

Kersting, M. (2009). Profit durch Personalauswahl – warum sich eine qualitativ hochwertige Personalauswahl langfristig rechnet.
In A. Gourmelon, C. Kirbach & S. Etzel (Hrsg.), Personalauswahl im öffentlichen Sektor (2. neubearbeitete Aufl., 33-54). Baden-Baden: Nomos.
(Erstauflage 2005, S. 15-34).

Andreas Gourmelon/Christine Kirbach/
Stefan Etzel (Hrsg.)

2009

Personalauswahl im öffentlichen Sektor

2. erweiterte und überarbeitete Auflage



Nomos

Profit durch Personalauswahl – warum sich eine qualitativ hochwertige Personalauswahl langfristig rechnet

Dr. Martin Kersting

Die meisten Prozesse einer Organisation werden mittlerweile unter Rationalisierungsgesichtspunkten betrachtet – nur die (internen und externen) Personalentscheidungen scheinen oft von jeglichen Rationalitätserwägungen ausgenommen zu werden. Personalentscheidungen werden häufig nach wie vor aus dem Bauch heraus getroffen oder durch zweifelhafte Personalauswahlverfahren unterstützt. Angesichts der großen Bedeutung, die der Mitarbeiter¹ für die Leistungsfähigkeit und somit letztendlich für die Wirtschaftlichkeit und den Erfolg einer Organisation hat, ist es erstaunlich, dass Personalentscheidungen häufig nicht überprüft werden – im Gegensatz zu anderen Investitionsentscheidungen der Organisation. Dabei übersteigen die Personalinvestitionen nicht selten deutlich diese anderen Investitionen. Gleichwohl wird für eine Sachinvestition oft ein aufwändiges Ausschreibungs- und Begutachtungsverfahren eingeleitet, während wesentlich kostenträchtigere Personalentscheidungen „freihändig“ getroffen werden. Ein Grund für diese Nachlässig- bis Fahrlässigkeit ist die weit verbreitete Ansicht, dass sich Kosten und Nutzen von Personalentscheidungen als weiche Faktoren nicht rechnen lassen. Diese Position ist auf den ersten Blick bequem, denn tatsächlich sind institutionelle Nutzenberechnungen schwierig und die notwendigen Festlegungen einzelner Kennziffern sind im Detail angreifbar. Langfristig schadet die Position aber den Personalverantwortlichen, die sich als strategischer Partner der Leitung letztendlich nur dann in einer Organisation behaupten können, wenn auch der finanzielle Wert ihrer Arbeit offensichtlich wird. Und die Position schadet der Organisation, die durch suboptimale Personalentscheidungsstrategien hinter ihren Möglichkeiten zurückbleibt. Schließlich schadet die Position auch den Mitarbeitern, die nicht ihren Fähigkeiten entsprechend eingesetzt werden und die mit „unpassenden“, leistungsschwachen oder demotivierten Kollegen zusammenarbeiten müssen. In dem vorliegenden Beitrag werden zunächst die Grundannahmen der Nutzenanalyse herausgearbeitet. Nach einem kurzen Überblick über die Möglichkeiten der Leistungsbeurteilung, die finanziellen Konsequenzen der Streuung beruflicher Leistung sowie der Treffsicherheit verschiedener Personalaus-

¹ Im Interesse der Kürze und Flüssigkeit des Textes wird bei der Bezeichnung von Personengruppen nur die männliche Form verwendet; gemeint sind jeweils Männer und Frauen.

wahlverfahren wird eine konkrete Methode zur Bestimmung des Nutzens von Personalauswahlverfahren vorgestellt. Anschließend werden einige in der Literatur dargestellte Ergebnisse solcher Nutzenanalysen referiert. Zum Abschluss des Beitrags wird die Methode der Nutzenanalyse auf ein konkretes, faktisches Beispiel aus dem Bereich des öffentlichen Dienstes angewandt. Alle Analysen zeigen, dass qualitativ hochwertige Personalauswahlverfahren langfristig zu einer erheblichen Verbesserung der finanziellen Situation einer Organisation beitragen. Der Begriff „Personalauswahl“ wird hier vereinfachend als Oberbegriff für Verfahren der Eignungsdiagnostik im Kontext der Personalentwicklung und -auswahl genutzt, auch Verfahren der Potenzialanalyse fallen darunter. Die Ausführungen beziehen sich sowohl auf Verfahren mit internen Mitarbeitern wie auf Verfahren mit externen Bewerbern.

Die Grundannahmen, die sich aus der Perspektive der Kosten-Nutzen-Analyse auf das Personalauswahlgeschehen ergeben, sind simpel:

- (1) Menschen unterscheiden sich in ihrer Berufsleistung. Schlichter formuliert:
Es gibt leistungsstarke und leistungsschwache Mitarbeiter.

Es dürfte keinen Personaler geben, der dieser Aussage widerspricht. Diese Unterschiede in der Berufsleistung einzelner Mitarbeiter schlagen sich auf die Effektivität und somit letztendlich auf betriebswirtschaftliche Kennzahlen der Organisation nieder.

- (2) Personalauswahlverfahren unterscheiden sich sowohl hinsichtlich ihrer Kosten als auch hinsichtlich ihrer Aussagekraft und Gültigkeit (Validität).

Dass sich die Aussagekraft von verschiedenen Auswahlverfahren deutlich unterscheidet, ist die Erkenntnis aus fast 100 Jahren entsprechender empirischer Forschung; gleichwohl behandeln viele Entscheider die Frage geeigneter Personalauswahlverfahren als Glaubensfrage. Dabei ist längst erwiesen, dass z. B. standardisierte Interviews aussagekräftigere Ergebnisse liefern als unstandardisierte (freie) Vorstellungsgespräche; dass Intelligenztests unter Validitätsgesichtspunkten hervorragende Instrumente der Personalauswahl sind usw. Es gibt anerkannte Richtlinien und Qualitätskriterien für eine qualitätsgerechte Personalauswahl, z. B. die DIN 33430 (DIN, 2002; Gourmelon im vorliegenden Buch; Kersting, 2008), es ist aber offensichtlich nicht deutlich, welcher (auch) finanzielle Schaden durch die Missachtung der Erkenntnisse und Qualitätsstandards entsteht.

Um den kostenbereinigten Nutzen von Auswahlverfahren bestimmen zu können, muss man zunächst eine Operationalisierung der Berufsleistung vornehmen und ihre finanziellen Dimensionen bestimmen. Außerdem benötigt man Informationen über den Grad der Übereinstimmung zwischen diesem Kriterium der Berufsleistung einerseits und dem mithilfe eines Personalauswahlverfahrens erfassten Eignungsgrad

einer Person andererseits. Diese Übereinstimmung nennt man Kriteriumsvalidität (oder -gültigkeit) des Verfahrens. Nach Berücksichtigung einiger weiterer Parameter, die weiter unten erläutert werden, kann man dann den kostenbereinigten Nutzen von Auswahlverfahren berechnen.

2.1 Die Beurteilung beruflicher Leistung

Es gibt leistungsstarke und leistungsschwache Mitarbeiter – so bereitwillig man dieser Aussage zustimmt, so schwer fällt es, diese Leistungsunterschiede zu bestimmen. Berufliche Leistung lässt sich auf unterschiedlichen Maßen abbilden, Schuler (2004, S. 6f.) unterscheidet (1) Verhalten, (2) Tätigkeiten, (3) Fähigkeiten/Eigenschaften, (4) Kenntnisse, (5) Ziele und (6) Ergebnisse.

Dette, Abele und Renner (2004) legen drei Ordnungsparameter zugrunde. Zunächst unterscheiden sie zwischen (1a) arbeits- oder aufgabenbezogenen Kriterien einerseits und (1b) laufbahnbezogenen Kriterien andererseits. Arbeits- oder aufgabenbezogene Berufserfolgskriterien sind aufgabenspezifisch, z. B. die Erreichung bestimmter Ziele bezüglich der Arbeitsmenge oder der Qualität des Umgangs mit Beschwerden usw. Laufbahnbezogene Kriterien betrachten hingegen die persönlichen Karriereerfolge über einen längeren Zeitraum hinweg. Als zweiten Ordnungsparameter betrachten Dette et al. (ebd.) die Datenart. Handelt es sich (2a) um neutrale Kennzahlen, (2b) um Vergleiche mit Bezugsstandards oder (2c) um Zufriedenheitseinschätzungen? Schließlich unterscheiden sie verschiedene Datenquellen, nämlich (3a) Akten/Dokumente, (3b) Fremdurteile (z. B. Vorgesetztenbeurteilungen) und (3c) Selbstberichte. Tabelle 1 zeigt das vollständige Modell.

