



Assessment Center: Ein valides Instrument zur Verhaltens- prognose von Führungskräften

Ein Assessment Center (AC) ist ein diagnostisches Instrument, welches in der Personalauswahl und -entwicklung sowie bei Potenzialanalysen eingesetzt wird (Thornton & Gibbons, 2009). Innerhalb eines ACs durchlaufen die Teilnehmer¹ verschiedene Übungen. Die Übungen stellen arbeitsrelevante Situationen dar. Während den Übungen beobachten und beurteilen mehrere geschulte Assessoren das Verhalten eines Teilnehmers nach zuvor definierten Bewertungsdimensionen (International Task Force on Assessment Center Guidelines, 2015).

ACs werden häufig für die Auswahl und Entwicklung von Führungskräften konzipiert (Gaugler, Rosenthal, Thornton, & Bentson, 1987; Munchus & McArthur, 1991) sowie für Mitarbeiter, bei denen kommunikative und überfachliche Anforderungen besonders wichtig sind (Obermann, 2013).

Der Aufbau und die Durchführung eines ACs gehen einher mit zeitlichen und finanziellen Ressourcen. Im Vergleich zu anderen Personalauswahlverfahren haben aber ACs den Vorteil, dass konkrete Verhaltensweisen und erfolgsentscheidende sowie überfachliche Eigenschaften wie die Motivationsfähigkeit von Menschen tatsächlich beobachtet und bewertet werden können (Obermann, 2013). Ein weiterer Vorteil ist, dass geschulte Assessoren die AC-Teilnehmer beobachten sowie bewerten und deshalb Selbsteinschätzungen durch die AC-Teilnehmer oder Fremdeinschätzungen durch die Arbeitskollegen oder Mitarbeiter der AC-Teilnehmer entfallen können. Denn ein Problem bei der Selbsteinschätzung stellen die Fehlbeurteilungen durch Selbstpräsentationseffekte dar (z.B. Ones & Viswevaran, 1998). Ein Problem bei der Einschätzung durch Mitarbeitende stellen Verzerrungen durch die emotionale Beziehung der Mitarbeitenden zum AC-Teilnehmer dar (Brown & Keeping, 2005). Die Assessoren im AC stehen aber in der Regel in keinem Abhängigkeitsverhältnis zu den AC-Teilnehmern und sollten diese deshalb möglichst objektiv bewerten können. Eine möglichst objektive Messung der Anforderungsdimensionen sollte eine akkurate Vorhersage der Kriterien (z.B. Indikatoren von beruflichem Erfolg) begünstigen.

¹ Im vorliegenden Artikel werden zur einfacheren Lesbarkeit nur die männlichen Formen verwendet, die weiblichen Formen sind eingeschlossen.



„Ein bisschen Theorie schadet nie“: Qualitätskriterien des ACs – und wie sie von adt zürich gelebt werden

Gütekriterien

In der Psychologie gibt es verschiedene anerkannte Kriterien, nach denen die Güte eines Verfahrens beurteilt werden kann. Die Kriterien unterteilt man in Haupt- und Nebengütekriterien (Lienert & Raatz, 1998). Zu den Hauptgütekriterien gehören die Objektivität, Reliabilität und Validität. Die Hauptgütekriterien lassen sich weiter differenzieren. Einen guten Überblick der Haupt- und Nebengütekriterien gibt Bühner (2011). Im Folgenden werden die Hauptgütekriterien und deren Zusammenhang mit ACs erläutert.

