

Validität von Assessment-Centern in deutschen Unternehmen

Vergleich von Unternehmensdaten mit einer bestehenden Metaanalyse

Im Jahr 1998 erschien eine umfassende Studie, in der die Validitäten von insgesamt 19 verschiedenen Auswahlverfahren metaanalytisch untersucht wurden. Für Assessment-Center ergab sich nur eine mittlere Validität, aktuellere Folgestudien konstatieren für AC-Verfahren sogar eine abnehmende Validität. Der Validitätsverlust hat nicht nur die Wissenschaft, sondern auch die Praxis irritiert. Vor diesem Hintergrund prüfen die Autoren die Validitätsgeneralisierung von AC aus deutschen Unternehmen.



Wechselsprung über Leitkegel: Ein neuer Eignungstest für Berufsfeuerwehren berücksichtigt die körperlich unterschiedlichen Voraussetzungen von Frauen und Männern.

Um zuverlässige Personalauswahlentscheidungen fällen zu können, sind treffsichere Auswahlverfahren notwendig. Die Vorhersagegüte eines Verfahrens wird grundsätzlich durch einen Validitätskoeffizienten abgebildet, der den statistischen Zusammenhang zwischen der Kandidatenleistung im Auswahlverfahren (Prädiktor) und einem späteren beruflichen Erfolgsindikator (Kriterium) beschreibt. Der Korrelationskoeffizient kann dabei zwischen 0.0 (kein Zusammenhang) und 1.0 (perfekter Zusammenhang) variieren, wobei im Bereich der Personaldiagnostik Korrelationen im Bereich um die $r = .50$ als hervorragend gelten (Kersting 2008, 73). Demnach sollten Kandidaten, die im Auswahlverfahren besser abgeschnitten haben, auch später erfolgreicher im Beruf agieren. Kann ein solcher Zusammenhang nicht nachgewiesen werden, hätte das Auswahlverfahren seinen ursprünglichen Zweck verfehlt und würde als nicht prognostisch valide eingestuft werden.

DIE AUTOREN

NILS BENIT ▶ Master of Science, Psychologie, wissenschaftlicher Mitarbeiter und Promovend am Institut für Psychologie der Universität Hildesheim



PROF. DR. PHIL. RENATE SOELLNER ▶ Lehrstuhl für Forschungsmethoden und Evaluation, Institut für Psychologie, Universität Hildesheim



DAS ASSESSMENT-CENTER

Ein international weitverbreitetes Auswahlverfahren, das sich seit längerer Zeit auch im deutschsprachigen Raum steigender Beliebtheit erfreut (vgl. Höft / Obermann 2010), ist das Assessment-Center (AC). Obwohl sich die Durchführung von AC ressourcenintensiv gestaltet, scheint vor allem die multimethodale und augenscheinliche Ausrichtung dieses Verfahrens die Anwender zu überzeugen. Die eingesetzten Übungen (z. B. Rollenspiel, Gruppendiskussion) weisen einen erkennbaren Bezug zu den Anforderungen des zu besetzenden Arbeitsplatzes auf, was sich einerseits positiv auf die Akzeptanz des Verfahrens aus Arbeitgeber- und Arbeitnehmerseite auswirkt und andererseits eine hohe Erwartung in die Validität setzt (Schuler

et al. 2007). Die Argumentation für oder gegen ein Auswahlverfahren ist also stark von dessen prognostischer Vorhersagegüte für spätere berufliche Erfolgskriterien geleitet.

INTERNATIONALE FORSCHUNG

Eine der ersten und berühmtesten Studien zur Validität eines AC wurde in der American Telephone and Telegraph Company (AT & T) durchgeführt (Bray / Grant 1966). In dieser Einzelstudie ergab sich ein zufriedenstellender Validitätskoeffizient von $r = .46$. Erst in den 1980er-Jahren legten Thornton et al. (1987) die erste Metaanalyse zur Validität von AC vor. Diese Auswertungstechnik integriert empirische Befunde mehrerer Einzelstudien aus der Literatur und ermöglicht so eine repräsentative und generalisierbare Schätzung der Validität (zum typischen Ablauf einer Metaanalyse vgl. Abb. 1). Die Autoren berücksichtigten insgesamt 50 Validierungsstudien und analysierten eine mittlere korrigierte Validität von $\rho = .37$.