Tabelle 1: Ordnungsparameter zur Bestimmung des beruflichen Erfolgs (nach Dette et al., 2004, S. 174)

B e z u g s k r i t e r i u m	Datenquelle		Akten/ Dokumente	Fremdurteil	Selbstbericht
	Datenart				
	spez. Ar- beit	neutrale Kenn- zahlen	z. B. Stück- zahlen		z. B. Stückzahlen
		Vergleich mit Bezugsmaßstab		z. B. Leistungsbeurteil- ung	z. B. Leistungs- einschätzung
		Zufriedenheit			z. B. Zufrieden- heit mit der Tätigkeit
	globale Laufbahn	neutrale Kenn- zahlen	z. B. Gehalt		z. B. Gehalt
		Vergleich mit Bezugsmaßstab		z. B. Beförde- rungs- würdigkeit	z. B. Erfolg im Vergleich zu Kol- legen
		Zufriedenheit			z. B. Zufrieden- heit mit der Laufbahn

Die aktuellen multiperspektivischen Ansätze (360-Grad-Beurteilungen, siehe z. B. Scherm, 2005) differenzieren stärker innerhalb der Quellen des Fremdurteils. Neben der Beurteilung von „oben“ (durch den Vorgesetzten) und der Selbstbeurteilung können auch Urteile von Kollegen, Mitarbeitern, Kooperations- und Projektpartnern sowie Kunden eingeholt werden.

Inhaltlich wird neuerdings auch zwischen der aufgabenbezogenen (*task performance*) und der umfeldbezogenen (*citizenship behavior*) Leistung unterschieden (Organ, Podsakoff & MacKenzie, 2006). Umfeldbezogene Leistung dient zwar letztendlich auch der Aufgabenerfüllung, die Wirkung ist aber eher indirekt, z. B. durch (ansteckenden) Enthusiasmus bei der Arbeit und Unterstützung anderer.

Gleichgültig welches Kriterium man zur Messung des Berufserfolgs heranzieht: Kaum ein Kriterium kann vollständig überzeugen. Die meisten unterliegen einer Verzerrung (siehe z. B. Etzel & Küppers, 2002, S. 189f.). Dennoch gilt: Auch Kriterien, die nicht vollständig überzeugen können, sind immer noch besser als überhaupt keine Kriterien. Die Kombination verschiedener Kriterien zu einem Composite- oder Overall-Kriterium (siehe Etzel & Küppers, ebd.) kann sinnvoll sein, setzt aber ebenfalls brauchbare Einzelkriterien voraus. Viele Kriterien sind in bestimmten Kontexten

ten mehr oder weniger wertlos. Für den öffentlichen Dienst ist eine Operationalisierung des Berufserfolgs über das Gehalt beispielsweise nur bedingt aussagekräftig, da sich die Bezahlung an der Position und nicht an der Leistung orientiert. Auch Stückzahlen kommen im öffentlichen Dienst häufig nicht infrage. Es ist schwer anzugeben, welchen Beitrag zum Gemeinwohl eine Arbeit erbracht hat. Gleichwohl kann der öffentliche Dienst nicht aus betriebswirtschaftlichen Erwägungen ausgeklammert werden, er steht im Gegenteil unter Wirtschaftlichkeitsdruck. Um die Streuung der Berufsleistung und deren finanzielle Auswirkungen für Bereiche zu bestimmen, die nicht direkt produktions- und verkaufsorientiert sind, wurden Schätzverfahren eingeführt, die sich am Geldwert des Outputs oder des Gewinns orientieren. Diese gehen davon aus, dass Experten (z. B. Vorgesetzte) die Leistungsunterschiede zwischen den Mitarbeitern in jedem Beruf ebenso bestimmen können wie den Nutzen dieser Leistungen für die Organisation. Konkret werden die Experten beispielsweise aufgefordert, den finanziellen Wert der von einem durchschnittlichen Mitarbeiter geleisteten beruflichen Tätigkeit für die Organisation zu schätzen. Eine entsprechende Aufgabe lautet etwa: „Bitte beziffern Sie die Kosten, die anzusetzen sind, wenn man die von einem durchschnittlichen Mitarbeiter eines spezifischen Bereichs geleisteten Dienstleistungen von einer Fremdfirma beziehen würde.“ Bei Organisationen, bei denen sich der finanzielle Nutzen der Tätigkeit nur schwer bestimmen lässt, z. B. bei der Polizei, kann die Bestimmung auch über die Kostenseite vorgenommen werden. So fragte man beispielsweise die Experten, wie viel Personal und Material ein Mitarbeiter für die Bewältigung einer definierten Aufgabe benötigt. Dabei zeigte sich, dass überdurchschnittlich leistungsstarke Mitarbeiter ungleich weniger Personal und Material für die qualitativ hochwertige Bewältigung einer Aufgabe benötigen als leistungsschwache Mitarbeiter.

2.2 Die Streuung der beruflichen Leistung und ihre finanziellen Konsequenzen

Auf welche Art man die Berufsleistung auch erfasst, sofern man zuverlässige und gültige Messungen vornimmt, stellt man immer eine erhebliche Streuung der beruflichen Leistung fest. Schmidt und Hunter (1998) fassen die Ergebnisse zahlreicher wissenschaftlicher Untersuchungen zur Variabilität der Berufsleistung zusammen. Das Ausmaß der Streuung wird als Standardabweichung bestimmt. Je nach Niveau der Arbeitstätigkeit fanden sich Standardabweichungen zwischen 19 % und 48 %. Bei einfachen Tätigkeiten produzierte ein überdurchschnittlicher Mitarbeiter im statistischen Mittel 19 % mehr Output als der durchschnittliche Mitarbeiter. Bei Facharbeitern übertraf die Produktion der überdurchschnittlichen Mitarbeiter die des Durchschnitts um 32 %, und ein überdurchschnittlicher Manager erzielte für seine Organisation 48 % mehr Output als der Durchschnitt. Grundsätzlich gilt also: je komplexer die Tätigkeit, desto größer die Leistungsstreuung.

Aus zahlreichen Forschungsarbeiten (siehe Schmidt, Hunter & Pearlman, 1982) konnten mittlerweile die folgenden „Faustregeln“ abgeleitet werden:

- Der jährliche Wert des Outputs eines durchschnittlichen Mitarbeiters ist in der Regel doppelt so hoch wie sein jährliches Entgelt.
- Die Standardabweichung der in Geldwert umgerechneten Leistung zwischen überdurchschnittlichen und durchschnittlichen Mitarbeitern beträgt zwischen 40 % und 70 % des jährlichen Entgelts.

Der genannte Unterschied bezieht sich auf die Differenz zwischen den besten Mitarbeitern und dem Durchschnitt. Nimmt man einen extremeren Vergleich vor und vergleicht die Leistung der überdurchschnittlich starken Mitarbeiter mit der Leistung der schwächsten Mitarbeiter, muss man die Differenzen verdoppeln, also von einer Streuung zwischen 80 % und 140 % ausgehen. Daraus ergibt sich die „klassische“ Erkenntnis:

Der leistungsstärkste Mitarbeiter erbringt in der Regel doppelt so viel wie der schwächste.