Objektivität

Das erste Hauptgütekriterium ist die Objektivität. Objektivität ist definiert als der Grad, in dem die Ergebnisse eines Tests unabhängig vom Untersucher sind (Bühner, 2011). In Zusammenhang mit ACs ist vor allem die Auswertungs- und die Interpretationsobjektivität relevant. Die Auswertungsobjektivität ist hoch, wenn verschiedene Auswerter zum selben Ergebnis kommen. Die Interpretationsobjektivität ist hoch, wenn verschiedene Auswerter zur gleichen Beurteilung der Testwerte kommen (Bühner, 2011). Durch eine gute Assessorenschulung und durch den Einsatz mehrerer Assessoren pro AC-Teilnehmer können diese Herausforderungen gemeistert und die Qualität der Ergebnisse gesteigert werden. Zudem hilft die Assessorenkonferenz am Ende jeder Übung oder am Ende des ACs, dass beide Assessoren allfällige Diskrepanzen in den Beurteilungen bereinigen können und schliesslich zur gleichen Bewertung kommen. Alle Assessoren von adt zürich haben eine umfangreiche Assessorenschulung absolviert und besitzen eine langjährige Erfahrung in der Tätigkeit als Assessor sowie als Führungskraft. Zudem setzt adt zürich bei jedem AC zwei Assessoren ein, um die Objektivität zu garantieren. Das Vier-Augen-Prinzip ist ein grundlegendes Element der ACs von adt zürich. Nach jeder Übung besprechen die Assessoren in der Assessorenkonferenz das beobachtete Verhalten und bereinigen allfällige Differenzen. Anschliessend schreibt jeder Assessor einen Bericht. Nach dem durchgeführten AC werden die zwei Berichte in einen Schlussbericht zusammengefasst und dem Auftraggeber bis spätestens 12:00 Uhr des Folgetags elektronisch versendet.



Reliabilität

Das zweite Hauptgütekriterium ist die Reliabilität. Reliabilität ist definiert als der Grad der Messgenauigkeit eines Messwertes (Bühner, 2011). Als Messwerte dienen beispielsweise Persönlichkeits- oder Verhaltensmerkmale (Lienert & Raatz, 1998). In Zusammenhang mit ACs ist es wichtig, dass vor der Durchführung des ACs eine Anforderungsanalyse durchgeführt wird. Gemäss Obermann (2013) sollen zwei Fragen beantwortet werden: Welche Anforderungen kommen auf den Stelleninhaber zu und mit welchen Kompetenzen, Eigenschaften sowie Verhaltensweisen ist diesen Anforderungen am besten zu begegnen? Die Genauigkeit der AC-Beurteilungen und schliesslich der Prognosen über den Berufserfolg ist abhängig von der Definition der Anforderungen und deshalb ist es von Bedeutung, die nötigen Anforderungsdimensionen vor der Durchführung eines ACs sorgfältig zu bestimmen und zu definieren. Um die Reliabilität von ACs zu sichern ist es ausserdem wichtig, dass jede Anforderungsdimension in mindestens zwei Übungen beurteilt wird. Dazu kann man eine Übungs-Dimensions-Matrix erstellen. Eine Matrix könnte beispielsweise so aussehen:

Übungen:	Interview	Komm. - Übungen	Test log.- analyt. Den- ken	Präsen- ta- tion	Fallstudien	Argument- Übung
Kompetenzen:						
Persönlichkeitsformat	✓	✓		✓	✓	✓
Leistungsorientierung / Belastbarkeit	✓	✓		✓	✓	✓
Kommunikationsfähigkeit	✓	✓		✓		✓
Kritik- und Konfliktfähigkeit	✓	✓			✓	✓
Führungs- und Lenkungsfähigkeit	✓	✓		✓	✓	
Komplexe Problemlösungsfähigkeit	✓		✓		✓	

Tabelle 1: Übungs-Dimensions-Matrix

adt zurich analysiert vor der Durchführung des ACs in Zusammenarbeit mit dem Auftraggeber, welche Kompetenzen beurteilt werden sollen. Danach werden massgeschneiderte Übungen – ebenfalls ein Markenzeichen von adt zurich – erstellt. Diese Übungen decken grösstenteils den tatsächlichen Aufgabenbereich des zukünftigen Stelleninhabers inhaltlich ab. Um die Messgenauigkeit der eingesetzten ACs zu steigern, wird jede Anforderungsdimension in mindestens zwei Übungen beurteilt. Nach dem Prinzip der Methodenvielfalt ist es wichtig, dass verschiedene Arten von Übungen eingesetzt werden. adt zurich stellt deshalb unterschiedliche Methoden und Verfahren zusammen und kombiniert diese miteinander. Dank dieser Vorgehensweise ist die Messgenauigkeit der eingesetzten ACs garantiert.