In der umfassenden Überblicksarbeit von Schmidt und Hunter (1998) wurden die Validitäten von 19 verschiedenen Auswahlverfahren metaanalytisch untersucht. Für das AC ergab sich erneut eine mittlere korrigierte Validität von $\rho = .37$. Aktuelle Folgestudien konstatieren der AC-Methode eine abnehmende Validität. So berichten Hardison und Sackett (2007) in ihrer Metaanalyse eine korrigierte mittlere Vorhersagekraft von $\rho = .26$. Hermelin et al. (2007) ermittelten auf Basis von 27 Einzelbefunden eine vergleichbar niedrige Validität von $\rho = .28$ (vgl. Holzenkamp et al. 2010). Der Validitätsverlust über die Zeit hinweg beunruhigt.

NATIONALE FORSCHUNG

Die skizzierten Befunde aus dem englischsprachigen Raum werfen zwei bedeutende Forschungsfragen auf, denen sich

Holzenkamp et al. (2010) widmeten. Die Autoren untersuchten, inwieweit die primär anglo-amerikanischen Befunde auf den deutschsprachigen Kulturkreis übertragen werden können und ob ein ähnlich negativer Trend in der Validitätsentwicklung auch hier zu beobachten ist (Holzenkamp et al. 2010, 19). Ihre Metaanalyse basierte auf 24 Koeffizienten aus 19 verschiedenen Einzelstudien aus dem deutschsprachigen Raum, die im Zeitraum von 1987 bis 2007 durchgeführt wurden. Die Stichprobengrößen der Einzelstudien variierten von 29 bis 546 und umfassten insgesamt 3 556 AC-Kandidaten.

Die in dieser Analyse berücksichtigten prognostischen Validitätskoeffizienten zeigten ebenfalls eine hohe Variabilität (-.07 bis .51). Die mittlere korrigierte Validität der untersuchten Verfahren betrug im Ergebnis $\rho = .36$. Deutschsprachige AC erwiesen sich damit als valide und aussagekräftige Verfahren (Holzenkamp et al. 2010). Es liegen bislang keine weiteren Metaanalysen aus dem deutschsprachigen Raum vor, die Rückschlüsse auf eine Validitätszunahme oder -abnahme für den deutschsprachigen Raum zulassen.

Ziel der vorliegenden Arbeit ist es, einen weiteren Beitrag zur Validitätsgeneralisierung von AC im deutschsprachigen Raum zu leisten. Hervorzuheben ist, dass in dieser Studie ausschließlich AC zur Auswahl von Führungskräften untersucht wurden. Im Vergleich zu den genannten Metaanalysen basieren außerdem die metaanalytischen Befunde dieser Studie auf Selbstauskünften deutscher Unternehmen, die im Rahmen einer Befragung erhoben wurden. Ein weiteres Ziel besteht somit in der Prüfung, ob die unterschiedlichen Datenerhebungsmethoden aus der vorliegenden Untersuchung (Selbstauskunft) und aus der Studie von Holzenkamp et al. (2010) (Literaturrecherche) bedeutsame Validitätsunterschiede hervorbringen.

METHODE UND VORGEHENSWEISE

Datengewinnung Im Rahmen einer im Herbst 2010 durchgeführten Studie zum Thema „Vorauswahl vor dem AC“ wurde an 250 deutsche Unternehmen postalisch ein Fragebogen geschickt.

Es wurden zuvor über 500 deutsche Unternehmen recherchiert, die AC für Führungskräfte durchführen und mindestens 150 Mitarbeiter beschäftigen. Jedes iden-

tifizierte Unternehmen wurde telefonisch kontaktiert, über die geplante Studie informiert und zur Bereitschaft an einer Teilnahme befragt. Von den 250 an der Studie interessierten Unternehmen schickten 109 einen ausgefüllten Fragebogen zurück. Neben der Erfassung verschiedener Vorauswahlverfahren wurden die befragten Personalverantwortlichen gebeten, falls ihr AC bereits evaluiert wurde, den Validitätskoeffizienten anzugeben.

