

„Is it all about intelligence?“

Kognitive Fähigkeitstests und Assessment Center: Ein neuer Blick auf eine kontrovers diskutierte Befundlage

Diana E. Krause und Martin Kersting

1 Einleitung

Heute gilt als unstrittig, dass sowohl kognitive Fähigkeitstests als auch Assessment-Center-Ratings hilfreiche Prädiktoren verschiedener Leistungskriterien im beruflichen Kontext sind. Strittig ist dagegen die Frage, ob Assessment Center (AC) die Prognosegüte verbessern, wenn man sie zusätzlich zu kognitiven Fähigkeitstests einsetzt. Wir werden nachfolgend die heterogene empirische Befundlage zu dieser Frage vergleichend darstellen und daraus die Kernthese ableiten, dass AC eine eigenständige leistungsbezogene Vorhersagekraft aufweisen, wenn sie zusätzlich zu kognitiven Fähigkeitstests Anwendung finden. Diese These überprüfen wir erstmals anhand einer Stichprobe des höheren Managements. Die Ergebnisse verweisen auf inkrementelle, prädiktive Validität von kognitiven Fähigkeitstests und AC für Top-Managementpositionen, so dass wir unsere Kernfrage „Is it all about intelligence?“ tendenziell mit: „nein“ beantworten.

2 Evidenz für den Zusammenhang zwischen AC-Urteilen und Kriterien

Eine Vielzahl von Studien dokumentiert, dass AC-Gesamturteile („Overall Assessment Ratings“, OAR) verschiedene Kriterien – wie etwa Gehaltsverbesserungen, beruflichen Aufstieg, Ausbildungserfolg, Arbeitsleistung – für ein breites Spektrum an Berufen und Organisationen valide prognostizieren (z. B. Lievens & Thornton, 2005; Thornton & Rupp, 2006). Die Ergebnisse der Metaanalysen für die Beziehung zwischen AC-Urteilen und berufsbezogenen Ergebnissen variieren erheblich, nämlich zwischen $r = .41$ (Schmitt, Gooding, Noe & Kirsch, 1984), $r = .37$ (Gaugler, Rosenthal, Thornton & Bentson, 1987), $r = .25$ bis $r = .39$ (Arthur, Day, McNelly & Edens, 2003), $r = .22$ (Aamödt, 2004) und $r = .31$ (Hardison & Sackett, 2007). Diese Unterschiede in den ermittelten Validitätsschätzungen lassen sich durch die Berücksichtigung unterschiedlicher Einzelstudien pro Metaanalyse, Unterschiede in der Qualität des AC (vgl. dazu Krause & Gebert, 2003a, b; Krause, Gebert & Thornton, 2007; Krause & Thornton, 2007) und die Anwendung divergierender metaanalytischer Verfahren erklären. Lievens und Thornton (2005) argumentieren, dass der vielzitierte Validitätskoeffizient ($r = .37$) nach Gaugler et al. (1987) möglicherweise eine Unterschätzung der Validität des AC darstellt, da sie eine vergleichsweise konservative (also hohe) Schätzung der

Reliabilität des Kriteriums (.86) in ihren Korrekturen bezüglich der Unreliabilität des Kriteriums vorgenommen haben. Sofern die Reliabilität des Kriteriums auf .52 geschätzt würde (Viswesvaran, Ones & Schmidt, 1996), würde die geschätzte Validität des AC von $r = .37$ auf $r = .47$ steigen.

3 Kriteriumsbezogene Validität von kognitiven Fähigkeitstests

Wenngleich Befunde dafür vorliegen, dass AC-Urteile in einem statistisch signifikanten und praktisch relevanten Ausmaß mit Leistungskriterien korrelieren, fallen diese Korrelationen generell niedriger aus als die Korrelationen zwischen kognitiven Fähigkeitstests und berufsbezogenen Kriterien. In diesem Zusammenhang wird häufig auf die Metaanalyse von Schmidt und Hunter (1998) verwiesen, in der die Korrelation zwischen kognitiven Fähigkeitstests und beruflicher Leistung ($r = .51$) höher beziffert wurde als die Korrelation zwischen AC-Urteilen und Arbeitsleistung ($r = .37$). Für Manager und anspruchsvolle technische Berufe fiel die Prognosegüte dabei höher aus als für Berufe mit geringer Komplexität. Für den europäischen Bereich berichten Salgado, Anderson, Moscoso, Bertua und De Fruyt (2003) eine Kriteriumsvalidität für kognitive Fähigkeitstests in Höhe von $r = .62$. Pynes und Bernadin (1989) zeigen an einer Stichprobe von Kandidaten für den mittleren Polizeivollzugsdienst, dass Scores kognitiver Fähigkeitstests höher mit dem Ausbildungserfolg ($r = .31$) korrelieren als AC-Urteile ($r = .14$). Angesichts der höheren kriteriumsbezogenen Validität von kognitiven Fähigkeitstests und des Umstands, dass diese im Vergleich zu ACn schlichtweg geringere Kosten implizieren, kann es aus der Validitätsperspektive nur einen Grund geben, AC zusätzlich zu kognitiven Fähigkeitstests einzusetzen: AC-Urteile müssten die Prognosegüte entscheidend (inkrementell) verbessern, wenn sie zusätzlich zu kognitiven Fähigkeitstests Anwendung finden sollen.

