

Spielwiese für Laien? Weshalb das Assessment-Center seinem Ruf nicht mehr gerecht wird

Die Verwendung des Assessment-Centers als eignungsdiagnostischer Methode hat in den letzten Jahrzehnten zugenommen. Doch gleichzeitig hat dieses Verfahren an psychometrischer Qualität verloren. Die ließe sich jedoch durch eine multimodale Gestaltung wieder verbessern.

Das Assessment-Center (AC) gehört heute zu den populärsten eignungsdiagnostischen Verfahren. Es wird zur Auswahl externer Bewerber wie zur Potenzialdiagnose bei unternehmensinternen Nachwuchskräften eingesetzt, es liefert die Entscheidungsgrundlage für Karrierewege und für Maßnahmen der Personalentwicklung. Überdies wird ihm Bedeutung für die Unternehmenskultur zugesprochen. Zum Beispiel durch die Förderung der Bereitschaft von Führungskräften, eine Mentorenrolle zu übernehmen, sowie zur Steigerung der Führungsfähigkeit, da Vorgesetzte durch ihre Rolle als Beobachter angehalten werden, auf erfolgsrelevantes Verhalten zu achten. AC-Ratgeberbücher für Praktiker und vor allem die für Stellenbewerber erzielten hohen Auflagen.

Trotz des großen Aufwands finden ACs heute wesentlich häufiger Anwendung als in den vergangenen beiden Dekaden. Abbildung 1 zeigt die Verwendungshäufigkeit in deutschen Unternehmen in den letzten Jahren. Die Daten lassen erkennen, dass für die Beschäftigtengruppen Auszubildende, Trainees und Führungskräfte ACs häufiger eingesetzt werden, und zwar in besonders ausgeprägtem Maße für die Gruppen der Nicht-Führungskräfte. Nennenswerte Zuwachsraten konnten ansonsten nur noch strukturierte Interviews erzielen, wohingegen Persönlichkeits- und Intelligenztests heute seltener eingesetzt werden als vor einem und vor zwei Jahrzehnten. Der wachsenden Einsatzhäufigkeit des ACs entspricht auch die positivere – wenngleich noch immer kritische – Praktikabilitätseinschätzung der Verwender im Vergleich zur vorgängigen Befragung, während Validitäts- und Akzeptanzeinschätzung nahezu unverändert hoch ausfallen. Dementsprechend planen etwa 20 Prozent der zum aktuellen Zeitpunkt befragten Unternehmen (Schuler, Hell, Trapmann, Schaar & Boramir, 2007), künftig ACs erstmalig einzuführen.

Assessment-Center haben also an Ansehen und Verbreitung gewonnen. Doch wie steht es mit ihrer

psychometrischen Qualität? Die erste große publizierte Validierungsstudie zum AC war die Untersuchung bei AT&T in den 50er-Jahren (Bray & Grant, 1966). Sie resultierte in einem unkorrigierten (!) Validitätskoeffizienten von $r = .46$ sowohl für Bewerber mit als auch für solche ohne College-Abschluss (Kriterium war das Erreichen einer Führungsposition innerhalb von acht Jahren). Der methodische Clou der Studie war das Geheimhalten der Diagnoseergebnisse, sodass der häufig erhobene Selffulfilling-Prophecy-Vorbehalt gerade bei dieser Untersuchung gegenstandslos ist. Nach vormaliger Anwendung im militärischen und nachrichtendienstlichen Bereich gewannen ACs in den darauffolgenden Jahrzehnten auch in zivilen Unternehmen rasch an Verbreitung, zumindest in den USA. Dadurch konnten Thornton, Gaugler, Rosenthal und Bentson (1987) eine metaanalytische Auswertung von 50 Einzelstudien mit 107 unabhängigen Validitätskoeffizienten vornehmen. Die Koeffizienten zeigten die erstaunliche Streubreite von $r = -.25$ bis $r = .78$. und wiesen daher schon darauf hin, dass neben den Stichprobenfehlern auch die methodische Qualität der Einzelstudien sehr unterschiedlich sein dürfte.

