

**Man kann nicht immer alles haben?
Vereinbarkeit von Validität und Akzeptanz
bei eignungsdiagnostischen Auswahlverfahren**

Dissertation

zur Erlangung des akademischen Grades

eines Doktors der Sozial- und Erziehungswissenschaften

„Doctor philosophiae“

der Universität Hildesheim

vorgelegt von

Nils Benit

(geb. 07.03.1981 in Bad Oldesloe)

Dekan der Fakultät für Sozial- und Erziehungswissenschaften der Universität Hildesheim Prof. Dr. Werner Greve

Vorsitz: Prof. Dr. Karl-Heinz Arnold

Gutachterin: Prof. Dr. Renate Soellner

Gutachter: Prof. Dr. Andreas Mojzisch

Tag der Disputation: 03.04.2013

Inhaltsverzeichnis

Zusammenfassung	iv
Abstract	viii
1 Einleitung	1
2 Forschungshintergrund	7
2.1 Das Assessment Center	7
2.2 Psychologische Testverfahren	9
2.2.1 Intelligenz	11
2.2.2 Persönlichkeit	15
2.3 Verwendungshäufigkeiten von eignungsdiagnostischen Auswahlverfahren	18
2.4 Akzeptanz von eignungsdiagnostischen Auswahlverfahren	19
3 Offene Forschungsfragen	25
4 Abschlussdiskussion	36
5 Fazit	53
Literatur	55
Danksagung	69
Eigenständigkeitserklärung	71
Anhang	72

Zusammenfassung

In Zeiten zunehmender Globalisierung und geringerer Verfügbarkeit von Fachkräften nimmt die Bedeutung einer fundierten Eignungsdiagnostik stetig zu (Kersting, 2011a). Dabei ist die unternehmerische Entscheidungsfindung zur Implementierung eines Auswahlverfahrens nicht nur von dessen prognostischer Vorhersagegüte für spätere berufliche Erfolgskriterien geleitet (König, Klehe, Berchtold & Kleinmann, 2010). Zunehmend rückt außerdem die wahrgenommene Akzeptanz des Verfahrens aus Bewerberperspektive in den Fokus der Unternehmensentscheidung (Anderson, 2004; Ryan & Huth, 2008). In einer Reihe von Studien konnte indes gezeigt werden, dass die meistakzeptierten Auswahlverfahren nicht gleichzeitig auch die validesten sind (Schuler, Hell, Trapmann, Schaar & Boramir, 2007). Während das Assessment Center (AC) gemäß Metaanalysen – insbesondere unter Berücksichtigung des zeitlichen und monetären Aufwands – als lediglich moderat valide zu beurteilen ist (Hermelin, Lievens & Robertson, 2007), erfreut es sich weltweit aufgrund seiner multimethodalen sowie berufsbezogenen Aufgabengestaltung steigender Beliebtheit und Verwendungshäufigkeit (Höft & Obermann, 2010). Im Gegensatz dazu vermögen psychologische Testverfahren (Intelligenz- und Persönlichkeitstests) berufliche Erfolgskriterien überaus valide vorherzusagen (Barrick & Mount, 1991, 2005; Kramer, 2009; Salgado & Anderson, 2003; Schmidt & Hunter, 1998), werden aber aufgrund ihrer häufig abstrakten und kontextfernen Gestaltung als wenig akzeptiert bewertet (Hausknecht, Day & Thomas, 2004) und infolgedessen nur selten zu eignungsdiagnostischen Zwecken eingesetzt (Ryan, McFarland, Baron & Page, 1999). Daher wird im Rahmen dieser Arbeit eine wichtige Erweiterung bisheriger Forschung in diesem Themenfeld vorgenommen, indem untersucht wird, wie sowohl für das AC als auch für psychologische Testverfahren die Vereinbarkeit von Validität und Akzeptanz realisiert werden kann.

So wird in der ersten Studie metaanalytisch die prognostische Validität von ACs aus dem deutschsprachigen Raum untersucht, um bewerten zu können, ob die hohe Verwendungshäufigkeit auch unter dem Aspekt der prognostischen Validität gerechtfertigt erscheint. Die analysierte mittlere korrigierte Validität von $\rho = .42$ weist auf eine zufriedenstellende Güte der untersuchten ACs hin, die allerdings psychologischen Testverfahren (insbesondere Intelligenztests) weiterhin hinsichtlich der Prognosefähigkeit unterlegen ist. Der Zertifizierungsprozess nach DIN 33430 (Deutsches Institut für Normung e.V., 2002) erwies sich in der Untersuchung als bedeutsamer Moderator in dem Sinne, dass Verfahren aus bereits zertifizierten oder sich noch im Zertifizierungsprozess befindenden Unternehmen eine höhere Validität aufwiesen als Verfahren aus Unternehmen, bei denen ein Zertifizierungsprozess noch nicht begonnen hatte ($\rho = .47$ vs. $\rho = .38$).