2.3. Zur Treffsicherheit verschiedener Personalauswahlverfahren

Angesichts dieser großen Streuungen in der Berufsleistung, die sich für eine Organisation finanziell auswirken, ist es evident, dass eine treffsichere Auswahl von leistungsstarken Mitarbeitern sich finanziell rechnet und dass der finanzielle Nutzen eines Personalauswahlverfahrens umso höher ist, je besser (treffsicherer) das eingesetzte Verfahren ist. Einzelne Personalauswahlverfahren unterscheiden sich erheblich in ihrer Vorhersagekraft. Die Treffsicherheit der aus verschiedenen Personalauswahlverfahren abgeleiteten Prognosen über den zukünftigen beruflichen Erfolg wird von Wissenschaftlern seit bald 100 Jahren immer wieder überprüft. Seit einiger Zeit ist es möglich, die Ergebnisse der zahlreichen Einzelstudien mithilfe der sogenannten Meta-Analyse in statistisch kontrollierter Form zusammenzufassen (siehe z. B. Schmidt & Hunter, 1998). Dabei zeigte sich, dass vor allem standardisierte und regelgeleitete Verfahren gute Zukunftsprognosen erlauben. So übertrifft das standardisierte Interview mit einem Validitätskoeffizienten von $r=.51$ hinsichtlich der Vorhersage des Berufserfolgs die Validität von unstrukturierten Vorstellungsgesprächen ($r=.38$). Zur Vorhersage des Ausbildungs- und Berufserfolgs sind besonders Intelligenztests hervorragend geeignet. Dies zeigte sich sowohl in den bei Schmidt und Hunter (1998) dargestellten nord-amerikanischen Meta-Analysen als auch in entsprechenden Arbeiten für den europäischen Raum. Salgado, Anderson, Moscoso, Bertua und de Fruyt (2003) analysierten die Ergebnisse europäischer Studien. In ih-

rer Meta-Analyse beziffern sie den Zusammenhang zwischen Intelligenz und Berufserfolg sogar auf $r=.62$, wobei dieser Berechnung allerdings extreme statistische Optimierungsprozeduren zugrunde lagen. Eine Metaanalyse der deutschen Studien zur Kriteriumsvalidität von Intelligenztests wurde von Hülshager, Maier und Stumpp (2007) berechnet. Die Autoren kommen auf einen Vorhersagewert von $r=.47$ für Ausbildungserfolg und $r=.53$ für Berufserfolg. Nach der neuesten metaanalytischen Übersicht von Kramer (2009) ergeben sich für die Intelligenz Koeffizienten von .62 für die Vorhersage von Arbeitsleistungen, .59 für die Vorhersage von Lernleistungen, .33 für die Vorhersage des Einkommens und .31 für die Vorhersage von beruflichen Entwicklungen.

Hinsichtlich der Gültigkeit der Aussagekraft bestimmter Personalauswahlverfahren kann man sich an den in der Fachliteratur berichteten Werten orientieren. Dies kann allerdings die kritische Prüfung im Einzelfall nicht ersetzen. Auch wenn beispielsweise Intelligenztests in der Regel sehr gute prognostische Validitäten aufweisen, gibt es einzelne Intelligenztests, die z. B. auf der Basis veralteter wissenschaftlicher Annahmen konstruiert wurden und deshalb untauglich sind (siehe z. B. Kersting, 2007).

Die Kosten eines Auswahlverfahrens sind relativ leicht nachzuvollziehen. Allerdings werden in der Praxis die Kosten häufig falsch kalkuliert, weil die internen Zeitaufwendungen nicht in Rechnung gebracht werden. Nach Art der Milchmädchenrechnung wird beispielsweise das Angebot, die Personalauswahl durch schriftliche Testverfahren zu unterstützen, als „zu teuer“ abgelehnt, um dann zeitaufwändige Vorstellungsgespräche mit allen Bewerbern zu führen. Ein Test kann als Gruppenverfahren mit z. B. 30 Bewerbern gleichzeitig durchgeführt werden. Bei Kosten in Höhe von beispielsweise 50 Euro pro getesteter Person würden somit Gesamtkosten von 1.500.- Euro entstehen. Mit einem Test könnte man treffsicher die fünf Bewerber herausfiltern, mit denen sich ein Eignungsinterview in der zweiten Stufe des Auswahlverfahrens lohnt. Stattdessen wird mit allen 30 Bewerbern ein Interview geführt, die Zeit für die 25 unnötigen Interviews à mindestens 30 Minuten, insgesamt also 25 mal 30 Minuten, also über 12 Stunden Arbeitszeit, wird aber nicht in die Rechnung einbezogen. Natürlich muss man aber nicht nur die monetären Kosten, sondern auch den Einsatz von Ressourcen transparent werden lassen. In der Regel sind aufwändige Verfahren wie standardisierte Eignungsinterviews und Assessment-Center teurer als handgestrickte, unstandardisierte Vorstellungsgespräche. Gleichwohl sind auch die zuletzt genannten, in Eigenregie durchgeführten Verfahren nicht kostenlos. Kennt man die Kosten und die Aussagekraft verschiedener Personalauswahlverfahren, kann man im nächsten Schritt prüfen, ob sich die Investition in etwaige Mehrkosten lohnt.

2.4 Kosten und Nutzen von Personalauswahlverfahren

Die bislang thematisierten Kennwerte der Streuung der Berufsleistung und der Qualität des Personalauswahlverfahrens sowie seiner Kosten sind zentral bei der Bestimmung des Nutzens von beruflichen Eignungsbeurteilungen. Ein häufig angewendetes Nutzenmodell geht auf Brogden (1949) sowie Cronbach und Gleser (1965) zurück und wird entsprechend als Brogden-Cronbach-Gleser-(BCG-)Nutzenmodell bezeichnet.

Der zusätzliche Nettonutzen einer qualitätsgesicherten Personalauswahl – hier im Vergleich zu einer Zufallsauswahl – wird in diesem Modell wie folgt in Geldeinheiten (z. B. Euro) berechnet.

$$\Delta(U) = N_A \times T \times r_{xy} \times \bar{Z}_x \times SD_y - C \times N_B$$

- N_A bezeichnet die Anzahl der eingestellten Personen
 T bezeichnet deren Verweildauer in der Organisation
 r_{xy} ist die Gültigkeit des Verfahrens, ausgedrückt als Korrelationskoeffizient
 $\bar{Z}_x$ ist der (standardisierte) Wert, den die letztlich ausgewählten Bewerber durchschnittlich im Auswahlverfahren erzielen
 SD_y ist die Standardabweichung der in der Organisation festgestellten Berufsleistung in Geldeinheiten
 C steht für die Kosten des Auswahlverfahrens pro Bewerber
 N_B beziffert die Anzahl der Bewerber

Abbildung 1: Brogden-Cronbach-Gleser-(BCG-)Nutzenmodell

Erläuterungsbedürftig sind die Werte $\bar{Z}_x$ und SD_y .

- $\bar{Z}_x$ ist der Wert, den die letztlich ausgewählten Bewerber durchschnittlich im Auswahlverfahren erzielen. Dies spiegelt die unterschiedliche „Strenge“ der jeweiligen Personalauswahl wider. Der Nutzen des Auswahlverfahrens hängt u. a. davon ab, ob nur „Top-Bewerber“ (z. B. die in einem Assessment-Center mit „sehr gut“ beurteilten Personen) eingestellt werden (hoher Prädiktorwert) oder (z. B. aufgrund eines Bewerbermangels) auch mittelmäßige oder gar schwache Bewerber zugelassen oder eingestellt werden mussten. Da die relativ leistungsschwachen Bewerber voraussichtlich auch schwache Berufsleistungen erbringen werden, ist der Nutzen eines Auswahlverfahrens umso geringer, je mehr „schwächere“ Bewerber akzeptiert werden. Weil die in der Praxis in Auswahlverfahren genutzten Beurteilungsskalen variieren (z. B. von „1“ bis „6“ oder von „0“ bis „100“), die Formel aber universell anwendbar sein soll, wird der entsprechende Wert vorab auf eine einheitliche Skala (z-Skala) transformiert, d. h. standardisiert. Die genutzte z-Wert-Verteilung geht praktisch von minus drei bis plus drei, mit einer Standardabweichung von eins. Der Wert „Null“ auf der z-Skala entspricht somit einem durchschnittlichen Wert. In der Regel stellt man aber keine „durchschnittlichen“ Bewerber ein, sondern Bewerber mit überdurchschnittlich guten Leistungen im Auswahlverfahren.
- SD_y , die Standardabweichung der Berufsleistung in Geldeinheiten, wurde weiter oben (Abschnitt 2.2) ausführlich thematisiert. Hinter diesem Kennwert verbirgt sich die Variabilität des Nutzens der Berufsleistung verschiedener Mitarbeiter. Ein Auswahlverfahren lohnt sich vor allem für Positionen, bei denen die Berufsleistungen und der damit verbundene finanzielle Nutzen für die Organisation stark streuen. Methoden zur Berechnung oder Schätzung dieser Streuung wurden weiter oben ebenfalls vorgestellt. Eine realistische, wenngleich konservative Voreinstellung für diesen Parameter ist eine Standardabweichung in Höhe von 40 % der Jahresgesamtbezüge.