Validität

Das dritte Hauptgütekriterium ist die Validität. Validität ist definiert als Ausmass, in dem ein Test das misst, was er zu messen beansprucht (Bühner, 2011). Validität wird unterteilt in Inhalts-, Kriteriums- und Konstruktvalidität. In Zusammenhang mit ACs ist das entscheidende Qualitätsmass die Kriteriumsvalidität. Der Grund dafür ist, dass die Kriteriumsvalidität die Beziehung zwischen dem Resultat eines Untersuchungsverfahrens, also in diesem Fall eines ACs, und einem externen Kriterium, also in diesem Fall der berufliche Erfolg oder die berufliche Leistung, beschreibt (Rupp, Thornton, & Gibbons, 2008). Metaanalytische Studien haben ergeben, dass man aufgrund der in einem AC gewonnenen Daten die berufliche Leistung der Teilnehmer voraussagen kann und ACs somit kriteriumsvalid sind (Becker, Höft, Holzenkamp, & Spinath, 2011; Gaugler et al., 1987).

Aufgrund des hohen Stellenwerts der Kriteriumsvalidität von ACs in Forschung und Praxis hat adt zurich im Rahmen des Rezertifizierungsprozesses eine Validitätsstudie der eigenen Berichte durchgeführt. Ziel war es herauszufinden, ob und in welchem Ausmass die Berichte von adt zurich kriteriumsvalid sind.



Die Validitätsstudie zur Rezertifizierung: Wie wurde vorgegangen?

Stichprobe und Ablauf

Die Stichprobe setzte sich zusammen aus 64 Berichten, die in den Jahren 2013 bis 2015 erstellt wurden. Von diesen 64 Berichten konnten 45 beurteilt werden. 19 Berichte konnten nicht beurteilt werden, weil der AC-Teilnehmer die Stelle nicht bekommen hat oder bereits wieder aus dem Unternehmen ausgeschieden ist. Die untersuchte Gruppe setzte sich also zusammen aus 45 Personen, die bei adt zurich ein eintägiges Einzel-Assessment durchlaufen haben. Bedingung für die Beurteilung der AC-Teilnehmenden war, dass diese mindestens sechs Monate beim Auftraggeber angestellt waren oder immer noch angestellt sind.

Die Auftraggeber setzten sich sowohl aus öffentlich-rechtlichen als aus privatwirtschaftlichen Organisationen zusammen. Die Beurteilungen erfolgten entweder durch den Vorgesetzten des AC-Teilnehmenden oder durch einen Personalverantwortlichen.



Als Grundlage für die Beurteilung der AC-Teilnehmer diente ein für diese Evaluationsstudie konzipiertes strukturiertes Interview sowie die nach den absolvierten ACs erstellten Berichte von adt zurich. Die jeweiligen Interviews fanden in den Räumlichkeiten der einzelnen Auftraggeber statt und dauerten je nach Anzahl Beurteilungen zwischen dreissig und sechzig Minuten.

Messungen

Der Fokus der Evaluation wurde auf den 1:1-Vergleich des AC-Anforderungsprofils beziehungsweise der AC-Dimensionen gelegt. Der betreffende Vorgesetzte oder Personalverantwortliche beurteilte die im AC erhobenen Dimensionen auf einer Skala von „ungenügend“ bis „sehr ausgeprägt“, wobei „erfüllt Erwartungen“ (mittlere Ausprägung) die optimale Passung zur Funktion darstellt. Anhand dieser Daten wurden die Werte aus den einzelnen Dimensionen mit den von adt zurich beurteilten Dimensionen in Zusammenhang gebracht.