In Studie 2 wird die erste Bestandsaufnahme von verwendeten Vorauswahlmethoden zur Nominierung potenzieller AC-Kandidaten¹ vorgenommen. Berücksichtigt man neben der lediglich moderaten Validität der ACs auch ihre kosten- und ressourcenintensive Konzeption sowie Durchführung (Gierschmann, 2005), erscheint die Forderung nach validen und ökonomischen Vorauswahlmethoden notwendig, um die Wirtschaftlichkeit von ACs weiterhin zu sichern (Schuler, 2007a; Warmke, 1985). Die Ergebnisse der Studie weisen darauf hin, dass fast alle deutsche Unternehmen *Bewerbungsunterlagen* zur Vorauswahl heranziehen, während validere Methoden wie *strukturierte Interviews* oder *psychologische Testverfahren* nur selten in diesem Selektionsschritt Anwendung finden. Zudem ergaben weitere Analysen, dass die Anzahl der verwendeten Vorauswahlmethoden von der Unternehmensgröße, dem Zertifizierungsstatus nach DIN 33430 und der diagnostischen Kompetenz der im Vorauswahl-Prozess verantwortlichen Personen abhängt. Eine höhere Anzahl an verwendeten Vorauswahlverfahren indiziert, dass auch die Wahrscheinlichkeit für den Einsatz valider Vorauswahlmethoden (z.B.: Interview, Testverfahren) ansteigt.

¹ Aus Gründen der besseren Lesbarkeit wird an Stelle der Doppelbezeichnungen für Personengruppen lediglich die männliche Form verwendet.

In Studie 3 wird geprüft, ob die berichtete Zurückhaltung deutscher Unternehmen Intelligenz- sowie Persönlichkeitstests im Rahmen der Eignungsdiagnostik einzusetzen (Ryan et al., 1999; Schuler et al., 2007), weiterhin besteht. Neben einer aktuellen Erfassung der Einsatzhäufigkeit von psychologischen Testverfahren für unterschiedliche Zielgruppen und Einsatzzwecke wird untersucht, ob wissenschaftlich generiertes Wissen über die prognostische Güte von Auswahlverfahren die Praxis erreicht hat und welche weiteren Faktoren den Einsatz von Testverfahren limitieren. Die Ergebnisse unterstreichen, dass deutsche Unternehmen psychologische Testverfahren im Vergleich zu anderen Ländern immer noch verhältnismäßig selten einsetzen (Ryan et al., 1999). Als bedeutsamster Beweggrund wurde von den befragten Personalverantwortlichen die geringe Augenscheinvalidität herkömmlicher Testverfahren bewertet. Weiterführende Analysen zeigten, dass insbesondere die Intransparenz der Testqualität den Einsatz beeinflusst. Mittels einer Datenbankrecherche wurde zusätzlich überprüft, ob wissenschaftliche Befunde über die Güte psychologischer Testverfahren durch das Medium *Zeitschrift* (als häufig genutzte Informationsquelle für Praktiker) auch in Deutschland nur unzureichend in die Praxis transportiert werden (vgl. Anderson, 2007; Rynes, Giluk & Brown, 2007). Hierfür wurden seit 2005 veröffentlichte Beiträge aus fünf praxisorientierten Zeitschriften für Human Resource Management und als Referenzmaßstab die Beiträge aus drei personalpsychologischen Fachzeitschriften hinsichtlich der Beitragshäufigkeit und Forschungsevidenz der Beiträge untersucht. Die Recherche ergab, dass in den Praktikermagazinen ein unzureichender Transfer wissenschaftlicher Befunde stattfindet, was als weiterer Grund für den geringen Praxiseinsatz dieser Verfahrensklasse angeführt werden kann.