Höft (2001, S. 644 f.) unterscheidet in der dargestellten Formel den Qualitätsterm (r_{xy} , $\bar{Z}_x$ und SD_y) und den Quantitätsterm (N_A und T). Nicht nur die Qualität (der Personalauswahlverfahren, der eingestellten Bewerber und der Berufsleistung) entscheidet, sondern auch quantitative Elemente spielen eine Rolle: Der Nutzen variiert in Abhängigkeit von der Anzahl der ausgewählten Personen und deren Verweildauer in der Organisation. Organisationen mit traditionell hoher Verweildauer der einzelnen Mitarbeiter, beispielsweise der öffentliche Dienst, profitieren besonders von einer qualitativ hochwertigen Personalauswahl bzw. können es sich am wenigsten erlauben, bei der Personalauswahl nachlässig vorzugehen. Von dem Produkt der ersten fünf Parameter, die den Bruttonutzen bestimmen, sind natürlich die Kosten (Produkt aus der Anzahl der untersuchten Bewerber (N_B) und der Verfahrenskosten pro Bewerber (C)) abzuziehen. Der Gesamtnutzen ergibt sich aus der Formel Qualität mal

Quantität minus Kosten. Dividiert man das Ergebnis der Gleichung durch die Anzahl der eingestellten Personen (N_A), erhält man den Nutzenzuwachs pro eingestellte Person.

Zu beachten ist, dass ein Verfahren keinen Zusatznutzen erbringt, sobald einer der Parameter den Wert Null annimmt. Die Validität (Gültigkeit) des Auswahlverfahrens ist direkt proportional zu seinem Nutzen.

Aktuellere Formeln zur Berechnung des Nutzenzuwachses (z. B. Boudreau, 1989; Cronshaw & Alexander, 1985) berücksichtigen weitere Parameter wie z. B. die Rekrutierungskosten, die Entwicklungskosten für das Verfahren sowie seine Einsatzhäufigkeit, die Ablehnung von Angeboten durch die Bewerber, den Einfluss der Probezeit, die Fluktuation, die Freisetzung von (leistungsschwächeren) Mitarbeitern die Diskontierung, die Inflationsrate und die notwendige Versteuerung der Gewinne. Darüber hinaus gibt es Bestrebungen, die Kosten-Nutzen-Analyse durch verbesserte Operationalisierungen auf Seiten des Kriteriums zu optimieren, indem nicht nur der Nutzen auf der Ebene der individuellen Leistungsunterschieden betrachtet wird, sondern auch auf der organisationalen Ebene (z. B. Steigerung der Leistung von Teams, Steigerung der Kundenzufriedenheit, siehe Görlich & Schuler, 2006). Dadurch werden die Berechnungsformeln der Dynamik des Organisationsgeschehens gerechter. Grundsätzlich unterscheiden sich die mit den komplexeren Modellen erzielten Ergebnisse aber nicht von den Ergebnissen, die mit dem Brogden-Cronbach-Gleser-(BCG-)Modell erzielt werden, sodass im Folgenden der Einfachheit halber Beispiele nach dem BCG-Modell vorgestellt werden.

Unter der URL <http://www.psy.uni-muenster.de/Psychologie.inst2/AEHell/personen/rowold/taetigkeit/organisation/kna.html> kann eine von Rowold (2007) erstellte Excel-Tabelle zur praktischen Durchführung von Kosten-Nutzen-Analysen abgerufen werden. Die Erläuterungen dazu finden sich bei Rowold (2007).

2.5 Anwendungsbeispiele aus der Literatur

Nachfolgend sollen einige Beispiele aus der Literatur referiert werden. Das Referat beschränkt sich auf deutschsprachige Publikationen, weil die zahlreichen nord-amerikanischen Studienergebnisse nicht ohne weiteres auf unsere Kulturverhältnisse übertragbar sind. In einem Anwendungsbeispiel vergleicht Kersting (2004) den Nutzen von standardisierten und unstandardisierten Eignungsinterviews. Angenommen eine Organisation will zwei Personen einstellen ($N_A=2$). Ausgewählt werden die beiden Personen aus einer Gruppe von insgesamt 14 Bewerbern ($N_B=14$). Das Jahresentgelt jeder eingestellten Person beträgt 50.000 Euro. Entsprechend den oben (Abschnitt 2.2) erläuterten Regeln, denen zufolge der jährliche Wert des Outputs eines durchschnittlichen Mitarbeiters in der Regel doppelt so hoch ist wie sein jährliches Entgelt, wird von einem durchschnittlichen Output von 100.000 Euro pro Mitarbeiter ausgegangen. Überdurchschnittliche Mitarbeiter leisten gegenüber durch-

schnittlichen Mitarbeitern den oben zitierten Erfahrungen zufolge mindestens 40 % mehr an Output, sodass die geschätzte „Standardabweichung der Berufsleistung“ auf 40.000 Euro angesetzt wird ($SD_y=40.000$). Es wird davon ausgegangen, dass die letztlich eingestellten Personen im Durchschnitt im Auswahlverfahren einen Wert erzielen, der eine Standardabweichung über dem Durchschnitt der Gesamtgruppe liegt ($z_x=1$), und dass die beiden ausgewählten Personen zehn Jahre in der Organisation verbleiben ($T=10$). Die Kosten für das standardisierte Eignungsinterview werden mit 500 Euro pro Person ($C_{\text{standardisiertes Eignungsinterview}}=500$) doppelt so hoch veranschlagt wie die Kosten des unstandardisierten Vorgehens ($C_{\text{unstandardisiertes Eignungsinterview}}=250$). Die Gültigkeitswerte der Verfahren werden der bei Schmidt und Hunter (1998) berichteten Literaturübersicht entnommen (siehe Abschnitt 1.2) und mit $r=.38$ für das unstandardisierte und mit $r=.51$ für das standardisierte Eignungsinterview angesetzt ($r_{xy \text{ unstandardisiertes Eignungsinterview}}=.38$; $r_{xy \text{ standardisiertes Eignungsinterview}}=.51$). Dem Anwendungsbeispiel zufolge ist also das unstandardisierte Interview kostengünstiger als die standardisierte Vorgehensweise, erlaubt dafür aber weniger gültige Aussagen. Hinsichtlich aller übrigen Parameter wird von den gleichen Bedingungen für die beiden Verfahren ausgegangen. Setzt man nun die Werte in die oben genannte Gleichung ein, so erzielt man die folgenden Ergebnisse.

Nutzenzuwachs (Zuwachs gegenüber einer Zufallsauswahl) bei der Anwendung des ...

a) unstandardisierten Eignungsinterviews =

$$2 * 10 * 0.38 * 1 * 40.000 - 250 * 14 = 300.500 \text{ Euro}$$

b) standardisierten Eignungsinterviews =

$$2 * 10 * 0.51 * 1 * 40.000 - 500 * 14 = 401.000 \text{ Euro}$$

Abbildung 2: Nutzenzuwachs (Zuwachs gegenüber einer Zufallsauswahl) durch unstandardisierte und standardisierte Interviews

Man erkennt, dass das auf den ersten Blick doppelt so teure Personalauswahlverfahren des standardisierten Eignungsinterviews gegenüber der vermeintlich kostengünstigeren Variante des unstandardisierten Vorgehens einen finanziellen Vorteil in Höhe von (400.100 € - 300.500 € =) 100.500 € erbringt.

Holling (1998) gibt eine Übersicht über 20 Studien zum Nutzen von Verfahren der beruflichen Eignungsbeurteilung, die auf unterschiedlichen Methoden der Bestimmung der Variabilität der Berufsleistung und des Nutzens der Personalauswahl basieren und die den enormen finanziellen Nutzen qualitativ hochwertiger Personalauswahlverfahren bestätigen. In einer eigenen Studie berechnet Holling (ebd.) den zusätzlichen Nutzen der Einführung von Assessment-Centern in Relation zu bislang genutzten unstandardisierten Eignungsinterviews anhand der Auswahl von Außen-

dienstmitarbeitern einer Versicherung. Der Autor schätzte dabei die Gültigkeit der Eignungsinterviews auf $r=.15$ und ging von einer nur geringfügig höheren Gültigkeit von $r=.26$ für die Assessment-Center aus. Unter realistischen Annahmen zu den Rahmenbedingungen und bei einer auf objektiven Verkaufszahlen beruhenden Schätzung der Standardabweichung der beruflichen Leistung ergab die äußerst konservative Schätzung des Gesamtnutzens (also nach Abzug der Kosten für Entwicklung und Anwendung) des Assessment-Centers einen Betrag in Höhe von fast 400.000 Euro.