Es wurden immer alle im Bericht erwähnten Dimensionen bewertet. Die Korrelationen wurden aber aus folgenden Dimensionen errechnet, weil diese die gängigsten sind:

- Persönlichkeitsformat (Echtheit und Ausstrahlung/Authentizität, Integrität und Loyalität)
- Belastbarkeit, Leistungsorientierung
- Kommunikationsfähigkeit
- Kritik- und Konfliktfähigkeit
- Führungs- und Lenkungsfähigkeit
- Logisch-analytisches ganzheitliches und konzeptuelles Denken, Problemlösungsfähigkeit

Errechnet wurden die Produkt-Moment-Korrelationen r . Die Produkt-Moment-Korrelation r beschreibt den Zusammenhang zwischen zwei intervallskalierten Variablen (Bortz & Döring, 2013). In Bezug auf die Evaluation beschreibt die Produkt-Moment-Korrelation den Zusammenhang zwischen der Dimensionsbeurteilung seitens adt zurich und der Beurteilung des Auftraggebers. Die entsprechenden r 's sind in der Tabelle 2 weiter unten ersichtlich. Zudem wurde eine Gesamtbeurteilung des Berichts erfragt. Der Kunde musste einschätzen, inwiefern die Einschätzung des Kandidaten seitens adt zurich zutraf. Die Gesamtbeurteilung erfolgte auf einer Skala von „nicht zutreffend“ bis „ausgeprägt zutreffend“. Die entsprechenden Ergebnisse mit und ohne Prozentzahlen sind in den Abbildungen 1 und 2 weiter unten ersichtlich.



Die Ergebnisse der Validitätsstudie: adt zurich Berichte sind ausgesprochen treffsicher

Es wurden 45 Berichte beurteilt. 25 ACs wurden als Auswahl-Assessments, 17 ACs als Entwicklungs-Assessments und drei ACs als Potenzialanalysen eingesetzt. Von den 45 beurteilten Personen waren 37 Männer und acht Frauen. Das Durchschnittsalter der Kandidaten betrug 46.6 Jahre.

In der folgenden Tabelle befinden sich die berechneten Produkt-Moment-Korrelationen r .

Persönlichkeitsformat (Echtheit und Ausstrahlung/Authentizität, Integrität und Loyalität)	$N = 42, r = .744$
Belastbarkeit, Leistungsorientierung	$N = 30, r = .866$
Kommunikationsfähigkeit	$N = 45, r = .918$
Kritik- und Konfliktfähigkeit	$N = 41, r = .903$
Führungs- und Lenkungsfähigkeit	$N = 44, r = .728$
Logisch-analytisches ganzheitliches und konzeptuelles Denken, Problemlösungsfähigkeit	$N = 27, r = .940$

Tabelle 2: Produkt-Moment-Korrelationen r der einzelnen Dimensionen (1:1-Vergleich auf der Basis des AC-Anforderungsprofils resp. der AC-Dimensionen), N = Anzahl beurteilte Dimensionen



In den folgenden Abbildungen befinden sich die Gesamtbeurteilungen der Berichte:

