

Werner Sarges

# WARUM ASSESSMENT CENTER HÄUFIG ZU KURZ GREIFEN

**- UND ZUDEM MEIST DAS FALSCHES ZU MESSEN VERSUCHEN <sup>1</sup>**

*Aktualisierte Version vom 05.03.2015*

In diesem Beitrag werden – mit Blick auf die Praxis des Assessment-Center (AC)–Ansatzes – zwei Problembereiche kritisch behandelt: einerseits das verfahrensmäßig zu homogene und damit *zu kurz greifende AC-Konzept*; und andererseits die nach wie vor bestehende Schwachstelle *Beurteilung des Kandidatenverhaltens*. Die vorgetragenen Überlegungen basieren auf langjährigen Erfahrungen des Autors in und mit der Praxis und deren Reflexion mit Bezug auf den Stand der Forschung.

## ZUR VERKÜRZUNG DES ASSESSMENT-CENTER-KONZEPTS

Bedauerlicherweise hat sich in AC–Guidelines und in großen Teilen der Praxis die methodisch verkürzte AC–Konzeption breit gemacht, nur verhaltensbezogene, arbeitsprobenartige Aufgaben zuzulassen. Und vor gut einem Jahrzehnt kam auch noch die von der American Psychological Association (1999) für 2000–2010 proklamierte ‚Decade of Behavior‘ (im Anschluss an die vorherige ‚Decade of the Brain‘) einer solchen auf Verhaltensaufgaben reduzierten AC–Auffassung entgegen. Diesem Verständnis folgen auch Kanning, Pöttker und Gelléri (2007) bei ihrer Aufstellung wichtiger Standards für Prozesse der Gestaltung und Durchführung von ACn. Indes: Selbst wenn sie die Einbeziehung weiterer Verfahren bei ACn als Option erwähnen, so mahnen sie – ebenso wenig wie Kleinmann (2013) – diese nicht eigens als einen Standard an, der zugleich konstitutiv für das Gesamtverfahren wäre.

Vorsicht! Derartig amputierte AC sparen wichtige psychologische Bereiche, wie etwa den des Antriebs, allein methodenbedingt systematisch aus. Schon Atkinson (1978) schätzte die Bedeutung der Motive für den Berufserfolg generell höher ein als die der Fähigkeiten. Bei Führungskräften im besonderen dürften sie eine noch größere Rolle spielen; schließlich war eine solche Vermutung der wesentliche Grund dafür, dass gerade diese Personengruppe so früh in den Fokus der Motivforschung geriet: Einerseits wird es mit aufsteigender Position immer schwieriger, Erfolg versprechende Könnens–Faktoren zu identifizieren, andererseits werden Wollens–Faktoren umso einflussreicher, je unabhängiger ein Funktionsträger über das „Ob“, „Was“ und „Wie“ seiner Handlungen bestimmen kann (Krug & Rheinberg, 1987). Im Gegensatz zum Alltagsverständnis verstehen wir hier unter Motiven nicht konkrete Ziele, sondern Inhaltsklassen überdauernder Handlungsziele, also Bedürfniskerne wie ‚Leistung‘, ‚Macht‘ oder ‚Anschluss‘, die Menschen über die gesamte Lebensspanne hinweg mehr oder weniger

---

<sup>1</sup> Dieser Beitrag ist die in 2013 aktualisierte und überarbeitete Fassung eines Beitrags des Verfassers mit gleichem Titel, der zuerst 2009 erschienen ist in: *Zeitschrift für Arbeits- und Organisationspsychologie*, Jg. 53, S. 79–82.

stark und ausdauernd anstreben (sensu Heckhausen in der Tradition von McClelland; vgl. Heckhausen & Heckhausen, 2006). Dies mit Hilfe von verhaltensbezogenen Aufgaben in 1–2 Tagen abschätzen zu können ist schlicht unmöglich. Dazu sind biographische Informationen und Testverfahren nötig. Ohne diese beiden zusätzlichen Verfahrensklassen kann man übrigens auch nicht ernsthaft die für Führungskräfte so erfolgsrelevante Fähigkeit erfassen, stabile Netzwerke aufbauen, pflegen und verfeinern zu können, um vertrauensbasierte Gefolgschaften zu ermöglichen.