Die gleiche Fragestellung (Nutzen der Einführung von Assessment-Centern gegenüber der Anwendung unstandardisierter Eignungsinterviews) untersuchten Stehle und Barthel (1984) für den Bereich der Einstellung von Führungsnachwuchskräften für die chemische Industrie. Bei der Einstellung von zwölf Hochschulabsolventen für ein Traineeprogramm ergaben sich unter den von den Autoren gesetzten Rahmenbedingungen durch die Einführung des auf den ersten Blick teureren Assessment-Centers langfristig Einsparungen von Personalbeschaffungs- und Ausbildungskosten in Höhe von einer halben Million Euro.

Den langfristigen Nutzen der Einführung eines umfangreichen Auswahlprogramms zur jährlichen Einstellung von 25 Personen im Bereich Forschung und Entwicklung berechneten Schuler, Funke, Moser und Donat (1995) mit knapp 3,8 Millionen Euro.

Höft (2001) berechnet den Netto-Nutzen für den Fall der Auswahl von 120 (zukünftigen) Bankkaufleuten aus einem Pool von 585 Bewerbern. Ausgehend von einer Standardabweichung des Nutzens der Leistung eines Bankkaufmanns in Höhe von 23.000 Euro sowie einer Verweildauer von zehn Jahren ergibt sich selbst für ein Testverfahren von mittlerer Validität ($r=.35$) ein Nutzenzuwachs (gegenüber der Zufallsauswahl) von fast fünf Millionen Euro.

Etzel und Küppers (2002) berechnen für den Fall von 148 eingestellten Personen mit einer extrem hohen Standardabweichung der in der Organisation festgestellten Berufsleistung einen Netto-Nutzen von mindestens 388 Mio. Schweizer Franken, maximal aber sogar 1.266 Mio. Schweizer Franken, wenn ein neues (computergestütztes) Testverfahren gegenüber dem bisher verwendeten Verfahren (Interview) zum Einsatz kommt.

Wenn in einem typischen Unternehmen des deutschen Mittelstands mit 1.400 Mitarbeitern unstrukturierte Vorstellungsgespräche durch strukturierte Eignungsinterviews ersetzt würden, ergäbe dies laut Stephan und Westhoff (2002) einen Nutzenzuwachs von 150.000 Euro pro Jahr.

Über den Nutzen eines halbtägigen Assessment-Centers zur Auswahl von Call-Center-Mitarbeitern berichten Rowold und Mönninghoff (2007). Die Studie ist besonders interessant, weil hier das umfassendere, aber auch aufwändigere Kosten-Nutzen-Analyseverfahren von Boudreau angewendet wurde. Die Validität des Verfahrens für die Vorhersage der Leistung der Mitarbeiter wurde für das Inbound-Geschäft regressionsanalytisch auf $R=.19$ geschätzt. Den Besonderheiten der Bran-

che entsprechend wurde eine Verweildauer der Call-Center-Agents von nur 18 Monaten angenommen. Trotz dieser eher ungünstigen Rahmenbedingungen berechnen die Autoren über den Zeitraum von 18 Monaten einen Nutzen des Assessment-Centers von 40.549 Euro im Inbound, das Verhältnis von Kosten zu Nutzen betrug 1:2.8.

3 Ein Praxis-Beispiel: Zur Aussagekraft und zum Nutzen eines Assessment-Centers hinsichtlich der Vorhersage des Erfolgs in der Polizei-Führungsakademie

Die Logik sogenannter Bewährungskontrollen und die (auch monetäre) Nützlichkeit von qualitativ hochwertigen Personalauswahlverfahren kann anhand eines faktischen, konkreten Beispiels aufgezeigt werden. Kersting (2003) hat ein Assessment-Center zur Auswahl von Führungskräften der Polizei evaluiert. Die Ergebnisse des von der Deutschen Gesellschaft für Personalwesen (DGP) durchgeführten Assessment-Centers wurden in einigen Bundesländern als Grundlage einer Zulassungsentcheidung herangezogen. Dabei ging es um die Zulassung zum Aufstieg vom gehobenen in den höheren Polizeivollzugsdienst. Bei der Evaluation sollte vor allem geklärt werden, ob sich die aufgrund des Verfahrens getroffenen Prognosen bewährt haben. Als Evaluationskriterium galt dabei der Ausbildungserfolg an der Polizei-Führungsakademie. Bei dem zweitägigen Assessment-Center wurden sowohl klassische Übungen (z. B. Präsentation, Gruppendiskussion) als auch Intelligenztestaufgaben verwendet. Das Verfahren ist bei Kersting (2003) ausführlich beschrieben, die vorliegende Darstellung beschränkt sich auf die Ergebnisse zur Treffsicherheit, die als Ausgangspunkt der nachfolgenden Nutzenanalyse gelten. Krause und Kersting (2008) sowie Krause, Kersting, Heggstad und Thornton (2006) weisen anhand des hier zugrunde liegenden Datensatzes nach, dass herkömmliche Assessment-Center-Beurteilungen über Intelligenztests hinaus inkrementelle Validität aufklären können.

3.1 Vorhersagekraft des Assessment-Centers

Tabelle 2 stellt die Verteilung der aus dem Assessment-Center abgeleiteten abschließenden Eignungsbeurteilungen der Verteilung des Kriteriums (dem Jahre später erzielten Abschluss an der Polizei-Führungsakademie) in Form einer Kreuztabelle deskriptiv gegenüber. Für jede „Zelle“ der Tabelle sind drei unterschiedliche Angaben verzeichnet, die am Beispiel der äußerst rechten Zelle in der ersten Zeile (Gesamtergebnis des Assessment-Centers: „den Anforderungen nur teilweise entsprechend“ und dem Ergebnis der Ausbildung an der Polizei-Führungsakademie: „nicht bestanden“) kurz erläutert werden sollen.

Tabelle 2: Kreuzklassifikation der Ergebnisse im Assessment-Center und des späteren Ausbildungserfolgs an der Polizei-Führungsakademie

Gesamtergebnis des Assessment-Centers	Abschluss der Polizei-Führungsakademie			
	„gut“ (*)	„befriedigend“	„ausreichend“	„nicht bestanden“
den Anforderungen nur teilweise entsprechend		9 (8,0%)	7 (6,3%)	2 (1,8%)
		50,0%	38,9%	11,1%
		12,5%	31,8%	66,7%
den Anforderungen weitgehend entsprechend – untere Gruppe	3 (2,7%)	34 (30,4%)	10 (8,9%)	1 (0,9%)
	6,3%	70,8%	20,8%	2,1%
	20,0%	47,2%	45,5%	33,3%
den Anforderungen weitgehend entsprechend – obere Gruppe	8 (7,1%)	23 (20,5%)	5 (4,5%)	
	22,2%	63,9%	13,9%	
	53,3%	32,0%	22,7%	
den Anforderungen entsprechend	4 (3,6%)	6 (5,3%)		
	40%	60%		
	26,7%	8,3%		
Gesamt N=112	15 (13,4%) (100%)	72 (64,3%) (100%)	22 (19,6%) (100%)	3 (2,7%) (100%)

(*) „sehr gut“ kam als Beurteilungsnote nicht vor.

- Die Angabe in der äußerst rechten Spalte der ersten Zeile der Tabelle bezieht sich auf die absolute Anzahl von zwei Personen. Die erste Zahl ist eine absolute Nennung, während die in Klammern aufgeführte Zahl den zugehörigen prozentualen Wert angibt. Zwei der insgesamt 112 analysierten Personen (also 1,8 % der Untersuchungsgruppe) haben im Assessment-Center die Wertung

„den Anforderungen nur teilweise entsprechend“ erzielt und die Prüfung an der Polizei-Führungsakademie „nicht bestanden“.

- Die Zahl darunter bezieht sich nur auf die Teilnehmer, die im Assessment-Center das Ergebnis „den Anforderungen nur teilweise entsprechend“ erzielt haben: 11,1 % der 18 Personen, die im Assessment-Center ein Ergebnis des Assessment-Centers dieser Kategorie erzielten, haben die Prüfung nicht bestanden (Zeilenprozent).
- Die letzte, grau hinterlegte Angabe pro Zeile bezieht sich nur auf die Personen, die in der Prüfung „durchgefallen“ sind: 66,7 % derjenigen, die in der Prüfung „durchgefallen“ sind, waren bereits im Assessment-Center mit einer schlechten Beurteilung, nämlich „den Anforderungen nur teilweise entsprechend“, bewertet worden (Spaltenprozent).