Enttäuschend: deutlich verringerte Validität im Laufe von 20 Jahren

Tatsächlich hat sich die Qualität der Studien als validitätswirksam erwiesen, ebenso wie die Zahl der einbezogenen diagnostischen Verfahren und die Beteiligung von Psychologen als Beurteiler. Thornton et al. errechneten einen unkorrigierten, bezüglich Stichprobengröße gewichteten durchschnittlichen Validitätskoeffizienten von $r = .29$; der reliabilitäts- und streuungskorrigierte Wert wurde mit $p = .37$ ermittelt. Dieser Validitätskoeffizient $p = .37$ galt für die folgenden beiden Jahrzehnte als maßgeblicher

*Professor Dr. Heinz Schuler,
Inhaber des Lehrstuhls für
Psychologie, Universität Hohenheim,
und Wissenschaftlicher
Leiter der Managementberatung
S&F Personalpsychologie
schuler@uni-hohenheim.de*



Fotoatelier Ariane, Stuttgart

Eignungsdiagnostik

	Auszubildende		Trainees	Führungskräfte		
	techn.	kfm.		untere	mittlere	obere
Assessment-Center	21,3 (+ 21,3) [+ 21,3]	31,3 (+ 20,3) [+ 30,2]	50,0 (+ 10,0) [+ 31,2]	17,1 (+ 3,1) [+ 10,8]	17,4 (+ 2,4) [+ 9,9]	10,1 (- 1,9) [+ 4,2]
Persönlichkeitstest	6,4 (- 3,6) [- 3,6]	6,3 (- 3,7) [- 6,1]	6,3 (- 0,7) [± 0,0]	4,1 (- 2,9) [- 0,7]	6,6 (- 1,4) [- 3,8]	7,6 (+ 2,6) [- 4,2]
Intelligenztest	23,4 (- 4,6) [- 26,6]	31,3 (- 3,7) [- 26,0]	4,2 (- 3,8) [- 16,6]	0,8 (- 3,2) [- 4,0]	1,7 (- 0,3) [- 4,3]	0,8 (+ 0,8) [- 5,1]
Strukturiertes Interview durch Fachabteilung	28,7 (+ 13,7)	26,8 (+ 6,8)	42,7 (+ 10,7)	56,9 (+ 20,9)	54,5 (+ 17,5)	50,4 (+ 15,4)

Abb. 1: Einsatzhäufigkeit ausgewählter eignungsdiagnostischer Verfahren
 Angegeben ist der Prozentsatz der Unternehmen, die für die betreffende Beschäftigten-
 gruppe die jeweiligen Auswahlverfahren einsetzen (N = 125); in runder Klammer
 Veränderungswerte gegenüber 1993 (N = 105); in eckiger Klammer Veränderungswerte
 gegenüber 1985 (N = 120); Daten aus Schuler, Hell, Trapmann, Schaar & Boramir (2007)

Referenzwert für die prognostische Validität des ACs. Er lag niedriger, als aufgrund früherer Einzelstudien erwartet wurde, und enttäuschte vor allem viele Praktiker. Diese allerdings entnahmen der Metaanalyse insbesondere die Botschaft, dass ACs auch zu besseren Ergebnissen führen können, und warben deshalb unverdrossen mit Validitätswerten bis zu $r = .70$ und mehr (ohne freilich Belege dafür vorlegen zu können). Das bedenkliche Nebenergebnis, dass sich die durchschnittliche Validität des ACs innerhalb der rund 20 Jahre, die von der Thornton-Studie umfasst wurden, verringert hat, wurde so gut wie gar nicht zur Kenntnis genommen.

Gefahr: AC ist als Verfahren nicht mehr konkurrenzfähig

Immerhin liegt ein Wert von $p = .37$ noch im konkurrenzfähigen Bereich im Vergleich zu anderen Auswahlverfahren. Aber es sollte noch schlimmer kommen: Hardison und Sackett (2007) führten eine Nachfolgeuntersuchung durch, welche die nachfolgenden 20 Jahre umspannte. Aus 106 Validitätskoeffizienten ermittelten

sie eine durchschnittliche Validität von $p = .26$ und damit erheblich weniger als den bisher gültigen Referenzwert von $.37$. Die Koeffizienten aus den beiden Metaanalysen sind insofern nicht gut vergleichbar, als Hardison und Sackett Korrekturen nur bezüglich mangelnder Kriteriumsreliabilität vorgenommen haben, Thornton et al. dagegen zusätzlich in Bezug auf Streuungseinschränkungen. Zwei Möglichkeiten bieten sich, die Ergebnisse beider Metaanalysen dennoch zu vergleichen, nämlich die unkorrigierten Koeffizienten gegenüberzustellen ($r = .22$ vs. $.29$) oder Thorntons Daten ebenfalls ohne Streuungseinschränkungen zu berechnen, was zu Vergleichswerten von $p = .26$ vs. $.34$ führt. Es kommt dabei zu einem ähnlichen Bild: Neuere Studien berichten geringere Validität als ältere Studien. Bedenklich ist, dass auch die neuere Analyse eine negative Korrelation zwischen Publikationsjahr und Validitätshöhe aufzeigt ($r = -.25$), dass sich also der Abwärtstrend innerhalb des erfassten Zeitraums fortsetzt. Damit gerät das Assessment-Center in Gefahr, kein ernsthaft konkurrenzfähiges eignungsdiagnostisches Verfahren

mehr zu sein, zumal die Qualität anderer Verfahren – insbesondere des Einstellungsinterviews – verbessert wurde und neue brauchbare Verfahren, etwa Situational-Judgment-Tests, hinzugekommen sind.