Die vierte Studie führte die aus Studie 3 gewonnenen Erkenntnisse weiter und untersucht, wie die aus Unternehmenssicht so bedeutsame Augenscheinvalidität von Testverfahren durch Manipulation des Testmaterials optimiert werden kann. Es wurde daher geprüft, ob die Akzeptanz eines Matrizen-tests allein durch eine berufsbezogene Symbolgestaltung erhöht werden kann und der Test trotzdem testtheoretischen Gütek-

riterien genügt. Hierfür wurde der BOMAT (Hossiep, Turck & Hasella, 2001), ein valider, aber abstrakter Matrizen test, mit dem BBM (berufsbezogener Matrizen test), einem für Forschungszwecke entwickelten Matrizen test mit berufsbezogenen Symbolen, kontrastiert. Beide Verfahren lassen sich als reliabel und valide bewerten. Die berufsbezogene Gestaltung erwies sich aus Bewerbersicht hinsichtlich der Augenscheinvalidität als überlegen, bei gleichwertiger Messqualität, Kontrollierbarkeit und Belastungsfreiheit (AKZEPT!-L: Kersting, 2008). In Folge dieser Befunde wurde ein berufsbezogener kognitiver Leistungstest (ANAKONDA) konstruiert und entwickelt, der sich in ersten Validierungsstudien als reliabler und valider Test erwies und eine höhere Akzeptanz als ein allgemeiner Intelligenztest (IST-2000R: Amthauer, Brocke, Liepmann & Beauducel, 2001) erlangte.

Die Gesamtheit der Befunde aus allen Manuskripten wird diskutiert sowie Implikationen der Ergebnisse für die weitere Forschung und die Praxis werden abgeleitet. Zusammenfassend kann konstatiert werden, dass die vorliegende Arbeit eine Erweiterung bisheriger Untersuchungen in diesem Themenfeld darstellt, da sie einerseits neue Erkenntnisse über die Generalisierbarkeit der prognostischen Validität von Assessment Centern in deutschen Unternehmen sowie deren eingesetzten Vorauswahlmethoden hervorbringt. Andererseits wird neues Wissen über die Einsatzhäufigkeit bzw. das Verständnis von psychologischen Testverfahren in deutschen Unternehmen, den Transfer wissenschaftlicher Erkenntnisse in die Praxis sowie Möglichkeiten zur Akzeptanzerhöhung dieser Verfahrensklasse generiert.

Schlüsselwörter: Akzeptanz, Assessment Center, Intelligenz, Personalauswahl, Persönlichkeit, psychologische Testverfahren, Validität

Abstract

In times of increasing globalization and low availability of specialists, the importance of well-founded aptitude diagnostics is ever increasing (Kersting, 2011a). When companies decide on implementing a selection procedure, not only the predictive validity of such procedures is crucial for entrepreneurial success (König, Klehe, Berchtold & Kleinmann, 2011), but also the applicant's perceived acceptance of such selection procedures (Anderson, 2004; Ryan & Huth, 2008). Numerous studies reveal that the most accepted selection procedures are not the most valid (Schuler, Hell, Trapmann, Schaar & Boramir, 2007). Meta analyses indicate that Assessment Centers (AC) can only be evaluated as moderately valid, especially, when amount of economic expenditure and time are considered (Hermelin, Lievens & Robertson, 2007). However, this selection procedure is globally increasing in popularity and is employed with rising frequency due to its multi-method approach and job-oriented task design (Höft & Obermann, 2010). In contrast, psychological tests such as intelligence and personality tests are capable of predicting job success factors exceedingly well (Kramer, 2009; Salgado & Anderson, 2003; Schmidt & Hunter, 1998), but are regarded as less acceptable due to their abstract and not context-derived design (Hausknecht, Day & Thomas, 2004). Therefore, they are rarely used for aptitude diagnostics (Ryan, McFarland, Baron & Page, 1999). Thus, an important expansion of previous research will be carried out in this paper when possibilities to increase compatibility of validity and acceptance for both ACs as well as psychological tests are being analyzed.

In the first study, the predictive validity of ACs in German speaking regions will be meta-analytically examined. This way we can assess whether the high frequency of using ACs is justifiable taking into account its predictive validity. The analyzed mean-corrected validity of $\rho = .42$ exhibits a satisfying quality of the examined ACs. However psychological tests (especially intelligence tests) exceed ACs by far in their predictive capability. In this study, the certification process of DIN 33430 (German Institute for

Standardization, 2002) proved to be an important moderator. ACs which had either already been certified or were in the certification process exhibited a higher validity than ACs from companies that have not started the certification process ($\rho = .47$ vs. $\rho = .38$).