Um die Treffsicherheit der im Assessment-Center aufgestellten Prognosen zu bestimmen, ist die Zahl in der zweiten Zeile pro Zeile ausschlaggebend. Die Organisation will erfahren, inwieweit sie sich darauf verlassen kann, dass ein Bewerber, der im Assessment-Center sehr gut abgeschnitten hat und als „den Anforderungen entsprechend“ klassifiziert wurde, auch tatsächlich den Anforderungen der Ausbildung und des Berufs genügt. Die Kreuztabelle ist diesbezüglich ein Beleg für die Treffsicherheit des Assessment-Centers: Tendenziell gilt, dass die im Assessment-Center positiv beurteilten Personen auch an der Polizei-Führungsakademie erfolgreich waren. Mit zunehmend schlechterem Ergebnis im Assessment-Center nimmt auch die durchschnittliche Prüfungsleistung ab. Besonders interessant ist die Gruppe der Personen, die im Assessment-Center (relativ) schlecht abgeschnitten hat und lediglich das Ergebnis „den Anforderungen nur teilweise entsprechend“ erzielen konnte. Aufgrund der Ergebnisse im Assessment-Center bestanden hier seitens der Assessoren deutliche Bedenken gegenüber einer Zulassung. Tatsächlich haben 50 % dieser Gruppe die Prüfung entweder nur mit der schlechtesten Note „ausreichend“ bestanden oder sind „durchgefallen“. Keine einzige Person aus dieser Gruppe konnte die Note „gut“ erzielen. Das Risiko, die Ausbildung nicht zu bestehen, ist bei diesen Personen größer als bei den im Assessment-Center positiver beurteilten Personen. In der Regel ist man bei wichtigen und kostspieligen Personalentscheidungen nicht risikofreudig. Unsichere Kandidaten werden zugunsten aussichtsreicherer Bewerber zurückgewiesen. Dass die Gruppe der mit „den Anforderungen nur teilweise entsprechend“ beurteilten Personen mit drei Personen einen so geringen Umfang hat, ist nicht verwunderlich. Zahlreiche Bewerber, die im Assessment-Center mit diesem Ergebnis oder deutlich schlechter (mit „den Anforderungen weitgehend nicht entsprechend“ und „den Anforderungen nicht entsprechend“) beurteilt wurden, sind erst gar nicht zur Ausbildung zugelassen worden. Bei den 112 Personen, deren Ergebnisse hier ausgewertet werden, handelt es sich um eine Auslese aus einer Gruppe von fast 600 Assessment-Center-Teilnehmern (siehe unten). Setzt man gedanklich den klaren Zusammenhangstrend fort, so ist zu erwarten, dass die zahlreichen Be-

werber, die auf Grund ihrer Leistung im Assessment-Center schlecht beurteilt wurden, in der Prüfung versagt hätten, wenn man sie zur Ausbildung zugelassen hätte.

In einem weiteren Analyseschritt wurde der statistische Zusammenhang zwischen den Ergebnissen des Personalentscheidungsverfahrens und dem Punktwert der Abschlussprüfung als Kriterium in korrelativer Form bestimmt. Um mögliche Missverständnisse bei der Interpretation zu vermeiden, wird hier die logisch positive Korrelation berichtet, obwohl rein nominell der unterschiedlichen Polung der beiden Skalen wegen der Zusammenhang negativ ausfällt. Der in Tabelle 2 dargestellte Zusammenhang entspricht einer Rangkorrelation von $r=.46$. In Tabelle 2 sind sowohl die Ergebnisse des Assessment-Centers als auch die Ergebnisse der Prüfung zu Kategorien zusammengefasst worden. Dem Vorteil der Übersichtlichkeit steht der Nachteil gegenüber, dass möglicherweise bestehende oder ausbleibende Zusammenhänge auf einer feiner auflösenden Betrachtungsebene nicht entdeckt werden. Außerdem sind die Werte auf dieser groben Auflösungsstufe nicht normalverteilt. Schließlich bietet die Kategorienbildung Raum zur Manipulation, indem jeweils die „passenden“ Werte zu Kategorien gebündelt werden. Zusätzlich wurde daher auch der exakte Zusammenhang zwischen den Ergebnissen im Assessment-Center auf der Ebene aller möglichen Ergebnisstufen einerseits und den exakten Punktwerten (mit zwei Nachkommastellen) auf der Seite der Kriterien andererseits bestimmt. Die entsprechende Korrelation beträgt $r=.47$, sodass der zuerst für die Kategorien berichtete Wert in Höhe von $r=.46$ nicht auf die genannten Artefakte zurückgeführt werden kann.

3.2 Nutzen des Assessment-Centers

Der Zusammenhang zwischen den Assessment-Center-Ergebnissen und dem Erfolg an der Polizei-Führungsakademie ist deutlich zu erkennen – gleichwohl gibt es auch Fälle, bei denen die Prognosen auf Grund des Assessment-Centers sich nicht bewährt haben. Lohnt sich angesichts dieser nicht perfekten Ergebnisse der mit dem Assessment-Center verbundene (finanzielle) Aufwand überhaupt?

3.2.1 Nutzen der Personalauswahl in Bezug auf die Ausbildungskosten

Das Assessment-Center unterstützt die Polizei bei der Entscheidung, welche Bewerber zum Aufstieg in den höheren Polizeivollzugsdienst zugelassen werden. Zu Beginn der Laufbahn des höheren Dienstes steht in der Regel eine zweijährige Ausbildung, ein Jahr davon als Studium an der Polizei-Führungsakademie. 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. Kurzfristig gilt es aber, mit dem Assessment-Center die Investition der zweijährigen Ausbildung abzusichern. Wie hoch ist diese Investition?

Der Polizei-Führungsakademie standen im Jahr der Analyse pro Jahr 6,4 Millionen Euro für die Ausbildung zur Verfügung. Ein Kalenderjahr umfasst in der Regel 204 Studientage, ein Studientag kostet somit rund 31.000 Euro. Pro Jahr befinden sich ca. 150 Führungskräfte in der Ausbildung, d. h. die Ausbildung einer Führungskraft kostet pro Tag ca. 200 Euro. Ein Studienjahr mit 204 Tagen kostet somit rund 41.000 Euro pro Person. Die Kosten des zweiten Ausbildungsjahres sind schwerer zu berechnen, da die angehenden Führungskräfte hier (individuell) verschiedene Praxis-Stationen durchlaufen. Eine konservative Schätzung liegt bei 29.000 Euro pro Person und Jahr. Beide Ausbildungsjahre zusammen kosten somit 70.000 Euro allein an Ausbildungskosten pro Person.

Hinzu kommen die Personalkosten, denn die Beamten erhalten während ihrer Ausbildung weiterhin ihre Bezüge. Die Kosten variieren u. a. in Abhängigkeit von Dienstgrad und Alter. Ein typischer Fall dürfte ein Beamter der Besoldungsgruppe A 11 sein. Auf seine durchschnittlichen Jahresbezüge von damals 39.014 Euro sind jährlich jeweils 30 Prozent (1) Personalgemeinkosten (=11.704 Euro) sowie (2) Versorgungszuschläge (weitere 11.704 Euro) aufzurechnen, macht insgesamt 62.422 Euro. Veranschlagt man weitere Personalnebenkosten (z. B. Beihilfe) mit 2.129 Euro erhält man eine Summe von 64.551 Euro pro Jahr. Für beide Ausbildungsjahre somit aufgerundet ca. 130.000 Euro pro ausgebildeter Person.

Der Einfachheit halber bleiben zahlreiche weitere Kosten (Vertretungskosten, Reisekosten usw.) hier unberücksichtigt, sodass die hier aufgestellte Analyse die tatsächlichen Kosten unterschätzt.

Pro Person ergeben sich für die zweijährige Ausbildung somit ca. 70.000 Euro Ausbildungskosten und ca. 130.000 Euro Personalkosten, insgesamt also Kosten in Höhe von ca. 200.000 Euro pro Person. Bezogen auf die Analysegruppe von 112 Personen wird mit dem Assessment-Center somit eine Investition von ca. 22,4 Millionen Euro abgesichert.