4 Inkrementelle Validität von Assessment Center-Urteilen

Schon vor 30 Jahren sahen sich Klimoski und Brickner (1987) zu dem Argument veranlasst, dass AC zwar funktionieren, letztlich aber wenig mehr als Intelligenz messen. Heute fragen wir erneut: „Is it all about intelligence?“ Dabei gehen wir davon aus, dass dieses Bedingungsgefüge von der Komplexität der Leistungsdomäne und der Beziehung zwischen AC-Urteilen und kognitiven Fähigkeitstests abhängt. Der Leistungsbereich vieler Berufe ist als komplex zu bezeichnen (Campbell, McCloy, Oppler & Sager, 1993); dies gilt selbstverständlich insbesondere für Managementtätigkeiten. Aus diesem Grund ist es plausibel, zu erwarten, dass nicht nur ein Konstrukt, sondern verschiedene Konstrukte einen eigenständigen Beitrag zur Leistungsprognose leisten können.

Während Schmidt und Hunter (1998) keine inkrementelle Validität von AC-Ratings gegenüber kognitiven Fähigkeitstests erwarten, fanden Dayan, Kasten und Fox (2002) gegenteilige empirische Evidenz: An einer Stichprobe israelischer Anwärter für den Polizeivollzugsdienst zeigten die Autoren, dass AC-Ratings in der Tat die Prognosegüte steigern, wenn sie zusätzlich zu kognitiven Fähigkeitstests eingesetzt werden.

Das OAR war in dieser Studie signifikant mit verschiedenen Ausbildungskriterien korreliert, wie dem Ausbildungserfolg ($r = .34$), den mittleren Dimensionsurteilen durch Peers ($r = .14$), dem zukünftigen Berufserfolg ($r = .43$), der Arbeitsleistung gemessen durch Vorgesetztenurteile ($r = .25$) und regelmäßigen Evaluationen durch Vorgesetzte ($r = .24$). Ferner wies das OAR signifikante inkrementelle Validität gegenüber allgemeiner Intelligenz auf.

Mit Blick auf unsere Frage ist auch die Beziehung zwischen AC-Urteilen und kognitiven Fähigkeiten genauer zu betrachten. Schmidt und Hunter (1998) schätzen die korrigierte Korrelation zwischen AC-Ratings und Intelligenztests auf $r = .50$. Eine Metaanalyse von Scholz und Schuler (1993) verweist indes auf eine Korrelation von $r = .33$ zwischen allgemeiner Intelligenz und AC-Leistung. Andere metaanalytische Ergebnisse zeigen ebenfalls eine enge Beziehung zwischen dem OAR und kognitiven Fähigkeiten in Höhe von $r = .67$ (Collins, Schmidt, Sanchez-Ku, Thomas, McDaniel & Le, 2003). Aufgrund derartiger Befunde argumentieren Schmidt und Hunter (1998), dass AC-Urteile mit hoher Wahrscheinlichkeit keine inkrementelle Validität aufweisen ($r = .02$), sofern diagnostische Information aus kognitiven Fähigkeitstests verfügbar ist.

Wir möchten mit der nachfolgend dargestellten Studie diese Fragestellung in einem neuen Kontext analysieren: Es werden Top-Managementpositionen betrachtet, bei denen für den Berufserfolg mehrere Konstrukte relevant sein dürften, z. B. kognitive Fähigkeiten und Führungsqualitäten. Berufliche Leistung, insbesondere aber die Management- und Führungsleistung (vgl. dazu Krause, 2004), sind keine eindimensionalen Konstrukte (Campbell et al., 1993; Sarges, 2006). Der Wert des AC für das Verständnis und die Prognose der Führungsleistung in Kombination mit kognitiven Fähigkeitstests dürften von der Art der betrachteten Leistungsdomäne abhängig sein. Eine Veranschaulichung kann an dieser Stelle hilfreich sein: Kognitive Fähigkeitstests dürften so genannte „can do“-Aspekte der Leistungsdomäne vorhersagen, während AC-Urteile so genannte „will do“-Aspekte der Leistungsdomäne prognostizieren. Insofern dürften AC-Urteile andere Leistungsaspekte messen als kognitive Fähigkeitstests. Dies sollte insbesondere deutlich werden, wenn auf Seiten der Kriterien Kontextaspekte (Motowidlo & Van Scotter, 1994) berücksichtigt werden. AC-basierte Urteile sollten daher im Bereich von Managementtätigkeiten inkrementelle, prädiktive Validität gegenüber kognitiven Fähigkeitstests aufweisen, da der Berufserfolg hier nicht nur eine Funktion der kognitiven Fähigkeiten des Managers ist, sondern von unterschiedlichen Fähigkeiten und Eigenschaftsausprägungen abhängt. In Bezug auf die kriteriumsbezogene Validität des AC fragen wir deshalb: Weisen AC-Urteile (OAR) für die Prognose des Ausbildungserfolgs von Top-Managern der Polizei einen Zusatznutzen auf, wenn sie kombiniert mit kognitiven Fähigkeitstests eingesetzt werden?