Da nun AC bevorzugt für die Auswahl, Allokation und Entwicklung von Führungskräften und deren Nachwuchs herangezogen werden (seltener für die von Experten o.a.), sollten in solchen Veranstaltungen wirklich ausreichende Breitbandfassungen der leistungs- und karriererelevanten Eigenschafts- und Verhaltensmuster angestrebt werden, d.h. es sollten außer Verhaltenssimulationen (in Einzel-, Zweier- und Gruppenaufgaben) auch – wenn immer möglich – Interviews und Tests einbezogen werden. Eine einzelne Methode allein kann gar nicht jedes der interessierenden Merkmale vollständig erfassen; daher sollte man möglichst viele Merkmale mit unterschiedlichen Verfahren zu messen suchen, um Informationen zu erhalten, die sich ergänzen und dann zu einem valideren Gesamtbild zusammengefügt werden können (Schuler, 2000; Templer, 1995). Schon lange plädiere ich deshalb dafür (in der Praxis wie in der Wissenschaft), AC wieder als multimethodale Verfahrensgruppe aufzufassen und zu handhaben, so wie es paradigmatisch bei AT&T schon in den 1950er Jahren als frühe Voraussicht späterer Einsichten realisiert wurde (Sarges, 2001) – womit ich mich in guter Gesellschaft mit anderen praxisorientierten Wissenschaftlern befinde.

Für Schuler etwa sind Assessment Center „ein Fall von Qualitätsverlust“ (2009, S. 189f.). Er begründet das mit abfallenden Validitätskoeffizienten: von .47 über .37 auf inzwischen .26 im Verlauf der letzten ca. 50 Jahre. Eine Erklärung dieser Abnahme der prognostischen Validität sieht er einerseits darin, dass AC in weiten Teilen inzwischen zu einer „Spielwiese der Laiendiagnostik“ verkommen seien, also in mangelnder psychologischer Expertise der Designer und Assessoren von ACn. Vor allem aber schlägt auch er vor, AC wieder als multiples eignungsdiagnostisches Verfahren zu begreifen und dabei „den Gedanken der Multimodalität ernst zu nehmen und – wie es ehemals der Fall war – wieder verschiedene Teilverfahren von nachweislich inkrementeller Validität zu kombinieren“ (S. 190).

Es wäre daher zu begrüßen, wenn das Prinzip „Multimethodalität“ zum zentralen Standard für AC avancieren würde. Glücklicherweise sprechen die Ergebnisse der AC-Studie von Obermann, Höft und Janke (2008) dafür, dass Persönlichkeits- und Intelligenztests auch im Rahmen von ACn auf dem Vormarsch sind und damit die Zeiten zum Propagieren eines (wieder) erweiterten AC-Konzepts inzwischen günstiger geworden sind.

## **ZUM IGNORIEREN DER SCHWACHSTELLE ‚BEURTEILUNG‘**

Viele Arbeits- und Organisationspsychologen (z.B. auch Kanning et al., 2007) fordern, dass AC dem Multi-Trait-Multi-Method-Ansatz (MTMM) folgen sollten, und zwar in der Weise, dass jede Anforderungs- bzw. Beurteilungsdimension in mindestens zwei voneinander unabhängigen Verhaltensaufgaben (meist irreführenderweise „Übungen“ genannt) untersucht werden müssten. Dies aber spiegelt mehr einen paradigma-diktierten Methodenzwang wider als eine empirisch begründbare Zweckmäßigkeit (vgl. dazu auch Hoffman & Meade, (2012). Denn spätestens seit 1982 (Sackett & Dreher) ist bekannt und problemlos replizierbar, dass die Beurteilungen gleicher Dimensionen in verschiedenen Übungen wesentlich geringer un-

tereinander korrelieren als die Beurteilungen verschiedener Dimensionen in der gleichen Übung. Offenbar spielen weniger die zu beobachten angestrebten Merkmale eine Rolle als die unterschiedlichen Verhaltensaufgaben. Beobachter differenzieren demnach nicht hinreichend zwischen den diversen Beobachtungsdimensionen, sondern geben pro Proband nur eine eher globale Einschätzung ab – vermutlich als Antwort auf die implizit vereinfachte Frage „Wie gut war der Proband in dieser Übung?“. Dies würde die stark überzogene Merkmalskohärenz ganz einfach mit dem Halo-Effekt erklären, wenn man Kahneman's (2011) Überlegungen zu dem kognitiven „System 1“ mit seiner Tendenz zu voreiligen Schlussfolgerungen heranzieht, etwa die automatische Operation „Nur was man gerade sieht oder weiß, zählt“ (S. 112ff.).

