

Werner Sarges

WARUM ASSESSMENT CENTER HÄUFIG ZU KURZ GREIFEN

- UND ZUDEM MEIST DAS FALSCHES ZU MESSEN VERSUCHEN ¹

Aktualisierte Version vom 09.05.2011

In ihrem Beitrag „Assessment Center-Praxis in deutschen Großunternehmen – Ein Vergleich zwischen wissenschaftlichem Anspruch und Realität“ (Kanning, Pöttker & Gelléri, 2007) in dieser Zeitschrift begründen die Autoren die Aufstellung wichtiger Standards für Prozesse der Gestaltung und Durchführung von Assessment Centern (ACn); des weiteren berichten sie über eine Klärung der Einhaltung dieser Standards in der AC-Praxis mittels einer eigenen Repräsentativ-Umfrage bei über 200 deutschen Großunternehmen. Dies ist für ein dynamisches und wenig transparentes Diagnostik-Feld ein wertvoller Informations- und Orientierungsgewinn, der fraglos Anerkennung verdient.

Gleichwohl möchte ich in der dadurch angestoßenen Diskussion zweierlei kritisch anmerken, nämlich

1. dass das dem o.g. Beitrag zugrunde liegende AC-Konzept zwar weit verbreitet ist, aber eigentlich zu kurz greift, und
2. dass die von der Forschung schon lange und immer wieder festgestellte Schwachstelle ‚Beurteilung‘ bei der Aufstellung einiger Standards offenbar ignoriert wurde.

Bei meinen Überlegungen stütze ich mich auf langjährige Erfahrungen in und mit der Praxis und deren Reflexion, stets auch im Hinblick auf den jeweiligen Stand der Forschung.

ZUR VERKÜRZUNG DES AC-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 neuerdings kommt 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 et al.; selbst wenn sie die Einbeziehung weiterer Verfahren bei ACn als Option erwähnen, so mahnen sie diese nicht eigens als einen Standard an, der zugleich konstitutiv für das Gesamtverfahren wäre.

¹ erschienen in: *Zeitschrift für Arbeits- und Organisationspsychologie*, 2009, Jg. 53 (Heft 2), S. 79–82.

Vorsicht! Derartig amputierte AC sparen wichtige 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 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 kann gar nicht jedes der interessierenden Merkmale erfassen; und zur Vermeidung weiterer Monomethodenfehler sollte man möglichst viele Merkmale mit unterschiedlichen Methoden zu messen suchen, um sich ergänzende Informationen zu erhalten, die 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 50er Jahren als frühe Voraussicht späterer Einsichten realisiert wurde (Sarges, 2001) – womit ich mich in guter Gesellschaft mit etlichen praxisorientierten Wissenschaftlern befinde.

Es wäre daher zu begrüßen, wenn Kanning et al. das Prinzip der Multimethodalität als zentralen Standard für AC in ihren Katalog aufnahmen. Im übrigen spricht die Tatsache, dass nach der jüngsten AC-Studie (Obermann, Höft & Janke, 2008) Persönlichkeits- und Intelligenztests auch im Rahmen von ACn auf dem Vormarsch sind, dafür, dass die Zeiten zum Propagieren eines (wieder) erweiterten AC-Konzepts inzwischen günstiger geworden sind.

ZUM IGNORIEREN DER SCHWACHSTELLE ‚BEURTEILUNG‘

Bei Standard 5 fordern Kanning et al., dass AC dem Multi-Trait-Multi-Method-Ansatz folgen sollten, und zwar in der Weise, dass jede Anforderungsdimension 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. Denn spätestens seit 1982

(Sacket & Dreher) ist bekannt und problemlos replizierbar, dass die Beurteilungen gleicher Dimensionen in verschiedenen Übungen geringer untereinander korrelieren als die Beurteilungen verschiedener Dimensionen in der gleichen Übung. Offenbar spielen weniger die zu beobachtenden Merkmale eine Rolle als die unterschiedlichen Verhaltensaufgaben. Beobachter differenzieren also nicht hinreichend zwischen den verschiedenen Dimensionen, sondern geben pro Proband eine eher globale Einschätzung ab, quasi als Antwort auf die implizit vereinfachte Frage „Wie gut war der Proband in dieser Übung?“

Auch aufmerksamen und erfahrenen Praktikern war dieses hartnäckig-robuste Phänomen schon lange augenfällig: Beobachter beurteilen – ganz gleich, welche und wie viele Anforderungsdimensionen ihnen vorgelegt werden – mehr oder weniger gesamthaft die Leistung eines Probanden in der ganz konkreten Verhaltensaufgabe, vermutlich 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 Merkmalsdimensionen, 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. Allerdings wäre das seitens der Forschung noch empirisch zu überprüfen.

Nun haben Sacket und Dreher selbst nach ihrer o.g. Entdeckung der im MTMM-Design ermittelten mangelnden Konstruktvalidität vorgeschlagen, auf globale 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 Übung 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 Beratungsinstitut jedenfalls praktiziert derartige Beurteilungsvereinfachungen seit Beginn der 90er Jahre im Rahmen einer auch in anderer Hinsicht modifizierten AC-Konzeption (Lern-Potential-AC: Feedback schon während des Assessment-Centers; Sarges, 1992; 1995, S. 735), und wir können einen solchen Wechsel im Beurteilungsprozedere nur weiterempfehlen. Doch damit ist das trotz 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).

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 einer oder zweier Personen, 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 Management-Literatur (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 kommt (Bungard, 1987).

Eine Umsetzung dieser Empfehlungen in die Standards von Kanning et al. hätte dann außer dem Aufgeben von Standard 5 eine entsprechende Modifikation der Standards 10 und 13 zur Folge.

LITERATUR

- American Psychological Association (1999). ‘Decade of Behavior’ moves forward. *APA MONITOR Online*, 30 (3). <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.
- 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.
- Heckhausen, J. & Heckhausen, H. (2006). Motivation und Handeln: Einführung und Überblick. In J. Heckhausen & H. Heckhausen (Hrsg.), *Motivation und Handeln* (S. 1–10; 3. Aufl.). Heidelberg: Springer.
- 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.
- 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.
- Obermann, C., Höft, S. & Janke, O. (2008). *Was machen die DAX100-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.
- Templer, K.-J. (1995). Zusammenhänge zwischen Aufgabentypen beim Assessment Center. *Zeitschrift für Arbeits- und Organisationspsychologie*, 39, 179–181.