ABLAUF EINER METAANALYSE		
Vorbereitungsphase	Hauptphase	Abschlussphase
<ul style="list-style-type: none"> ▶ Formulierung der Forschungsfrage mit Festlegung der Prädiktor- / Kriteriumsvariablen ▶ Literaturrecherche nach vorher festgelegten Qualitätskriterien ▶ Kodierung der Studien nach vorher festgelegten Kriterien 	<ul style="list-style-type: none"> ▶ Artefaktkorrektur der Ausgangsdaten ▶ Schätzung der generalisierbaren Effektgröße ▶ Suche und Analyse von Moderatorvariablen 	<ul style="list-style-type: none"> ▶ Ergebnispräsentation ▶ Ergebnisinterpretation ▶ Implikationen für Praxis und Forschung

Abb. 1

tifizierte Unternehmen wurde telefonisch kontaktiert, über die geplante Studie informiert und zur Bereitschaft an einer Teilnahme befragt. Von den 250 an der Studie interessierten Unternehmen schickten 109 einen ausgefüllten Fragebogen zurück. Neben der Erfassung verschiedener Vorauswahlverfahren wurden die befragten Personalverantwortlichen gebeten, falls ihr AC bereits evaluiert wurde, den Validitätskoeffizienten anzugeben.

Berichtete Validitätskennziffern, die nicht als Korrelationskoeffizienten vorlagen (z. B. Prozentangaben) oder als unrealistisch einzustufen waren (z. B. $r > 1$) wurden nicht in die Analyse eingeschlossen. Durchgängig wurde die berufliche Leistung als Außenkriterium angegeben.

Analyse der generalisierten Validität mit Artefaktkorrektur Zur Berechnung wurde das metaanalytische Verfahren von Hunter und Schmidt (2004) verwendet, das trotz partiell fehlender Artefaktinfor-

mationen (z. B. Varianzeinschränkung, Prädiktor-, Kriteriumsreliabilität, Stichprobenfehler) eine Analyse zur Bestimmung der Validitätsgeneralisierung erlaubt. Die berichteten Koeffizienten wurden anhand der Stichprobengröße und der Kriteriumsreliabilität korrigiert. Die Korrektur der Stichprobenfehler bewirkt, dass auf großen Stichproben basierende Koeffizienten stärker gewichtet werden. Die mangelnde Kriteriumsreliabilität (z. B. Vorgesetzter schätzt die Mitarbeiterleistung nicht perfekt ein) wurde zwecks Vergleichbarkeit nach dem von Holzenkamp et al. (2010) berichteten korrigierten und gemittelten Koeffizienten für das Kriterium „Indikatoren für Berufserfolg“ mit .83 angesetzt. Auf eine Korrektur der Prädiktorreliabilität wurde verzichtet, da wie in der Studie von Holzenkamp und Kollegen „die prädiktive Validität real existierender und nicht statistisch aufgewerteter ACs“ von besonderem Forschungsinteresse sind (Holzenkamp et al., 2010, 20). Speziell im Bereich der Personalauswahl sollte ebenso die Varianzeinschränkung korrigiert werden, da grundsätzlich nur von den eingestellten Bewerbern Prädiktor- und Kriteriumswerte vorliegen. Einige Unternehmen gaben auch die Besteherquote für ihr AC an. Da jedoch unklar war, ob die berichteten Validitätskoeffizienten bereits hinsichtlich dieser Varianzeinschränkung korrigiert worden sind, wurde keine weitere Korrektur vorgenommen.

Zur Überprüfung der Validitätsgeneralisierung wurden neben den Vertrauens-

intervallen, die die Variabilität der korrigierten Validitäten beschreiben, auch der Kreditabilitätswert betrachtet, welcher überprüft, ob 90 Prozent der individuellen Studien von null verschieden sind und dementsprechend auf eine positive sowie generalisierbare Validität hinweist (Hunter / Schmidt 2004).