ACs werden von Nicht-Psychologen durchgeführt

Wie ist es zu erklären, dass ein Verfahren an Qualität abgenommen und gleichzeitig verstärkten Zuspruch gefunden hat? Hier wird die These vertreten, dass dieser scheinbare Gegensatz auf eine gemeinsame Ursache zurückgeht: Das AC ist zur Spielwiese der Laiendiagnostik geworden. Dazu kommt: ACs arbeiten weit unter ihrer methodischen Kapazität. Sie werden vornehmlich von Nicht-Psychologen durchgeführt. Die Augenscheinmäßigkeit der eingesetzten Teilverfahren hat sich durchgesetzt. In vielen Fällen wird ausschließlich mit simulationsorientierten Teilverfahren („Übungen“) gearbeitet. Der Anforderungsbezug steht ebenso in Frage wie die vorgebliche Organisationspezifität. Ergebnisse zur psychometrischen Qualität werden gering bewertet oder gar nicht zur Kenntnis genommen. Mehrere Beiträge in Schuler (2007) zeigen, dass dieser Trend in deutschsprachigen Ländern noch stärker ausgeprägt ist als beispielsweise in den USA. In einer Einzelstudie wurde als Wiederholungsreliabilität von in deutschen Unternehmen durchgeführten ACs $r = .41$ ermittelt (Kelbetz & Schuler, 2002). Kleinmann (1997) fand eine Paralleltestreliabilität von sogar nur $r = .34$, wenn den Teilnehmern ein dimensionsbezogenes Feedback gegeben wurde.

Auch die Forschung hat ihren Beitrag zur Stagnation der AC-Entwicklung geleistet: Auf rigide Weise wurde die vermeintlich geringe Konstruktvalidität des Verfahrens verfolgt, um zu immer wieder neuen Bestätigungen der Entdeckung zu kommen, dass zwar mit verschiedenen Beurteilungsdimensionen innerhalb einer Aufgabe das Gleiche erfasst wird (was geringe diskriminante Validität anzeigt), nicht aber mit den gleichen Dimensionen

H. Schuler Spielwiese für Laien?

über verschiedene Aufgaben hinweg. Dieses stabile Ergebnis hat auch viele Psychologen unter den Praktikern dazu verführt, Beurteilern Unmögliches abzuverlangen, nämlich die trennscharfe Verwendung der Dimensionen und die Trennung von Beobachtung und Bewertung. Die empirische Bestätigung der Vermutung, dass verschiedene Aufgaben oder Übungen in unterschiedlichem Maße geeignet sind, bestimmte Fähigkeits- oder Urteilsdimensionen zu erfassen beziehungsweise dass Dimensionen in unterschiedlichen Aufgabenkontexten unterschiedliche Bedeutung haben (Guldin & Schuler, 1997), wurde nicht weiter zur Kenntnis genommen. Lance (2007) kommt auf aktueller empirischer Basis zum gleichen Ergebnis und plädiert für eine Umorientierung: Methodeneffekte sollten nicht länger als Fehlervarianz (miss-) verstanden werden, sondern als Ausdruck ebendieser Situationsspezifität des Verhaltens. Das vermeintliche Rätsel: „Wie kann ein eignungsdiagnostisches Verfahren bei geringer Konstruktvalidität hohe prognostische Validität aufweisen?“, hat sich damit in beiden Behauptungen als unzutreffend erwiesen.

Was tun? Maßnahmen zur Verbesserung von ACs

Was kann nun zur (Wieder-)Verbesserung des ACs getan werden? Alle Maßnahmen des Noch-mehr-vom-Gleichen haben sich als unwirksam erwiesen. Der hier vorgeschlagene Weg ist der, das AC als multiples eignungsdiagnostisches Verfahren ernst zu nehmen. Das heißt erstens, seine Entwicklung, Durchführung und Prüfung an den gleichen Standards zu orientieren, die generell in der psychologischen Diagnostik gelten. Zweitens bedeutet es, die Multiplizität beim Wort zu nehmen. Multipel heißt, es müssen verschiedene Teilverfahren kombiniert werden. Ob Teilverfahren verschieden sind, bemisst sich nicht nur nach dem Augenschein, aus psychometrischer Sicht bemisst es sich nach ihrer inkrementellen Validität, also nach ihrem Validitätszuwachs untereinander. Die inkrementelle Validität augenschein-

lich verschiedener AC-Übungen (zum Beispiel Gruppendiskussion, Rollenspiel, Präsentation) hat sich als äußerst gering erwiesen. Sie kann zwar durch geeignete Gestaltung erhöht werden, stößt aber an enge Grenzen.