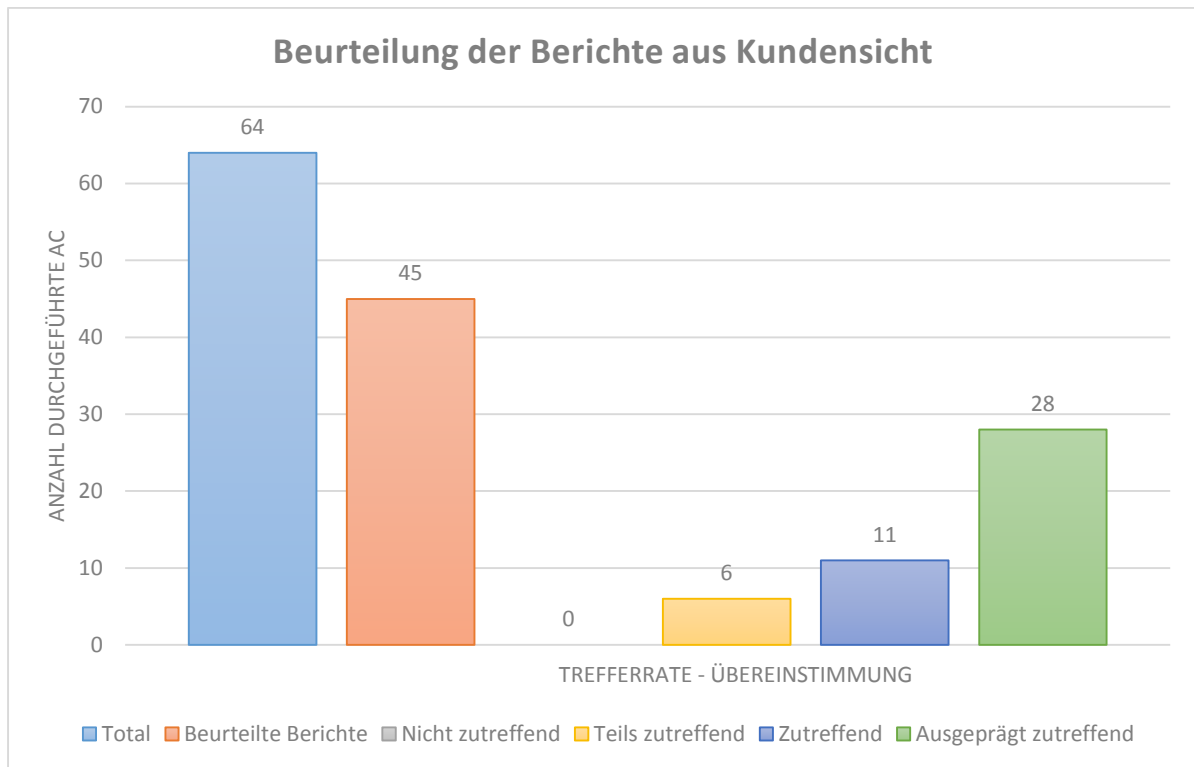


Abbildung 1: Gesamtbeurteilung zur Aussagekraft von 64 durchgeführten Assessments (Kundensicht).

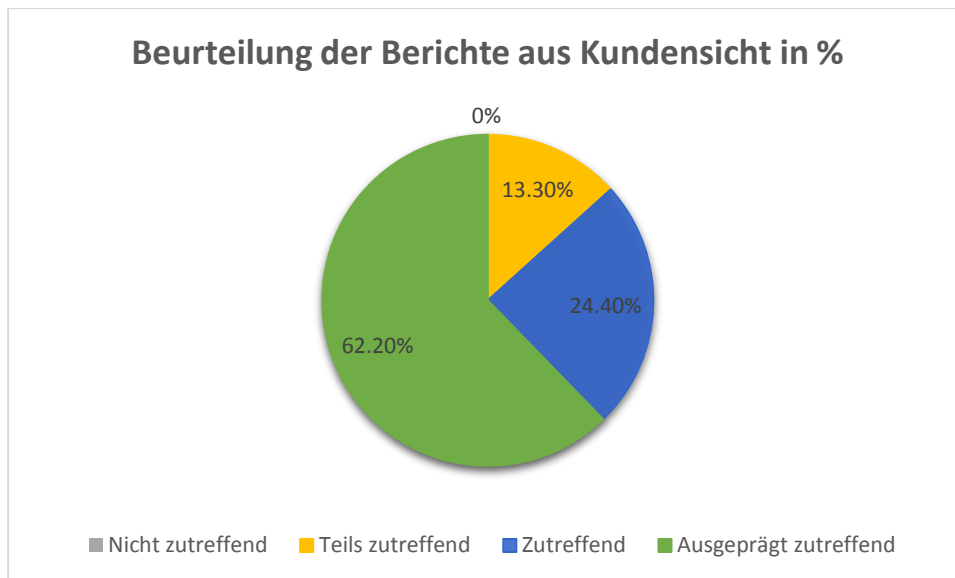


Abbildung 2: Gesamtbeurteilung zur Aussagekraft der 45 bewerteten Berichte in Prozent (Kundensicht).