5 Methode

Unsere Stichprobe basiert auf 91 männlichen Personen aus einer Gruppe von insgesamt ca. 700 Führungskräften, die sich für den höheren Polizeivollzugsdienst beworben hatten. Angehörige des höheren Polizeivollzugsdienstes besetzen Top-Management-

positionen der Polizei in verschiedenen deutschen Bundesländern. Grundlage der Entscheidung über die Beförderung in diese Top-Managementebene (zunächst: Zulassung zur Polizei-Führungsakademie, PFA) waren die Ergebnisse eines zweitägigen AC sowie eines kognitiven Fähigkeitstests. Von den ca. 700 Personen im Auswahlverfahren wurden 112 Kandidaten zur PFA zugelassen. Daten zum Ergebnis im kognitiven Fähigkeitstest stehen lediglich für 91 Personen zur Verfügung, weshalb sich unsere Analysen auf diese 91 Bewerber für das Top-Management der Polizei beziehen. Diese Personen waren zwischen 27 und 43 Jahre alt ($M = 33$, $SD = 3$ Jahre und 6 Monate).

Für ein besseres Verständnis sollen nachfolgend einige der potenziellen Tätigkeiten genauer skizziert werden, die den Personen nach erfolgreicher Bewältigung der Polizei-Führungstätigkeit offen stehen. Der Polizeivollzugsdienst gliedert sich in die mittlere, gehobene und höhere Laufbahn, wobei nur ein äußerst geringer Anteil (ca. 2 %) der Beamten in den höheren Dienst gelangt. Die Beamten des höheren Dienstes werden überwiegend aus Mitarbeitern des gehobenen Dienstes rekrutiert, die Auswahl folgt dem Prinzip der Bestenauslese. Zu Beginn der Laufbahn des höheren Dienstes steht in der Regel eine zweijährige Ausbildung, ein Jahr davon als Studium an der PFA. Mit der Auswahl wird somit zunächst über die Zulassung zur Aufstiegsausbildung entschieden, aber letztendlich ist die Entscheidung auch eine Voraussetzung für die Zulassung zur Polizeiführung. Es ist deutlich darauf hinzuweisen, dass die Führungskräfte der Polizei des höheren Dienstes mit Führungskräften in For-Profit-Organisationen vergleichbar sind. Dies gilt in Bezug auf die Bewältigung komplexer Aufgaben (z. B. das Management der polizeilichen Reaktionen auf eine Geiselnahme oder die Sicherung der Castor-Transporte), die Leitungsspanne (Angehörige des höheren Polizeivollzugsdienstes führen mehrere hundert bis mehrere tausend unterstellte Mitarbeiter) und die finanzielle Verantwortung (Entscheidungen, die mehrere Millionen EUR umfassen können).

Zur Diagnose von verschiedenen Facetten kognitiver Fähigkeiten sowie von Wissensdomänen wurde ein psychometrischer Eignungstest der Deutschen Gesellschaft für Personalwesen e. V. (DGP, vgl. www.dgp.de) eingesetzt. In Anlehnung an Thurstone und Thurstone (1946) beinhaltet der Test mehrere Aufgabengruppen zur Messung (1) verbaler und (2) numerischer Fähigkeiten sowie (3) der Wahrnehmungsgeschwindigkeit/des Arbeitstempos. In Anlehnung an Cattells Konzept der kristallinen Intelligenz kamen ferner (4) Aufgabengruppen mit Kenntnisfragen zu verschiedenen Wissensdomänen zum Einsatz (für statistische Angaben zur Qualität der Aufgabengruppen und zur faktoriellen Struktur vgl. Kersting, 1994, S. 53, 1999, S. 182). Faktorenanalysen zeigten, dass die vier kognitiven Fähigkeiten auch auf der empirischen Ebene unterscheidbar sind; jedoch kann außerdem ein G-Faktor kognitiver Fähigkeiten identifiziert werden. Daher haben wir durch Durchschnittsbildung der Standardscores in jedem Test eine Variable – kognitive Fähigkeiten – gebildet, wobei jede der vier Fähigkeiten mit dem gleichen Gewicht repräsentiert wurde.