Analyse der Moderatorvariablen Im Rahmen anschließender Moderatoranalysen wird untersucht, ob nicht auf diese Artefakte zurückzuführende Unterschiede (Restvarianz) zwischen den Einzelstudien existieren. Zur Überprüfung der Restvarianz wurde auf die 75-Prozent-Regel nach Hunter und Schmidt (2004) zurückgegriffen. Mittels Methode der Fraktionierung werden potenzielle Moderatorvariablen in zwei Substichproben aufgeteilt und getrennt voneinander metaanalytisch untersucht. Sollten sich die mittleren

valide erweisen. Da die Datengewinnung der vorliegenden Studie im Rahmen einer Unternehmensbefragung zum Thema „Vorauswahl vor dem AC“ stattfand, liegen nur wenige AC-spezifische Informationen vor, sodass lediglich die Variable DIN-33430-Zertifizierung als Moderator geprüft wurde.

ERGEBNISSE

Insgesamt wurden von den befragten Personalverantwortlichen 23 verwertbare und unabhängige Validitätskoeffizienten berichtet, die von $r = .18$ bis $r = .68$ reichten. Die Gesamtstichprobe umfasste 3 977 AC-Kandidaten, wobei die angegebenen Stichprobengrößen der jeweiligen AC beträchtlich variierte ($N = 30$ bis $N = 700$). Der hinsichtlich der Stichprobengröße gemittelte Va-

lidityskoeffizient beträgt $r = .38$. Nach zusätzlicher Korrektur der Kriteriums-unreliabilität resultiert eine prognostische Validität von $\rho = .42$. Das Konfidenzintervall erstreckt sich von .39 bis .45. Der 90-Prozent-Kreditabilitätswert lag mit .31 deutlich über null, sodass von einer verfahrens-unabhängigen Validitätsgeneralisierung gesprochen werden kann. In der vorliegenden Untersuchung lassen sich 39 Prozent der Gesamtvarianz auf die korrigierten Artefakte zurückführen, sodass gemäß der 75-Prozent-Regel eine Analyse potenzieller Moderatorvariablen sinnvoll erscheint. Tabelle 1 nennt die generellen Befunde der Metaanalyse und der Moderatoranalyse.

METAANALYTISCHE AUSWERTUNG UND MODERATORANALYSE				
Analyse		DIN-33430-Zertifizierung		
		gesamt	ja	nein
	k (Anzahl der verwendeten Koeffizienten)	23	9	14
	N (Stichprobengröße)	3977	1925	2052
	r (mittlere Reliabilität der Kriteriumsmaße)	.83	.83	.83
Korrigiert bzgl. Stichprobengröße	r	.38	.42	.34
	VAR (Restvarianz der Koeffizienten)	.006	.010	.001
	SD (Standardabweichung der Koeffizienten)	.08	.10	.01
Zusätzlich korrigiert bzgl. Kriteriumsreliabilität	ρ (nach Artefakten mittlere korrigierte Validität)	.42	.46	.37
	VAR	.01	.012	.001
	SD	.09	.11	.01
Kreditabilitätsintervall (CR) 80 %	LB (untere Grenze)	.31	.32	.36
	UB (obere Grenze)	.53	.60	.39
Konfidenzintervall (CI) 95 %	LB	.39	.42	.33
	UB	.44	.50	.41

Tab. 1

Validitäten der beiden Gruppen unterscheiden und die Konfidenzintervalle nicht überlappen, kann von einer moderierenden Wirkung ausgegangen werden.