Immerhin ist es erfreulich, dass wenigstens inzwischen – lange nach der kritischen Publikation von Sackett und Dreher (1982) – Einsichten wie die folgende formuliert und publiziert werden: „An assessment centre is composed of several exercises which are carefully selected to cover specific job-related competences. Consequently, they place different psychological demands on the assesseees. For instance, one might expect an assessee to behave differently – even inconsistently – in a one-to-one role-play as compared to in a group discussion“ (Lievens, 2009, S. 109).

Dass das Verhalten von Kandidaten in Assessment Centern zwangsläufig hochgradig übungsspezifisch ist, war als hartnäckig-robustes Phänomen allen aufmerksamen Praktikern und mehr noch den die Praxis beratenden Wissenschaftlern schon seit über 3 Jahrzehnten augenfällig: Beobachter beurteilen – ganz gleich, welche und wie viele Beurteilungsdimensionen ihnen vorgelegt werden – mehr oder weniger gesamthaft die Leistung eines Probanden in der ganz konkreten Verhaltensaufgabe, wahrscheinlich nach den in der Organisation offen oder verdeckt herrschenden Gütemaßstäben für die entsprechenden Rollen- oder Verhaltensmuster.

Nimmt man diese akkumulierte – wissenschaftliche wie praktische – Erfahrung wirklich ernst, ergibt sich als Konsequenz, den Fokus von der Messung der Reaktionsseite, den klassisch nach dem Eigenschaftsansatz konzipierten Beurteilungsdimensionen, wegzunehmen und auf die Gestaltung der Stimulusseite, d.h. die Verhaltensaufgaben, zu richten. Je leistungs- und karrierekritischer – also je inhaltlich und ökologisch valider in dieser Hinsicht – die Auswahl und Gestaltung der simulativen Verhaltensaufgaben ist, umso besser dürften Prognosen, aber auch Diagnosen von Stärken und Schwächen für Entwicklungsmaßnahmen möglich werden.

Nun haben Sackett und Dreher nach ihrer o.g. Entdeckung der im MTMM-Design ermittelten mangelnden Konstruktvalidität selbst vorgeschlagen, auf einzelne Beurteilungsdimensionen zu verzichten und stattdessen die Leistungsgüte der Probanden in Verhaltenssimulationen einer Vielzahl wichtiger Managerrollen (Verhandeln, Beraten, Untersuchen, Überzeugen etc.) einschätzen zu lassen. Dieser Vorschlag wurde von der Wissenschaft erst spät und auch nur spärlich aufgegriffen (vgl. Lance, 2007), ist aber in der Praxis von einigen Beratern und Organisationen – entgegen der Auffassung des seinerzeitigen wie heutigen Mainstreams – schon früh explizit umgesetzt worden. In diesen Fällen schätzen die Beurteiler einfach ein, welchen Gütegrad das Rollenverhalten des Probanden X in der Aufgabe Y erreicht hat, und vielleicht auch noch, woran sie das festmachen und was in ihren Augen (noch) fehlt. Systemisch betrachtet bedeutet das: Man arbeitet auf diese Weise nicht *gegen* das implizite Beurteilungsverhalten von Beobachtern, sondern *mit* ihm.