Das AC wurde von der DGP spezifisch für den höheren Polizeivollzugsdienst konzipiert und durchgeführt. Das Anforderungsprofil sah (zusätzlich zu den kognitiven Fähigkei-

ten) folgende acht Dimensionen vor: Kommunikationsfähigkeit, soziale Kompetenz, Sicherheit/Belastbarkeit, Zielorientierung, Sachbezogenheit, Aktivität/Dynamik, Einfallsreichtum/Flexibilität, Führungsverhalten und Motivation. Diese Dimensionen wurden durch vier Übungen im AC erfasst: Einzelvorstellung (eignungsdiagnostisches Interview), Kurzreferat (Präsentation), führerlose Gruppendiskussion und Presseschau/ Problemlöseszenario (für Details zu den Übungen und zur Übungs-Dimensions-Matrix vgl. Kersting, 2003; Krause, Kersting, Heggestad & Thornton, 2006).

Aus dem AC wurden drei Urteilsarten abgeleitet. Erstens wurde zusätzlich zu den Einzelurteilen (pro Dimension und Übung) der Beobachter ein abschließendes Dimensionsurteil für jeden Kandidaten gebildet. Das Dimensionsurteil reflektiert den Mittelwert der Konsensurteile einer Dimension (z. B. Führungsverhalten) über alle Übungen hinweg, in denen die jeweilige Dimension beurteilt wurde. Zweitens wurde ein Übungsurteil errechnet, das den Mittelwert aller Konsensurteile über verschiedene Dimensionen hinweg innerhalb jeder Übung (z. B. führerlose Gruppendiskussion) repräsentiert. Drittens wurde durch den Mittelwert aller Konsensurteile über alle Dimensionen hinweg ein AC-Gesamturteil für jeden Kandidaten gebildet. Grundlage der Personalentscheidung ist das abschließende Verfahrensurteil (OAR). Dieser Wert wird auf der Basis der Testleistungen einerseits und des AC-Gesamturteils andererseits bestimmt.

Das in dieser Studie genutzte Kriterium war die Abschlussnote bei der PFA. Die Abschlussnote setzt sich aus diversen Leistungstests zusammen, die während der zweijährigen Ausbildung an der PFA zu erbringen sind. Diese Abschlussnote reflektiert drei Leistungsarten: individuelle Leistungen während der zweijährigen Ausbildung (Gewichtung 30 %), einen themenspezifischen schriftlichen Teil (Gewichtung 45 %) sowie die mündlichen Abschlussprüfungen (Gewichtung 25 %). Im Hinblick auf die Taxonomie von Campbell et al. (1993) lässt sich diese Form der Leistungsprüfung als „task performance“ bezeichnen. Die metaanalytisch berichtete Reliabilität derartiger Prüfungen liegt bei $r = .52$ (Viswesvaran et al., 1996). Ferner reflektiert diese Form der Prüfung auch eine spezifische Wissensform, nämlich deklaratives Wissen (für genauere Informationen vgl. Kersting, 2003).

6 Ergebnisse

Tabelle 1 stellt die Interkorrelationen der Prädiktoren sowie die Korrelationen der Prädiktoren mit dem Erfolg bei der PFA zusammenfassend dar. Wie es in der Forschung zu kognitiven Fähigkeiten typisch ist (Ackerman, Beier & Boyle, 2005), variieren auch hier die Korrelationen zwischen den kognitiven Fähigkeiten erheblich ($r = .27$ bis $.62$). Übereinstimmend mit der AC-Forschung (Arthur et al., 2003; Chan, 1996) zeigen die Ergebnisse ferner signifikante Korrelationen zwischen den AC-Übungen ($r = .37$ bis $.72$) sowie den AC-Dimensionen ($r = .33$ bis $.83$).