Eine geeignete theoretische Basis für ein angereichertes AC könnte der multimodale oder trimodale Ansatz der Berufseignungsdiagnostik (Schuler, 2000) sein. Er stellt eine Konkretisierung der Idee des Kritischen Multiplizismus (Campbell & Fiske, 1959) in der Form dar, dass jedes Konstrukt durch verschiedene empirische Operationalisierungen erfasst wird. In der Personalpsychologie kann erwartet werden, dass durch einen multimodalen Ansatz die Validität und Generalisierbarkeit von Eignungs- und Leistungsdiagnosen verbessert wird. Die drei Modalitäten dieses Ansatzes sind Eigenschaften, Verhalten und Ergebnisse, ihre üblichen Operationalisierungsformen (eignungsdiagnostische Verfahren) sind Tests, Arbeitsproben und biografische Indikatoren (siehe Abbildung 2).

Die Vielfalt von Gesichtspunkten, bezüglich deren eignungsdiagnostische Verfahren klassifiziert werden können (Schuler, 2000), wird durch den trimodalen Ansatz reduziert und auf einen relativ einfachen Nenner ge-

bracht. Innerhalb jeder der drei Modalitäten gibt es immer noch eine Vielzahl von Variationsmöglichkeiten (beispielsweise kann die Testvorgabe schriftlich erfolgen oder die Form von Filmszenen haben, die Reaktion kann in aktuellem Handeln oder aber nur in symbolischer Form erfolgen etc.). Wesentlich ist jedoch, dass nicht alle eingesetzten Teilverfahren nur einer einzigen der drei Modalitäten zugehören.

Trimodaler Ansatz erreicht bessere Ergebnisse

Zwei Beispielverfahren für die praktische Umsetzung dieser Systematik werden bei Schuler (2007) dargestellt, ihre Validierung wird von Görlich, Schuler, Becker und Diemand (2007) berichtet. Die Verfahrensbestandteile dieser ACs oder „Potentialanalysen“ wurden entsprechend dem trimodalen Ansatz aus einer jeweils größeren Zahl von Einzelverfahren (auf der Basis gründlicher Anforderungsanalysen) nach Vorprüfungen ausgewählt. Zielsetzung war dabei nicht nur die prognostische Validität der Diagnosen, sondern auch die für jede Person gemäß ihren Interessen und Fähigkeiten angemessene „Funktionsfeld“-Zuordnung sowie die Akzeptabilität und Praktikabilität der Verfahren. Die Assessment-Center umfassten 13 beziehungsweise 14 Ein-

Abb. 2: Modalitäten und Diagnoseverfahren im trimodalen Ansatz der Berufseignungsdiagnostik

Modalität	Eigenschaften	Verhalten	Ergebnisse
Beispiele	Leistungsmotivation Soziale Kompetenz Raumvorstellung	Ziele setzen Verhandeln Techn. Fertigkeiten	Zielerreichung Umsatz Technisches Produkt
Typische Diagnoseverfahren	Tests	Arbeitsproben	Biografische Indikatoren
Beispiele	Motivationsinventar Persönlichkeitstest Kognitiver Fähigkeitstest	Planungsaufgabe Rollenspiel Hands-on-Arbeitsprobe	Zeugnisse Biografische Fragen Z.B. Nachweis erfolgreicher Reparaturen

Eignungsdiagnostik

zelterverfahren und konnten in jeweils einem Tag durchgeführt werden.

Die Ergebnisse der Validierung erfüllten oder übertrafen sogar die Erwartungen: Die unkorrigierten Validitätskoeffizienten für verschiedene Teilgruppen liegen zwischen $r = .40$ und $.48$, die korrigierten Werte betragen zwischen $p = .52$ und $.67$. Vergleicht man die Determinationskoeffizienten, das heißt die als Quadrat der Korrelationskoeffizienten bestimmte Varianzaufklärung im Kriterium, so ergibt sich, dass durch diese Potenzialanalyseverfahren je nach gewähltem Referenzwert (Thornton et al., 1987, oder Hardison & Sackett, 2007) zwischen der doppelten und der 6,5-fachen Prognosequalität im Vergleich zum herkömmlichen AC erzielt wird. Dementsprechend hoch ist der monetäre Nutzen der auf dieser Grundlage getroffenen Personalentscheidungen. Die ermittelten Zufriedenheitsdaten belegen, dass die psychometrische Qualität eines Verfahrens seiner Akzeptanz nicht im Wege stehen muss.