Unser Institut jedenfalls praktiziert derartige Beurteilungsvereinfachungen seit Beginn der 90er Jahre im Rahmen einer auch in anderer Hinsicht modifizierten AC-Konzeption (Lern-Potenzial-AC, Sarges, 1992; 1995, S. 735); wir können einen solchen Wechsel im Beurteilungsprozedere nur weiterempfehlen. Doch damit ist das trotzig zähe Festhalten an den komplizierteren, aber in die Sackgasse führenden Beurteilungsverfahren wohl kaum aufzuweichen, eher vielleicht mit einem berufsethischen Appell: „Es ist an der Zeit, den Wert der Forschungsergebnisse des letzten Vierteljahrhunderts anzuerkennen und Beurteilungen umzuorientieren“ (Lance, 2007, S.122). Auch Schuler (2009) ist inzwischen der Meinung, dass sich die dimensionsbezogenen Beurteilertrainings als unwirksam erwiesen haben – eine Einsicht, die zeitgleich auch bei Jackson (2007) gereift ist, und folgerichtig empfehlen er und seine Kollegen (Jackson, Ahmad, Grace & Yoon, 2011) – wie zuvor schon Sackett und Dreher (1982), Sarges (1992,1995, 2009) und Lance (2007) –, bei der Konzeption von ACn von einer Orientierung an Dimensionen zu einer Orientierung an Aufgaben fortzuschreiten.

Welchen Nutzen bringt eine solche Umorientierung? Nach all unseren Erfahrungen reduziert sie die kognitive Belastung der Beobachter durch die vorher aufwändige Beobachtungs- und Bewertungsadministration erheblich und eröffnet damit die nötigen Freiräume zu einer umfassenderen Sichtweise. Wie durch ein Makroskop erweitert sich dann der Blick: von kleinteiligen Verhaltensfacetten auf großflächigere Verhaltensmuster nicht nur eines oder zweier Akteure, sondern auch der anderen Spieler in der Assessment-Arena, und das zugleich in ihrer Einbettung in die soziale und kommunikative Dynamik der jeweiligen Verhaltensaufgabe.

Darauf müssen die Beobachter natürlich in geeigneter Weise vorbereitet werden, ganz im Sinne einer Bemerkung Einsteins zu Heisenberg: „Die Theorie bestimmt, was wir beobachten“, hier natürlich bescheidener: Das Verhaltens-Konzept bestimmt, worauf wir achten. Die relevanten Konzepte sind aber fairer Weise nicht nur den Beobachtern zugänglich zu machen, sondern auch den Probanden, um dysfunktionalem „Herrschaftswissen“ vorzubeugen und – wie für Leistungstests allemal zu empfehlen – die Chancengleichheit der Probanden zu erhöhen. Dazu müssen diese hinreichend lange vor dem AC ermuntert werden, Einiges an favorisierter praktischer Managementliteratur (incl. Videos) zu konsultieren, insbesondere zu Gesprächsführung, Moderation, Präsentation, Problemlösen und Verhandeln. Dies, damit sie sich methodisch kompetenter machen können und ohne raten zu müssen wissen, welche Verhaltenserwartungen gehegt werden – was schließlich einer Reduktion der Fehlervarianz zugute kommen dürfte (Bungard, 1987).

## LITERATUR

- American Psychological Association (1999). 'Decade of Behavior' moves forward. *APA MONITOR Online*, 30 (3). Verfügbar unter: <http://www.decadeofbehavior.org/index.cfm>
- Atkinson, J.W. (1978). Motivational determinants of intellectual performance and cumulative achievement. In J.W. Atkinson & J.O. Rynor (Eds.), *Personality, motivation, and achievement* (pp. 221–242). Washington: Hemisphere.
- Becker, N., Höft, S., Holzenkamp, M. & Spinath, F.M. (2011). The predictive validity of assessment centers in German-speaking regions – A meta-analysis. *Journal of Personnel Psychology*, 10, 61–69.
- Bungard, W. (1987). Zur Problematik von Reaktivitätseffekten bei der Durchführung eines Assessment Centers. In H. Schuler & W. Stehle (Hrsg.), *Assessment Center als Methode der Personalentwicklung* (S. 99–125). Stuttgart: Verlag für Angewandte Psychologie.