Mit Blick auf unsere Fragestellung sind insbesondere die Korrelationen zwischen den Prädiktoren (kognitive Fähigkeiten, AC-Übungen, AC-Dimensionen) und dem Erfolg bei der PFA interessant (vgl. Tab. 1, Zeile 19). Von einer Ausnahme abgesehen, sind

Tabelle 1: Korrelationen der Prädiktoren (Kognitive Fähigkeiten, AC-Übungen, AC-Dimensionen) und Korrelationen der Prädiktoren mit dem Erfolg bei der Führungskräfteakademie der Polizei

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
1 Kognitive Fähigkeiten																			
2 Verbale Fähigkeiten	-.87***	-																	
3 Numerische Fähigkeiten	.79***	.60***	-																
4 Wahrnehmungsgeschwindigkeit/Arbeitstempo	.82***	.62***	.59***	-															
5 Wissen	.64***	.50***	.27**	.30**	-														
AC-Übungen																			
6 Einzelvorstellung	.09	.17*	-.10	.06	.17*	-													
7 Kurzreferat	.35***	.34***	.20*	.22*	.35***	.39***	-												
8 Führerlose Gruppendiskussion	.06	.17*	-.13	.07	.09	.67***	.44***	-											
9 Presseschau/Problemszenario	.02	.12	-.08	.00	.04	.51***	.37***	.72**	-										
AC-Dimensionen																			
10 Kommunikationsfähigkeit	.75***	.34***	.04	.11	.29**	.76***	.72***	.62***	.45***	-									
11 Soziale Kompetenz	.14	.19*	-.02	.13	.14	.55***	.38***	.80***	.78***	.48***	-								
12 Sicherheit/Belastbarkeit	-.00	.09	-.20*	.01	.10	.71***	.55***	.76***	.73***	.71***	.66***	-							
13 Zielorientierung/Sachbezogenheit	.16	.23**	-.04	.13	.17*	.76***	.70***	.70***	.64***	.83***	.64***	.73**	-						
14 Aktivität/Dynamik	.07	.06	.01	.04	.11	.44***	.34***	.58***	.73***	.39***	.55***	.56***	.52***	-					
15 Einfallsreichtum/Flexibilität	.13	.20*	-.03	.10	.13	.51***	.64***	.71***	.66***	.59***	.59***	.62***	.70***	.58***	-				
16 Führungsverhalten	.11	.19*	-.04	.06	.14	.65***	.34***	.76***	.80***	.51***	.77***	.72***	.64***	.67***	.63***	-			
17 Motivation	.01	.09	-.12	.02	.03	.85***	.29**	.54***	.40***	.63***	.42***	.60***	.61***	.33**	.42***	.52***	-		
18 AC Gesamtwertung	.19	.33	-.06	.06	.21	.80***	.36***	.60***	.51***	.75***	.57***	.88***	.64***	.71***	.65***	.62***	.52***	-	
19 Erfolg bei der PFA	.35	.30	.20**	.30	.50**	.63***	.31	.50***	.47***	.71***	.53***	.81***	.51***	.50***	.60***	.52***	.49***	.38***	-

Anmerkungen: Korrelationen nach Pearson. Einseitige Signifikanz. PFA = Polizei Führungskräfteakademie. N = 91; * p < .05. ** p < .01. *** p < .001.

alle anderen Prädiktoren signifikant positiv mit dem Erfolg verbunden. Für die kognitiven Fähigkeiten gilt, dass verbale Fähigkeiten ($r = .50$) und Wahrnehmungsgeschwindigkeit/Arbeitstempo ($r = .50$) am höchsten mit dem Erfolg bei der PFA korrelieren. Der Vergleich dieser Korrelationen mit einer früheren Untersuchung für Berufseinsteiger der Polizei (Pynes & Bernadin, 1989) zeigt, dass die Korrelationen zwischen kognitiven Fähigkeiten und dem Kriterium für die Top-Managementebene der Polizei höher ausfallen. Signifikante positive Korrelationen können auch für die vier AC-Übungen und den Erfolg bei der PFA gezeigt werden. Die Korrelationen variieren zwischen $r = .19$ und $r = .46$. Auch die AC-Dimensionen (mit der Ausnahme Motivation) sind signifikant positiv mit dem Erfolg bei der PFA korreliert: Die Korrelation zwischen Zielorientierung/Sachbezogenheit und dem Erfolg bei der PFA fällt im Vergleich zu allen anderen AC-Dimensionen am höchsten aus ($r = .41$). Zusammenfassend bestätigen die Korrelationen die Funktionalität kognitiver Fähigkeiten, AC-Übungen und AC-Dimensionen für den Erfolg bei der PFA.

Um unsere Fragestellung zu prüfen, haben wir eine hierarchische Regressionsanalyse gerechnet. Im ersten Schritt wurden die kognitiven Fähigkeiten in den Regressionsansatz aufgenommen, im zweiten Schritt wurde zu den kognitiven Fähigkeiten das OAR hinzugefügt, wobei der Erfolg bei der PFA jeweils die Kriteriumsvariable darstellte. Es wurde geprüft, ob allein die kognitiven Fähigkeiten für den Erfolg maßgeblich sind oder ob die zusätzliche Berücksichtigung des OAR einen inkrementellen Varianzbeitrag zur Erfolgsvorhersage leistet. Der Zuwachs an Varianzaufklärung im Kriterium „Erfolg bei der PFA“ durch den zweiten Regressionsschritt stellt die kritische Prüfgröße für unsere zentrale Annahme dar, dass der Erfolg durch kognitive Fähigkeiten *und* das AC-Gesamturteil vorhergesagt werden kann.