Zusammenfassung

Die Korrelationen sind mit Werten von .728 bis .940 sehr hoch. Das zeigt, dass adt zurich die Kandidaten sehr treffsicher beurteilt und dass die von adt zurich eingesetzten ACs sehr valide sind. Auch die Werte aus den Abbildungen 1 und 2 mit den Gesamtbeurteilungen zeigen deutlich, dass die Bewertungen der Kandidaten seitens adt zurich grösstenteils richtig liegen. Knapp zwei Drittel der Beurteilungen treffen ausgeprägt zu, ein Viertel der Beurteilungen treffen zu und ein Siebtel der Beurteilungen treffen teilweise zu. Kein einziger Bericht wurde als „nicht zutreffend“ eingestuft.

Es kann verschiedene Gründe geben, weshalb die Berichte nicht immer ausgeprägt zutreffend sind. Einerseits sind die Beurteilungen aus den ACs eine Momentaufnahme und es ist natürlich möglich, dass sich der eine oder andere Kandidat nicht wie im beruflichen Alltag verhalten hat. Andererseits könnte es sein, dass sich die AC-Teilnehmer seit dem AC weiterentwickelt haben und bspw. Coachings oder andere Weiterbildungskurse besucht haben, die zu einer Verhaltensänderung geführt haben könnten.

Fazit: Das AC als Verfahren zur Auswahl und Entwicklung von Führungskräften hat sich als Methode bewährt. adt zurich setzt sehr valide ACs ein und aufgrund der Befolgung der Prinzipien und der Anwendung von unterschiedlichen und messgenauen Instrumenten sowie massgeschneiderten Übungen können die AC-Teilnehmer sehr präzise bewertet werden.

Alessandro Baia, Wissenschaftlicher Mitarbeiter adt zurich

Literatur

- 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. doi:10.1027/1866-5888/a000031
- Brown, D. J., & Keeping, L. M. (2005). Elaborating the construct of transformational leadership: The role of affect. *The Leadership Quarterly, 16*, 245-272. doi:10.1016/j.leaqua.2005.01.003
- Bortz, J., & Döring, N. (2013). *Forschungsmethoden und Evaluation*. Heidelberg: Springer-Verlag.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3. Aufl.). München: Pearson Studium Deutschland GmbH.
- Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology, 72*, 493-511. doi:10.1037/0021-9010.72.3.493
- International Task Force on Assessment Center Guidelines (2015). Guidelines and ethical considerations for assessment center operations. *Journal of Management, 41*, 1244-1273. doi:10.1177/0149206314567780
- Lienert, G. A., & Ratz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: Beltz Psychologie Verlagsunion.
- Munchus III, G., & McArthur, B. (1991). Revisiting the historical use of the assessment centre in management selection and development. *Journal of Management Development, 10*, 5-13. doi:10.1108/02621719110003999
- Obermann, C. (2013). *Assessment Center* (5. Aufl.). Wiesbaden: Springer. doi:10.1007/978-3-8349-3813-8
- Ones, D. S., & Viswesvaran, C. (1998). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human Performance, 11*, 245-269. doi:10.1207/s15327043hup1102&3_7
- Rupp, D. E., Thornton, G. I., & Gibbons, A. M. (2008). The construct validity of the assessment center method and usefulness of dimensions as focal constructs. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 116-120. doi:10.1111/j.1754-9434.2007.00021.x
- Thornton, G. C., & Gibbons, A. M. (2009). Validity of assessment centers for personnel selection. *Human Resource Management Review, 19*, 169-187. doi:10.1016/j.hmr.2009.02.002