Demgegenüber sind die Kosten für das Assessment-Center in Rechnung zu bringen. Ebenfalls vereinfachend gehen wir hier nur von den Kosten für das externe Beratungsunternehmen aus und übergehen die internen Personalkosten (Durchführungskosten sowie ex ante und ex post Kosten) sowie weitere Kosten durch z. B. logistischen Aufwand (Hotelkosten, Fahrtkosten usw.). Somit werden auch die Kosten des Assessment-Centers unterschätzt. Die Rechnung geht von einer Selektionsquote von .20 aus, d. h. jeweils einer von fünf Bewerbern wird ausgewählt. Bei der Vergabe der in der vorliegenden Analyse zur Verfügung stehenden 112 Plätze wurden dementsprechend 560 Personen mithilfe des Assessment-Centers begutachtet. Die Kosten für das mit der Assessment-Center-Durchführung beauftragte externe Beratungsunternehmen können für den Zeitpunkt der Analyse mit 1.500 Euro pro Person veranschlagt werden, bei 560 Personen ergibt dies eine Kostensumme von 840.000 Euro. Der abzusichernden Investition von ca. 22,4 Millionen Euro stehen

somit Kosten in Höhe von ca. 840.000 Euro für die Investitionsabsicherung entgegen, dies entspricht 3,75 % der abzusichernden Investition.

Zwei der drei Personen, welche die Ausbildung nicht bestanden haben, wurden zugelassen, obwohl sie im Assessment-Center vergleichsweise schlecht abgeschnitten hatten („den Anforderungen nur teilweise entsprechend“). Allein durch diese beiden Personen ist eine Fehlinvestition in Höhe von 400.000 Euro entstanden, die vermeidbar gewesen wäre, hätte man sich konsequenter an den Ergebnissen des Assessment-Centers orientiert. Positiv formuliert: Am Assessment-Center haben mehrere Hundert Personen teilgenommen, deren Verhalten und Leistung nicht überzeugen konnte. Die vorliegende Analyse zeigt, dass ein Großteil dieser Personen – wären sie zugelassen worden – in der Ausbildung mit hoher Wahrscheinlichkeit schlecht abgeschnitten hätte. Diese Fehlinvestitionen konnten durch das Assessment-Center vermieden werden.

3.2.2 Nutzen der Personalauswahl in Bezug auf die berufliche Leistung

Bislang wurden die Kosten des Assessment-Centers der abzusichernden Investition der Ausbildungskosten gegenübergestellt. Mit dem Assessment-Center soll aber nicht nur der Ausbildungserfolg, sondern letztendlich auch der *Berufserfolg* vorhergesagt werden. Der Berufserfolg ist ein komplexes und uneinheitlich definiertes Kriterium. Unbestritten dürfte aber sein, dass gute Ausbildungsleistungen nicht mit dem Berufserfolg gleichzusetzen sind; der Erfolg an der Polizei-Führungsakademie stellt also nicht das „*eigentliche*“ (ultimate) Kriterium dar.

Die monetären Konsequenzen der Streuung der beruflichen Leistung können vergleichsweise schlechter beziffert werden als die Ausbildungskosten. In Abschnitt 2.2 wurden allerdings die weitestgehend verallgemeinerbaren Erkenntnisse bisheriger Forschung dargestellt: Der jährliche Wert des Outputs eines durchschnittlichen Mitarbeiters ist in der Regel doppelt so hoch wie sein jährliches Entgelt. Und die Standardabweichung der in Geldwert umgerechneten Leistung zwischen überdurchschnittlichen und durchschnittlichen Mitarbeitern beträgt zwischen 40 % und 70 % des jährlichen Entgelts.

Weiter oben wurde erläutert, dass das jährliche Entgelt für einen durchschnittlichen Mitarbeiter des gehobenen Dienstes bei etwas unter 40.000 Euro liegt. Nach der Zulassung befinden sich die Polizisten im höheren Dienst. Für diese Laufbahn wird hier ein jährliches Entgelt von 50.000 Euro veranschlagt. Dieser statische Wert kann als Mittelwert des mit der Zeit zunehmenden Entgelts angesehen werden, das im höheren Dienst erzielt wird. Entsprechend der Regel, dass der jährliche Wert des Outputs eines durchschnittlichen Mitarbeiters in der Regel doppelt so hoch ist wie sein jährliches Entgelt, gehen wir von einem Output des durchschnittlichen Mitarbeiters in Höhe von 100.000 Euro aus. Die Standardabweichung der in Geldwert umgerechneten Leistung zwischen überdurchschnittlichen und durchschnittlichen Mitar-

beitern beträgt zwischen 40 % und 70 % des jährlichen Entgelts, im Folgenden wird von 40 % Streuung ausgegangen. Mit dieser somit konservativen Schätzung wird der tatsächliche Nutzen des Auswahlverfahrens eher unterschätzt. Ausgehend von einer 40 %-Streuung leistet ein überdurchschnittlicher Mitarbeiter einen Output in Höhe von 140.000 Euro pro Jahr, ein unterdurchschnittlicher Mitarbeiter einen Output in Höhe von 60.000 Euro pro Jahr. Wenn es mit dem Auswahlverfahren allein gelingt, durchschnittliche statt schwache Mitarbeiter einzustellen, ergibt sich eine Output-Differenz von 40.000 Euro pro Mitarbeiter und Jahr. Gelingt es mit dem Personalauswahlverfahren überdurchschnittliche an Stelle von schwachen Mitarbeitern zu identifizieren, beträgt die Output-Differenz 80.000 Euro im Jahr. Der öffentliche Dienst zeichnet sich durch eine hohe Verweildauer aus. Geht man von einer Verweildauer von 25 Jahren aus, beträgt der Geldwert des Leistungsunterschieds *pro Mitarbeiter* mindestens (Differenz Durchschnitt zu leistungsschwach) eine Million Euro und maximal (Differenz leistungsstark zu -schwach) zwei Millionen Euro. Mit der in Abbildung 1 dargestellten Formel lässt sich nun ausrechnen, welchen Nutzenzuwachs das Assessment-Center gegenüber einer Zufallsauswahl erbringt. Dabei wird von den folgenden Kennwerten ausgegangen:

N_A	112 eingestellte Personen
T	25 Jahre Verweildauer in der Organisation
r_{xy}	Gültigkeit des Assessment-Centers = .46 (dies war der Wert, der sich in Bezug auf den Ausbildungserfolg ergeben hat, siehe oben)
$\bar{z}_x$	2,17 ist der empirisch ermittelte Wert, den die letztlich ausgewählten Bewerber im Assessment-Center erzielt hatten
SD_y	die Standardabweichung wird auf 40.000 Euro angesetzt (siehe oben)
C	1.500 Euro Kosten des Assessment-Centers pro Bewerber
N_B	560 Assessment-Center-Teilnehmer

Eingesetzt in die Formel laut Abbildung 1 ergibt sich durch den Einsatz des Assessment-Centers gegenüber der Zufallsauswahl ein Nutzenzuwachs in Höhe von ca. 111 Mio. Euro (siehe Abbildung 3). Im Vergleich wird der Nutzenzuwachs für das unstandardisierte Interview berechnet. Die Validität des unstandardisierten Interviews wird entsprechend der bei Schmidt und Hunter (1998) berichteten Werte auf .38 geschätzt (siehe Abschnitt 2.2). Die Kosten für das Assessment-Center werden mit 1.500 Euro pro Person ($C_{\text{Assessment-Center}} = 1.500$) sechsmal so hoch veranschlagt wie die Kosten des unstandardisierten Vorgehens ($C_{\text{unstandardisiertes Eignungsinterview}} = 250$). Es wird also davon ausgegangen, dass das unstandardisierte Interview mit Kosten in Höhe von 250 Euro pro Person kostengünstiger ist als das Assessment-Center (1.500 Euro pro Person), dafür aber weniger treffsichere Aussagen erlaubt. Hinsichtlich aller übrigen Parameter wird von den gleichen Bedingungen für die beiden Verfahren ausgegangen. Setzt man nun die Werte in die oben genannte Gleichung ein, so ergibt

sich ein Nutzenvorteil des Assessment-Centers gegenüber dem unstandardisierten Interview in Höhe von fast 19 Mio. Euro (siehe Abbildung 3).

Abbildung 3: Nutzenzuwachs (Zuwachs gegenüber einer Zufallsauswahl) bei der Anwendung des ...