Es zeigte sich, dass die kognitiven Fähigkeiten ein signifikanter Erfolgsprädiktor im ersten Schritt sind ($R = .53$, $F = 34.36$, $p < .001$, $df_1 = 1$, $df_2 = 89$, $\beta_{\text{KognFäh}} = .53$, $N = 91$). Sie erklären 28 % der Erfolgsvarianz. Das Hinzufügen des AC-Gesamturteils im zweiten Regressionsschritt führt zu einem inkrementellen Varianzanstieg im Erfolgskriterium; das OAR erklärt zusätzliche 5 % der Erfolgsvarianz ($R = .57$, $F = 21.47$, $p < .001$, $df_1 = 2$, $df_2 = 88$, $\beta_{\text{KognFäh}} = .50$, $\beta_{\text{OAR}} = .23$, $N = 91$). Die Ergebnisse dieser Analyse bestätigen, dass AC-Gesamturteile inkrementelle Validität zusätzlich zu kognitiven Fähigkeitstests aufweisen. Ferner haben wir eine zweite hierarchische Regression berechnet, in der die Prädiktoren in umgekehrter Reihenfolge Eingang in den Regressionsansatz fanden (Schritt 1: OAR, Schritt 2: kognitive Fähigkeiten). Die Ergebnisse dieser Regressionsanalyse sind mit den Ergebnissen der ersten Regressionsanalyse vergleichbar: Sowohl kognitive Fähigkeiten als auch das OAR prognostizieren den Erfolg bei der PFA.

Darüber hinaus haben wir gefragt, ob die inkrementelle Validitätssteigerung des AC auch dann gegeben ist, wenn die kognitiven Fähigkeiten in separater Weise in den Regressionsansatz aufgenommen werden. Aus diesem Grund haben wir eine weitere hierarchische Regressionsanalyse gerechnet, in der die kognitiven Fähigkeiten einzeln im ersten Schritt in die Regression aufgenommen wurden ($R = .56$, $F = 9.86$, $p < .001$, $df_1 = 4$, $df_2 = 86$, $\beta_{\text{VerbaleFäh}} = .28$, $\beta_{\text{NumerFäh}} = .05$, $\beta_{\text{Wahrneh}} = .34$, $\beta_{\text{Wissen}} = .08$, $N = 91$)

und im zweiten Schritt zusätzlich zu den einzelnen kognitiven Fähigkeiten das OAR hinzugefügt wurde. Es zeigt sich, dass die Berücksichtigung des OAR zusätzlich zu den einzelnen kognitiven Fähigkeiten zu einem inkrementellen, prädiktiven Validitätszuwachs im Hinblick auf die Erfolgsprognose in Höhe von 4 % führt ($R = .60$, $F = 9.38$, $p < .001$, $df_1 = 5$, $df_2 = 85$, $\beta_{\text{VerbaleFäh}} = .22$, $\beta_{\text{NumerFäh}} = .00$, $\beta_{\text{Wahrneh}} = .34$, $\beta_{\text{Wissen}} = .05$, $\beta_{\text{OAR}} = .21$, $N = 91$). Vor dem Hintergrund der berichteten Ergebnisse wollen wir unsere eingangs gestellte Frage „Is it all about intelligence“ verneinen: Zusätzlich zu kognitiven Fähigkeiten sind AC-Urteile für die höhere Managementebene der Polizei erfolgsrelevant.