- Görn, A. (2002). *Optimierung von Assessment-Center-Verfahren in der Personalentwicklung auf der Grundlage theoretischer Erkenntnisse der Lern- und Leistungsmotivationsforschung*. Berlin: TENEA-Verlag.
- Heckhausen, J. & Heckhausen, H. (2006). Motivation und Handeln: Einführung und Überblick. In J. Heckhausen & H. Heckhausen (Hrsg.), *Motivation und Handeln* (3. Aufl.; S. 1–10). Heidelberg: Springer.
- Hoffman, B.J. & Meade, A. (2012). Alternate approaches to understanding the psychometric properties of Assessment Centers: An analysis of the structure and equivalence of exercise ratings. *International Journal of Selection and Assessment*, 20, 82–97.
- International Task Force on Assessment Center Guidelines. (2009). Guidelines and ethical considerations for assessment center operations. *International Journal of Selection and Assessment*, 17, 243–253.
- Jackson, D.J.R. (2007). Task-specific assessment centers: Evidence of predictive validity and fairness. Paper to the Society for Industrial and Organizational Psychology, New York. (April)
- Jackson, D., Ahmad, M.H., Grace, G. & Yoon, J. (2011). An alternative take on AC research and practice: Task-based assessment centers. In N. Povah & G.C. Thornton III (Eds.), *Assessment centres and global talent management* (pp. 33–46). Farnham, UK: Gower.
- Kahneman, D. (2011). *Schnelles Denken, Langsames Denken*. München: Pantheon.
- Kanning, U.P., Pöttker, J. & Gelléri, P. (2007). Assessment Center-Praxis in deutschen Großunternehmen – Ein Vergleich zwischen wissenschaftlichem Anspruch und Realität. *Zeitschrift für Arbeits- und Organisationspsychologie*, 51, 155–167.
- Kleinmann, M. (2013). Assessment-Center. In W. Sarges (Hrsg.), *Management-Diagnostik* (4., vollst. überarb. u. erw. Aufl.; S. 809–819). Göttingen: Hogrefe.
- Krug, S. & Rheinberg, F. (1987). Motivation von Führungskräften. In A. Kieser et al. (Hrsg.), *Handwörterbuch der Führung* (Sp. 1510–1520). Stuttgart: Poeschel.
- Lance, C.E. (2007). Weshalb Assessment Center nicht in der erwarteten Weise funktionieren. In H. Schuler (Hrsg.), *Assessment Center zur Potenzialanalyse* (S. 109–125). Göttingen: Hogrefe.
- Lievens, F. (2009). Assessment centres: A tale about dimensions, exercises, and dancing bears. *European Journal of Work and Organizational Psychology*, 18, 102–121.
- Obermann, C., Höft, S. & Janke, O. (2008). *Was machen die DAX-100-Unternehmen? Die große AkAC-Studie 2008*. Vortrag auf dem 7. Deutschen Assessment-Center-Kongress, Arbeitskreis Assessment Center e.V., 25.–26.9.2008 in Potsdam.
- Sackett, P.R. & Dreher, G.F. (1982). Constructs and assessment center dimensions: Some troubling empirical findings. *Journal of Applied Psychology*, 67, 401–410.
- Sarges, W. (1992). *Eine neue Assessment-Center-Konzeption: Das Lernfähigkeits-AC*. Vortrag auf der 34. Fachtagung der Sektion Arbeits-, Betriebs- und Organisationspsychologie im BDP, 25.–26.5.1992 in Bad Lauterberg (Harz).
- Sarges, W. (1995). Lern-Potential-AC. In W. Sarges (Hrsg.), *Management-Diagnostik* (2., vollst. überarb. u. erw. Aufl.; S. 728–739). Göttingen: Hogrefe.
- Sarges, W. (2001). Die Assessment Center-Methode – Herkunft, Kritik und Weiterentwicklungen. In W. Sarges (Hrsg.), *Weiterentwicklungen der Assessment Center-Methode* (2., überarb. u. erw. Aufl.; S. VII–XXXII). 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 Potentialbeurteilung* (S. 53–71). Göttingen: Hogrefe.

- Schuler, H. (2009). Arbeits- und Organisationspsychologie: Berufseignungsdiagnostik und Personalauswahl. In G. Krampen (Hrsg.), *Psychologie – Experten als Zeitzeugen* (S. 180–194). Göttingen: Hogrefe.
- Templer, K.-J. (1995). Zusammenhänge zwischen Aufgabentypen beim Assessment Center. *Zeitschrift für Arbeits- und Organisationspsychologie*, 39, 179–181.

---

als Download bereitgestellt vom  
*Institut für Management-Diagnostik*  
<http://www.sarges-partner.de/management-diagnostik.php>

---