Nutzenzuwachs (Zuwachs gegenüber einer Zufallsauswahl) bei der Anwendung des...	
a) Assessment-Centers =	$112 * 25 * 0.46 * 2.17 * 40.000 - 1500 * 560 = 110.958.400$, also ca. 111 Mio. Euro
b) unstandardisierten Eignungsinterviews=	$112 * 25 * 0.38 * 2.17 * 40.000 - 250 * 560 = 92.215.200$ also ca. 92 Mio. Euro
Nutzenvorteil Assessment-Center gegenüber unstandardisiertem Interview: Differenz: $110.958.400 - 92.215.200 = 18.743.200$ also ca. 19 Mio. Euro	

4 Fazit

Der Wertschöpfungsgedanke sollte auch im Bereich der Personalauswahlverfahren Beachtung finden. Stellt man die Kosten der Verfahren ihrem Nutzen gegenüber, so zeigt sich häufig, dass die vermeintlich preisgünstigen Varianten die Organisation langfristig teuer zu stehen kommen. Aber es gilt keinesfalls: je teurer, desto besser. Auf dem Markt der Personalauswahlverfahren tummeln sich auch im Hochpreissegment Scharlatane; in diesem Markt besteht kein eindeutiger Zusammenhang zwischen Qualität und Kosten. Die Qualitätskontrolle im Einzelfall bleibt keinem Personaler erspart. Eine Hilfestellung zur Qualitätskontrolle bietet beispielsweise die Checkliste zur DIN 33430 (Kersting, 2008, freier Download der Checkliste 1 unter: www.kersting-internet.de/DIN-Buch). Weiß man um die Qualität seiner Personalauswahl (und glaubt nicht nur an diese), so kann man ihren Nutzen in Zahlen fassen und in Euro beziffern. Eine qualitativ hochwertige Personalarbeit ist wie kaum eine andere Maßnahme geeignet, die wirtschaftliche Situation der Organisation langfristig zu verbessern. Finanzielle Nutzenerwägungen sollten zusätzlich zu den weiteren Vorteilen von hochwertigen Personalauswahlverfahren (z. B. Erhöhung der Zufriedenheit, Schutz vor Über- und Unterforderung, Zusatznutzen wie Auswirkungen auf die Personalentwicklungsarbeit usw.) offensiv herausgestellt werden. Die Investition in die Personalauswahlverfahren sollte in expliziter Konkurrenz zu den anderen Investitionen der Organisation gesehen und gerechnet werden. Eine nüchterne Analyse wird häufig zu dem Ergebnis führen, dass der Nutzen qualitativ hochwertiger Personalauswahlverfahren den Nutzen anderer Investitionen der Organisation übersteigt.

Alle Organisationen sollten die Wirtschaftlichkeit der Verfahren, die in der eigenen Organisation bislang präferiert werden, mit der Wirtschaftlichkeit möglicher Verfahrensalternativen vergleichen.

Literatur

- Boudreau, J. W. (1989). Selection utility analysis: A review and agenda for future research. In M. Smith & I. T. Robertson (Hrsg.), *Advances in selection and assessment* (S. 227-257). Chichester, West Sussex: Wiley.
- Brogden, H. E. (1949). When testing pays off. *Personal Psychology*, 2, 171-183.
- Cronbach, L. J. & Gleser, G. C. (1965). *Psychological tests and personal decisions*. Chicago II: University of Illinois Press.
- Cronshaw, S. F. & Alexander, H. A. (1985). *One answer to the demand for accountability: Selection utility as an investment decision. Organizational Behavior and Human Decision Processes*.
- Detle, D. E. & Abele, A. E. & Renner, O. (2004). Zur Definition und Messung von Berufserfolg. *Zeitschrift für Personalpsychologie*, 3, 170-183.
- DIN. (2002). *DIN 33430: Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen*. Berlin: Beuth.
- Etzel, St. & Küppers, A. (2002). *Innovative Management-Diagnostik*. Göttingen: Hogrefe.
- Görlich, Y. & Schuler, H. (2006). Personalentscheidungen, Nutzen und Fairness. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (2. Aufl., S. 797-840). Göttingen: Hogrefe.
- Höft, S. (2001). Erfolgüberprüfung personalpsychologischer Arbeit. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 617-651). Göttingen: Hogrefe.
- Holling, H. (1998). Utility analysis of personnel selection. An overview and empirical study based on objective performance measures. *Methods of Psychological Research Online*, 3, 5-24.
- Hülshager, U. R. & Maier, G. W. & Stumpp, T. (2007). Validity of general mental ability for the prediction of job performance and training success in Germany: A meta-analysis. *International Journal of Selection and Assessment*, 15, 3-18.
- 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. (2004). Kosten und Nutzen beruflicher Eignungsbeurteilungen. In L. F. Horne & U. Winterfeld (Hrsg.), *Eignungsbeurteilungen auf dem Prüfstand: DIN 33430 zur Qualitätssicherung* (S. 55-77). Heidelberg: Spektrum Akademischer Verlag.
- Kersting, M. (2007). Wenn Tests in die Jahre kommen. Probleme des Einsatzes überalterter Testverfahren. In C. Lorei (Hrsg.), *Polizei und Psychologie* (S. 565-577). Frankfurt: Verlag für Polizeiwissenschaft.
- Kersting, M. (2008). *Qualität in der Diagnostik und Personalauswahl: Der DIN Ansatz*. Göttingen: Hogrefe.
- Kramer, J. (2009). Allgemeine Intelligenz und beruflicher Erfolg in Deutschland. Vertiefende und weiterführende Metaanalysen. *Psychologische Rundschau*, 60, 82-98.

- Krause, D. & Kersting, M. (2008). „Is it all about intelligence?“ Kognitive Fähigkeitstests und Assessment Center: Ein neuer Blick auf eine kontrovers diskutierte Befundlage. In W. Sarges & D. Scheffer (Hrsg.), *Innovative Ansätze der Eignungsdiagnostik* (S. 343-352). Göttingen: Hogrefe.
- Krause, D. E. & Kersting, M. & Heggstad, E. D. & Thornton, G. C. (2006). Criterion validity of assessment centers and cognitive ability tests. An empirical study on the executive management level. *International Journal of Selection and Assessment*, 14, 360-371.
- Organ, D. W. & Podsakoff, P. M. & MacKenzie, S. B. (2006). *Organizational citizenship behavior. Its nature, antecedents and consequences*. Thousand Oaks, CA: Sage Publications.
- Rowold, J. (2007). Excel-Tabelle zur praktischen Durchführung von Kosten-Nutzen-Analysen für Assessment Center. In A. Süßmair & J. Rowold (Hrsg.), *Kosten-Nutzen-Analysen und Human Resources* (S. 176-180). Weinheim: Beltz.
- Rowold, J. & Mönninghoff, M. (2007). Kosten-Nutzen-Analyse von Assessment Centern: Fallbeispiele aus Call Centern. In A. Süßmair & J. Rowold (Hrsg.), *Kosten-Nutzen-Analysen und Human Resources* (S. 51-66). Weinheim: Beltz.
- 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.
- Scherm, M. (2005). *360-Grad-Beurteilungen*. 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 findings. *Psychological Bulletin*, 124, 262-274.
- Schmidt, F. L. & Hunter, J. E. & Pearlman, K. (1982). Assessing the economic impact of personnel programs of workforce productivity. *Personnel Psychology*, 35, 333-347.
- Schuler, H. (2004). Leistungsbeurteilung - Gegenstand, Funktionen und Formen. In H. Schuler (Hrsg.), *Beurteilung und Förderung beruflicher Leistung* (2. Aufl., S. 1-23). Göttingen: Hogrefe.
- Schuler, H. & Funke, U. & Moser, K. & Donat, M. (1995). *Personalauswahl in Forschung und Entwicklung. Eignung und Leistung von Wissenschaftlern und Ingenieuren*. Göttingen: Hogrefe.
- Stehle, W. & Barthel, E. (1984). Lohnen sich psychologische Auswahlverfahren? Eine Nutzen-Kosten-Analyse. *Personalwirtschaft*, 11/84, 381-386.
- Stephan, U. & Westhoff, K. (2002). Personalauswahlgespräche im Führungskräftebereich des deutschen Mittelstandes: Bestandsaufnahme und Einsparungspotenzial durch strukturierte Gespräche. *Wirtschaftspsychologie*, 4, 3-17.

Andreas Gourmelon/Christine Kirbach/
Stefan Etzel (Hrsg.)

2009

Personalauswahl im öffentlichen Sektor

2. erweiterte und überarbeitete Auflage



Nomos