7 Diskussion

Unsere Studie erweitert die bisherige AC-Forschung in dreierlei Hinsicht: Erstens bieten wir empirische Evidenz für die inkrementelle Validität von kognitiven Fähigkeitstests und AC-Urteilen auf der Top-Managementebene der Polizei. Dies ist zu betonen, denn die bisherige AC-Forschung gewann ihre Erkenntnisse anhand von Berufseinsteigerstichproben oder niedrigen und mittleren Managementpositionen (Chan, 1996; Dayan et al., 2002; Pynes & Bernadin, 1989). Zweitens belegen unsere Ergebnisse, dass kognitive Fähigkeiten auch für höhere Managementebenen bedeutsame Erfolgsprädiktoren sind. Nun zeigen aber Studien, dass kognitive Fähigkeitstest häufig nicht Bestandteil von AC sind (Thornton & Krause, 2007). Insofern wäre es für die zukünftige Personalpraxis empfehlenswert, mehr als bislang kognitive Fähigkeitstests im AC einzusetzen. Diese Argumentation steht im Einklang mit Ergebnissen, die den Nutzen kognitiver Fähigkeitstests betonen (Kuncel & Hezlett, 2007). Drittens verdeutlichen die Ergebnisse, dass AC einmalige Attribute messen können, wenn sie zusätzlich zu kognitiven Fähigkeitstests Anwendung finden. Bisherige Ergebnisse zu dieser Frage fallen widersprüchlich aus: Während einige Studien den prädiktiven Zusatznutzen von AC belegen (Dayan et al., 2002), verweisen andere Studien auf das Gegenteil (Schmidt & Hunter, 1998). Nach unseren Ergebnissen bietet das AC inkrementelle, prädiktive Validität für den Erfolg bei der PFA, wenn es in Kombination mit kognitiven Fähigkeitstests eingesetzt wird. Dieser Zusatznutzen legitimiert unserer Interpretation nach den Einsatz des kostenintensiven AC für Personalauswahl- und Personalbeförderungentscheidungen. Ferner weisen auch Kosten-Nutzen-Analysen (vgl. Kersting, 2005) darauf hin, dass sich der kombinierte Einsatz von kognitiven Fähigkeitstests und AC auszahlt.

Neben ihren Vorteilen ist unsere Untersuchung auch mit Problemen behaftet. Ein Problem manifestiert sich in Streuungseinschränkungen auf der Prädiktorseite. Es war weder möglich, das Ausmaß der Streuungseinschränkungen zu bestimmen, noch zu korrigieren, da die hierzu erforderlichen Rohdaten aus Gründen der Vertraulichkeit nicht zur Verfügung standen. Ein zweites Problem besteht darin, dass wir aufgrund des Nichtverfügens über die Rohdaten auch keine MTMM-Analysen rechnen konnten und daher keine Evidenz für die Konstruktvalidität der AC-Dimensionen bieten können. Ein drittes Problem ist mit unserer Kriteriumsmessung verbunden. Als Kriterium

stand uns lediglich der Ausbildungserfolg über die PFA, nicht aber der Berufserfolg bzw. die Arbeitsleistung im Beruf zur Verfügung. Allerdings bleibt diesbezüglich zu erwähnen, dass der Ausbildungserfolg eine notwendige Bedingung für die spätere Leistung „on the job“ ist. Ausbildungs- und Berufserfolg sind bedeutsam positiv korreliert (Dayan et al., 2002; Marcus, Goffin, Johnston & Rothstein, in press).

Literatur

- Aamodt, M. G. (2004). *Research in law enforcement selection*. Boca Raton, FL: Brown Walker.
- Ackerman, P. L., Beier, M. E. & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, 131, 30–60.
- Arthur Jr., W. A., Day, E. A., McNelly, T. L. & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology*, 56, 125–149.
- Campbell, J. P., McCloy, R. A., Oppler, S. H. & Sager, C. E. (1993). A theory of performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 35–70). San Francisco, CA: Jossey-Bass.
- Cattell, R. B. (1987). *Intelligence: Its structure, growth, and action*. Amsterdam: Elsevier.
- Chan, D. (1996). Criterion and construct validation of an assessment centre. *Journal of Occupational and Organizational Psychology*, 69, 167–181.
- Collins, J. M., Schmidt, F. L., Sanchez-Ku, M., Thomas, L., McDaniel, M. A. & Le, H. (2003). Can basic individual differences shed light on the construct meaning of assessment center evaluations? *International Journal of Selection and Assessment*, 11, 17–29.
- Dayan, K., Kasten, R. & Fox, S. (2002). Entry-level police candidate assessment center: An efficient tool or a hammer to kill a fly? *Personnel Psychology*, 55, 827–849.
- Gaugler, B. B., Rosenthal, D. B., Thornton, G. C. III & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology*, 72, 493–511.
- Hardison, C. M. & Sackett, P. R. (2007). Kriteriumsbezogene Validität des Assessment Centers: Lebendig und wohlauf? In H. Schuler (Hrsg.), *Assessment Center zur Potenzialanalyse* (S. 203–213). Göttingen: Hogrefe.
- Kersting, M. (1994). Personalauswahl in den neuen Bundesländern: Test ohne Grenzen oder Grenzen der Tests? Eine vergleichende Analyse der durchschnittlichen Testergebnisse in Ost und West (Teil II). *DGP-Informationen*, 53, 47–98.
- Kersting, M. (1999). *Diagnostik und Personalauswahl mit computergestützten Problemlösenszenarien? Eine Erörterung und ein empirischer Vergleich der Kriteriumsvalidität von Problemlösenszenarien und Intelligenztests*. Göttingen: Hogrefe.
- Kersting, M. (2003). Assessment Center: Erfolgsmessung und Qualitätskontrolle. In S. Höft & B. Wolf (Hrsg.), *Qualitätsstandards für Personalentwicklung in Wirtschaft und Verwaltung* (S. 72–93). Hamburg: Windmühle.
- Kersting, M. (2005). Profit durch Personalauswahl – warum sich eine qualitativ hochwertige Personalauswahl langfristig rechnet. In A. Gourmelon, C. Kirbach & S. Etzel (Hrsg.), *Personalauswahl im öffentlichen Sektor* (S. 15–34). Baden-Baden: Nomos.
- Krause, D. E. (2004). Influenced-based leadership as a determinant of the inclination to innovate and of innovation-related behaviors – An empirical investigation. *Leadership Quarterly*, 15, 79–102.
- Krause, D. E. & Gebert, D. (2003a). A comparison of assessment center practices in organizations in German-speaking regions and the United States. *International Journal of Selection and Assessment*, 11, 297–312.