In der Metaanalyse von Holzenkamp und Kollegen wurden zahlreiche Variablen als potenzielle Moderatoren geprüft (z. B. Anzahl der eingesetzten Instrumente, Dauer des AC, Alter der Teilnehmer. So lässt sich zudem zur generalisierten Validität prüfen, welche AC-Varianten oder Rahmenbedingen sich als besonders

Der Zertifizierungsprozess nach DIN 33430 erwies sich als bedeutsamer Moderator in dem Sinne, dass Verfahren aus bereits zertifizierten oder sich noch im Zertifizierungsprozess befindenden Unternehmen eine höhere Validität aufweisen als Verfahren aus Unternehmen, die sich noch nicht im Zertifizierungsprozess befinden ($\rho = .47$ versus $\rho = .38$). Es liegt keine Überschneidung der Konfidenzintervalle für die beiden Subgruppen vor, sodass von statistisch bedeutsamen Unterschieden, die nicht durch Artefakte aufgeklärt werden können, ausgegangen werden kann.

DISKUSSION

Der vorliegende Beitrag untersucht metaanalytisch auf Basis gesammelter Unternehmensdaten, ob die steigende Beliebtheit und Verwendungshäufigkeit von AC in Deutschland auch unter dem Aspekt der prognostischen Validität gerechtfertigt erscheinen. Unsere Ergebnisse weisen auf eine zufriedenstellende Vorhersagegüte von AC im deutschsprachigen Raum hin. Die nach Stichprobenfehler und Kriteriumsunsicherheit korrigierte Validität ($\rho = .42$) steht im Einklang mit den Befunden von Holzenkamp und Kollegen aus dem deutschsprachigen Raum ($\rho = .36$). Auch die fast identische Anzahl der eingeflossenen Koeffizienten ($k = 23$ versus $k = 24$) sowie die ähnlich großen Stichprobenumfänge ($N = 3977$ versus $N = 3556$) der beiden Studien begünstigen diese Schlussfolgerung. Die in

dieser Metaanalyse leicht höhere korrigierte Validität kann dabei entlang von vier Argumentationslinien erklärt werden.

1. Unsere Daten basieren auf AC für Führungskräfte. Es erscheint plausibel, dass von Unternehmen für die Eignungsdiagnostik höherer Positionen ein größerer Etat zur Verfügung gestellt wird und sich dieses in der Vorhersagegüte niederschlägt, denn professionelle Beobachterschulungen oder die Konzeption von AC-Übungen auf Basis einer ausführlichen Anforderungsanalyse sind zeit- und somit kostenintensiv. Sie führen aber im Resultat zu einer hochwertigeren Eignungsdiagnostik.
2. Von den befragten Personalverantwortlichen wurde ausschließlich die berufliche Leistung als Kriterium angegeben. Auch in der Metaanalyse von Holzenkamp et al. (2010) unterschieden sich die Validitäten in Abhängigkeit von der Art des Kriteriums (Indikatoren für Berufserfolg $\rho = .41$; objektive Leistungsindikatoren $\rho = .04$).
3. Aus den Unternehmensangaben war nicht erkennbar, ob die Validitätskoeffizienten bereits hinsichtlich Varianzeinschränkungen, Prädiktor- und / oder Kriteriumsunsicherheit korrigiert wurden. Hierdurch kann womöglich die korrigierte Validität überschätzt worden sein.
4. Es ist davon auszugehen, dass in unserer Studie relativ aktuell erhobene Validierungsdaten enthalten sind, da die Angaben der Personalverantwortlichen sich auf ihr aktuelles AC beziehen sollten. In jüngerer Vergangenheit ist eine zunehmende Vernetzung von Spezialisten aus Wissenschaft und Praxis (z. B.

Anzeige

Anzeige 10
Ge.on
1/2 Seite / 4c

Arbeitskreis Assessment Center e.V. oder Deutsche Gesellschaft für Personalführung e.V.) zu beobachten, die sich in einer fortschreitenden Professionalisierung und Qualitätserhöhung der AC-Methode niederschlagen sollte. Insbesondere die im Jahre 2002 verabschiedete DIN 33430 hat zu einer Standardisierung und Qualitätssicherung eignungsdiagnostischer Verfahren beigetragen.

