job interview selection decisions. *International Journal of Human Resource Management*, 24, 3512–3533. <https://doi.org/10.1080/09585192.2013.777675>
- Monnot, M. J. (2018). The effect of incentives on intrinsic motivation and employee attitudes: A multilevel study across nations and cultural clusters. *Thunderbird International Business Review*, 60, 675–689. <https://doi.org/10.1002/tie.21949>
- Neuliep, J. W. (2014). *Intercultural communication: A contextual approach* (6th ed.) Thousand Oaks, CA: Sage.
- Outtz, J. L. & Newman, D. A. (2010). A theory of adverse impact. In J. L. Outtz (Eds.), *Adverse impact: Implications for organizational staffing and high stakes selection* (pp. 53–94). New York: Routledge.
- Parsons, C. K. & Liden, R. C. (1984). Interviewer perceptions of applicant qualifications: A multivariate field study of demographic characteristics and nonverbal cues. *Journal of Applied Psychology*, 69, 557–568. <https://doi.org/10.1037/0021-9010.69.4.557>
- Pettigrew, T. F. (2016). In pursuit of three theories: Authoritarianism, relative deprivation, and intergroup contact. *Annual Review of Psychology*, 67, 1–21. <https://doi.org/10.1146/annurev-psych-122414-033327>
- Phills, C. E., Hahn, A. & Gawronski, B. (2020). The bidirectional causal relation between implicit stereotypes and implicit prejudice. *Personality and Social Psychology Bulletin*, 46, 1318–1330. <https://doi.org/10.1177/0146167219899234>
- Reeder, G. D., Vonk, R., Ronk, M. J., Ham, J. & Lawrence, M. (2004). Dispositional attribution: Multiple inferences about motive-related traits. *Journal of Personality and Social Psychology*, 86, 530–544. <https://doi.org/10.1037/0022-3514.86.4.530>
- Roulin, N., Bangerter, A. & Levashina, J. (2014). Interviewers' perceptions of impression management in employment interviews. *Journal of Managerial Psychology*, 29, 141–163. <https://doi.org/10.1108/JMP-10-2012-0295>
- Sackett, P. R., Zhang, C., Berry, C. M. & Lievens, F. (2021). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *Journal of Applied Psychology*. Advance online publication. <https://doi.org/10.1037/apl0000994>
- Schmitt, D. P., Allik, J., McCrae, R. R. & Benet-Martínez, V. (2007). The geographic distribution of Big Five personality traits: Patterns and profiles of human self-description across 56 nations. *Journal of Cross-Cultural Psychology*, 38, 173–212. <https://doi.org/10.1177/0022022106297299>
- Schneider, M. C. & Bos, A. L. (2014). Measuring stereotypes of female politicians. *Political Psychology*, 35, 245–266. <https://doi.org/10.1111/pops.12040>
- Schuler, H. (1992). Das Multimodale Einstellungsinterview. *Diagnostica*, 38, 281–300.
- Stroebe, W. & Insko, C. A. (1989). Stereotype, prejudice, and discrimination: Changing conceptions in theory and research. In D.

- Bar-Tal, C. F. Graumann & A. W. Kruglanski (Eds.), *Stereotyping and prejudice: Changing conceptions* (pp. 3–34). New York: Springer. https://doi.org/10.1007/978-1-4612-3582-8_1
- Swider, B. W., Barrick, M. R. & Harris, T. B. (2016). Initial impressions: What they are, what they are not, and how they influence structured interview outcomes. *Journal of Applied Psychology, 101*, 625–638. <https://doi.org/10.1037/apl0000077>
- Tajfel, H. & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations* (pp. 33–47). Chicago: Hall Publishers.
- Westhoff, K., Hagemeister, C., Kersting, M., Lang, F., Moosbrugger, H., Reimann, G. & Stemmler, G. (Hrsg.). (2010). *Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33430* (3., überarb. Aufl.). Lengerich: Pabst Science Publishers.
- Westhoff, K. & Strobel, A. (2011). Interview. In L. F. Hornke, M. Amelang & M. Kersting (Hrsg.), *Enzyklopädie der Psychologie: Themenbereich B, Methodologie und Methoden, Serie II, Psychologische Diagnostik* (S. 371–413). Göttingen: Hogrefe.
- Wilhelmy, A., Kleinmann, M., König, C. J., Melchers, K. G. & Truxillo, D. M. (2016). How and why do interviewers try to make impressions on applicants? A qualitative study. *Journal of Applied Psychology, 101*, 313–332. <https://doi.org/10.1037/apl0000046>
- Wirtz, M. A. & Caspar, F. (2007). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe.

Historie

Eingegangen: 30.01.2021
 Revision eingegangen: 23.07.2022
 Onlineveröffentlichung: 16.09.2022

Danksagung

Wir danken Gerhard Blickle für seine hilfreichen Rückmeldung zu einer früheren Fassung dieses Manuskripts.

Förderung

Open Access-Veröffentlichung ermöglicht durch die Saarländische Universitäts- und Landesbibliothek.

ORCID

Cornelius J. König
 <https://orcid.org/0000-0003-0477-8293>

Gamal Eid

AE Arbeits- und Organisationspsychologie
 Universität des Saarlandes
 Campus A1 3
 66123 Saarbrücken
 Deutschland
gamal.eid@uni-saarland.de