- Krause, D. E. & Gebert, D. (2003b). Unterschiede in der Durchführung zwischen internen und externen Assessment Centern – Eine empirische Analyse. *Zeitschrift für Arbeits- und Organisationspsychologie*, 47, 87–94.
- Krause, D. E., Gebert, D. & Thornton, G. C. III (2007). Interkulturelle Unterschiede in der Assessment Center Anwendung. In H. Schuler (Hrsg.), *Assessment Center zur Potenzialanalyse* (S. 359–378). Göttingen: Hogrefe.
- Krause, D. E., Kersting, M., Heggstad, E. D. & Thornton, G. C. III (2006). Incremental validity of assessment center ratings over cognitive ability tests. A study at the executive management level. *International Journal of Selection and Assessment*, 14, 360–371.
- Krause, D. E. & Thornton, G. C. III (2007). Kulturelle Vielfalt und Assessment Center: Ergebnisse einer Expertenstudie zu kulturellen Wertorientierungen und der Assessment Center-Anwendung. *Zeitschrift für Führung und Organisation*, 4, 223–230.
- Kuncel, N. R. & Hezlett, S. A. (2007). Standardized tests predict graduate students' success. *Science*, 315, 1080–1081.
- Lievens, F. & Thornton, G. C. III (2005). Assessment centers: Recent developments in practice and research. In A. Evers, O. Voskuijl & N. Anderson (Eds.), *Handbook of personnel selection* (pp. 243–264). London: Blackwell.
- Marcus, B., Goffin, R., Johnston, N. G. & Rothstein, M. G. (in press). Personality and cognitive ability as predictors of typical and maximum managerial performance. *Human Performance*.
- Motowidlo, S. J. & Van Scotter, J. R. (1994). Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology*, 79, 475–480.
- Pynes, J. E. & Bernadin, H. J. (1989). Predictive validity of an entry-level police officer assessment center. *Journal of Applied Psychology*, 74, 831–833.
- Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C. & De Fruyt, F. (2003). International validity generalization of GMA and cognitive abilities: A European community meta-analysis. *Personnel Psychology*, 56, 573–605.
- Sarges, W. (2006). Management-Diagnostik. In F. Petermann & M. Eid (Hrsg.), *Psychologische Diagnostik* (S. 739–746). Göttingen: Hogrefe.
- 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 finding. *Psychological Bulletin*, 124, 262–274.
- Schmitt, N., Gooding, R. Z., Noe, R. A. & Kirsch, M. (1984). Meta-analysis of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology*, 37, 407–422.
- Scholz, G. & Schuler, H. (1993). Das nomologische Netzwerk des Assessment Centers: Eine Meta-Analyse. *Zeitschrift für Arbeits- und Organisationspsychologie*, 37, 73–85.
- Thornton, G. C. III & Krause, D. E. (2007). *The design, execution, and evaluation of selection versus development assessment centers – A study in north and central America, Asia, and Europe*. Symposium paper on assessment center methods presented at the 22nd annual conference for the Society of Industrial and Organizational Psychology, New York City, U. S. A.
- Thornton, G. C. III & Rupp, D. E. (2006). *Assessment centers and human resource management*. Mahwah, NJ: Erlbaum.
- Thurstone, L. L. & Thurstone, T. G. (1946). *Primary mental abilities*. Chicago, IL: Science Research Associates.
- Viswesvaran, C., Ones, D. S. & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, 81, 557–574.

In:

Innovative Ansätze für die Eignungs- diagnostik

herausgegeben von

Werner Sarges und David Scheffer (2008)

ISBN 978-3-8017-2182-4

HOGREFE



GÖTTINGEN · BERN · WIEN · PARIS · OXFORD · PRAG
TORONTO · CAMBRIDGE, MA · AMSTERDAM · KOPE