Dieses Argument wird durch den Befund der Moderatoranalyse gestützt. Für im DIN-33430-Prozess sich befindende Verfahren konnte eine höhere Validität ($\rho = .46$) gefunden werden. Für AC aus Unternehmen, die sich noch nicht im Zertifizierungsprozess befinden ($\rho = .37$), ähneln unsere Validitäten stark denen von Holzenkamp und Kollegen ($\rho = .36$). Dieser Moderatorenbefund war erwartbar, da die DIN 33430 eine Mindesthöhe der prognostischen Validität für eine erfolgreiche Zertifizierung voraussetzt. Auch wenn dieser Befund auf den ersten Blick zirkulär wirkt, unterstreicht er eindrucksvoll den Einfluss einer nach wissenschaftlichen Empfehlungen und Kriterien durchgeführten Eignungsdiagnostik und verdeutlicht, dass die Durchführung von AC keine „Spielwiese für Laien“ sein sollte (vgl. Schuler 2007).

Unsere Ergebnisse unterstreichen die Notwendigkeit länderspezifischer Metaanalysen. Unterschiedliche Rahmenbedingungen (z. B. organisatorisches Umfeld oder das Verständnis von Personalauswahl) machen eine



Eignungstest für Studienanfänger im Fach Soziologie an der Ludwig-Maximilians-Universität München.

Generalisierung auf unterschiedliche (Unternehmens-) Kulturkreise fast unmöglich.

Abschließend stellt sich allerdings noch die Frage, welche Auswirkung die Erhebungsmethode auf die Ergebnisse hat. Eine literaturbasierte Metaanalyse überzeugt aufgrund der Berücksichtigung von vorher festgelegten Kriterien der Primärstudien, da verfahrens- sowie auswertungsbezogene Hintergrundinformationen den entsprechenden Originalbeiträgen entnehmbar sind. Eine auf Selbstauskünften basierende metaanalytische Befundintegration scheint insbesondere dann geeignet zu sein, wenn ein möglichst aktueller Stand in den Unternehmen abgebildet werden soll. Ob die Erhebungsmethode tatsächlich einen Bias hinsichtlich der Validitätsgeneralisierung bewirkt, kann abschließend noch nicht beantwortet werden. Die Befunde von Holzenkamp et al. (2010) sowie die Ergebnisse der vorliegenden Studie weisen aber gleichermaßen darauf hin, dass AC aus dem deutschsprachigen Raum aussagekräftige Verfahren darstellen. ●

Literatur

- Bray, D. W. / Grant, D. L. (1966): The assessment center in the measurement of potential for business management. *Psychological Monographs: General and Applied*, 80, 1–7
- Holzenkamp, M. et al. (2010): Wie valide sind Assessment Center im deutschsprachigen Raum? Eine Überblicksstudie mit Empfehlungen für die AC-Praxis, in: *Wirtschaftspsychologie aktuell*, 2, 17–25
- Höft, S. / Obermann, C. (2010): Praxiseinsatz von Assessment Centern im deutschsprachigen Raum: Eine zeitliche Verlaufsanalyse basierend auf den Anwenderbefragungen des Arbeitskreises Assessment Center e.V. von 2001 und 2008, in: *Wirtschaftspsychologie aktuell*, 2, 5–16
- Hunter, J. E. / Schmidt, F. L. (2004): *Methods of meta-analysis*, 2nd ed., Thousand Oaks, CA
- Kersting, M. (2008): Zur Zukunft von Assessment-Center-Verfahren. Möglichkeiten der Qualitätssicherung und -optimierung, in: *Personalführung*, 41 (12), 72–80
- 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, in: *Psychological Bulletin*, 124 (2), 262–274
- Schuler, H. (2007): Spielweise für Laien? Weshalb das Assessment-Center seinem Ruf nicht mehr gerecht wird, in: *Wirtschaftspsychologie aktuell*, 2, 27–30
- Schuler, H. et al. (2007): Die Nutzung psychologischer Verfahren der externen Personalauswahl in deutschen Unternehmen. Ein Vergleich über 20 Jahre, in: *Zeitschrift für Personalpsychologie*, 6 (2), 60–70