Das AC hat bei abnehmender Validität vermehrten Zuspruch gefunden. Beides hat die gleiche Ursache, nämlich die Vernachlässigung der psychometrischen Qualität zugunsten einer methodisch einseitigen, aber für diagnostische Laien ansprechenden und vermeintlich ohne diagnostische Kompetenz bewältigbaren Gestaltung. Es gibt Möglichkeiten, Assessment-Center wieder zu verbessern. Sie liegen

aber weder im dimensionsbezogenen Beurteilertraining noch in der Vermehrung gleichartiger Aufgaben, sondern in der multimodalen Gestaltung und der psychometrisch rigorosen Prüfung des Gesamtverfahrens und aller seiner Einzelkomponenten.

Weiterführende 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, 1–27.

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105.

Görlich, Y., Schuler, H., Becker, K. & Diemand, A. (2007). Evaluation zweier Potenzialanalyseverfahren zur internen Auswahl und Klassifikation. In H. Schuler (Hrsg.), *Assessment Center zur Potenzialanalyse* (S. 203–232). Göttingen: Hogrefe.

Guldin, A. & Schuler, H. (1997). Konsistenz und Spezifität von AC-Beurteilungskriterien: Ein neuer Ansatz zur Konstruktvalidierung des Assessment-Center-Verfahrens. *Diagnostica*, 43, 230–254.

Hardison, C. M. & Sackett, P. R. (2007). Kriteriumsbezogene Validität des Assessment Centers: lebendig und wohlauf? In H. Schuler (Hrsg.), *Assessment Center zur Potenzialanalyse* (S. 192–202). Göttingen: Hogrefe.

Kelbetz, G. & Schuler, H. (2002). Verbessert Vorerfahrung die Leistung im Assessment Center? *Zeitschrift für Personalpsychologie*, 1, 4–18.

Kleinmann, M. (1997). *Assessment Center. Stand der Forschung – Konsequenzen für die Praxis*. Göttingen: Hogrefe.

Lance, C. E. (2007). Weshalb Assessment Center nicht in der erwarteten Weise funktionieren. In H. Schuler (Hrsg.), *Assessment Center zur Potenzialanalyse* (S. 109–225). Göttingen: Hogrefe.

Schuler, H. (2000). Das Rätsel der Merkmals-Methoden-Effekte: Was ist „Potential“ und wie lässt es sich messen? In L. von Rosenstiel & T. Lang-von Wins (Hrsg.), *Perspektiven der Potenzialbeurteilung* (S. 53–71). Göttingen: Hogrefe.

Schuler, H. (Hrsg.). (2007). *Assessment Center zur Potenzialanalyse*. Göttingen: Hogrefe.

Schuler, H., Hell, B., Trapmann, S., Schaar, H. & Boramir, I. (2007). Die Nutzung psychologischer Verfahren der externen Personalauswahl in deutschen Unternehmen – ein Vergleich über 20 Jahre. *Zeitschrift für Personalpsychologie*, 6, 60–70.

Thornton III, G. C., Gaugler, B., Rosenthal, D. B. & Bentson, C. (1987). Die prädiktive Validität des Assessment Centers – eine Metaanalyse. In H. Schuler & W. Stehle (Hrsg.), *Assessment Center als Methode der Personalentwicklung* (S. 36–60). Göttingen: Verlag für Angewandte Psychologie.

WWW.GABLER.DE

Leitfaden für Assessment Center



Christof Obermann

Assessment Center

Entwicklung, Durchführung, Trends. Das Standardwerk mit originalen AC-Übungen

3. Aufl. 2006. ca. 440 S. Br. EUR 46,90 ISBN 978-3-8349-0208-5

Ja, ich bestelle Fax +49(0)611. 7878 - 412

Exemplare
Assessment Center
ISBN 978-3-8349-0208-5
EUR 46,90 zuzügl.
Versand EUR 3,32

Name, Vorname _____ Firma _____
Straße (bitte kein Postfach) _____ PLZ | Ort _____ 22107006
Datum | Unterschrift _____

Änderungen vorbehalten. Erhältlich im Buchhandel oder beim Verlag. Geschäftsführer: Andreas Kösters, Dr. Ralf Birkelbach, AG Wiesbaden HRB 9754.