Study 2 presents the first survey of employed pre-selection methods for the nomination of potential AC candidates. With respect to not only the moderate validity of ACs but also their resource- and cost-intensive conception and execution (Gierschmann, 2005), valid and economical pre-selection methods will be imperative to ensure economical efficiency of ACs in the future (Schuler, 2007a; Warmke, 1985). The study's results showed that almost all German companies use *application documents* as a pre-selection method while more valid methods such as *structured interviews* or *psychological tests* are rarely used at this stage of selection. In addition, further analysis showed that the number of pre-selection methods used depends on company size, certification status of DIN 33430, and diagnostic competence of the person responsible for the pre-selection process. A higher number of pre-selection methods used indicated an increasing probability of implementing valid pre-selection methods (e.g., interviews, psychological tests).

In Study 3, we investigated into the hesitance of German companies to use intelligence and personality tests as part of the personnel selection (Ryan et al., 1999; Schuler et al., 2007). In accordance with a recent survey on the use of psychological tests for different target groups and purposes, we also examined if scientifically generated knowledge of the predictive validity of selection procedures has reached the practicing field. Which factors limit the use of psychological tests was a further research question. The results underlined that German companies continue to employ psychological tests rarely compared to other countries (Ryan et al., 1999).

The personal advisors reported low face validity of conventional tests as the main reason for their rare use. Further analysis showed that low test employment was especially affected by hazy transparency of the test quality. By conducting database research, it was also examined whether scientific findings on the quality of psychological tests are

sufficiently transported into practice through the medium *journal* (an often used information source for practitioners) in Germany (see Anderson, 2007; Rynes, Giluk & Brown, 2007). For this purpose, published contributions since 2005 from five practice-oriented magazines for Human Resource Management and three personnel psychology journals were analyzed regarding frequency of contributions and research evidence. The findings indicated an insufficient transfer of scientific findings in the practitioner magazines compared to the personnel psychology journals. This might be an additional important reason for the rare utilization of psychological tests.

The fourth study based on the knowledge gained from Study 3 examined how the highly valued face validity of tests could be optimized by manipulating test material. We tested if the acceptance of a matrix test could be improved by using job-related symbols while still fulfilling test quality criteria (e.g., reliability, validity). For this reason, the BOMAT (Hossiep, Turck & Hasella, 2001), a valid but abstract matrix test was compared with the BBM (Business Based Matrix Test), a matrix test developed for research use with job-related symbols. Both instruments proved to be reliable and valid. The job-oriented design was superior in terms of face validity and equal value regarding perceived psychometric-quality, opportunity to perform, and perceived strain (AKZEPT!-L: Kersting, 2008).

In response to these findings, a completely new job-oriented intelligence test was developed (named ANAKONDA). This test has shown to be reliable and valid in first validation studies and has attained a higher acceptance level in contrast to a conventional intelligence test (IST-2000R: Amthauer, Brocke, Liepmann & Beauducel, 2001).

Abstract

Findings of all presented studies will be discussed in detail and implications from the results will be derived for further research and practice. In summary, this work extends previous research in the field by generating new knowledge on the generalization of predictive validity of ACs in German speaking regions and the used pre-selection methods at this selection stage. Moreover a deeper understanding of psychological tests regarding the frequency of use, the transfer of scientific findings into practice about their benefit and possibilities of increasing acceptance of these selection procedures in German companies is generated.

Keywords: acceptance, assessment center, intelligence, personality, personnel selection, psychological tests, validity

1 Einleitung

Technologische, wirtschaftliche und gesellschaftliche Veränderungen führten in den vergangenen Dekaden dazu, dass das *Humankapital* zunehmend in den Fokus der Unternehmenspolitik rückte. Unter Ökonomen herrschte Verblüffung, als der Begriff *Humankapital* als Unwort des Jahres 2004 gekürt wurde (Leitl, 2007). Sprachkritiker verwiesen auf die Gefahr, dass diese Begrifflichkeit Menschen zunehmend zu nur noch ökonomisch interessanten Größen degradieren würde. Eigentlich verbirgt sich jedoch folgende Idee hinter dem Begriff *Humankapital*: „Das wertvollste Kapital ist das in den Menschen investierte“ (Alfred Marshall, 1890; zitiert nach Dürndorfer, Nink & Wood, 2005, S. 16). Aus welcher Perspektive man nun auch den Begriff *Humankapital* betrachten mag, angesichts veränderter Gegebenheiten (z.B.: Globalisierung, demografischen Wandel) wird ersichtlich, dass der Faktor Mensch künftig mehr denn je den Erfolg eines Unternehmens maßgeblich mitbestimmen wird. Demzufolge gehört die Personalauswahl zu einer der wichtigsten Aufgaben des Personalmanagements in Unternehmen; denn personelle Fehlentscheidungen können zu erheblichen monetären Einbußen auf unternehmerischer Seite führen und den Unternehmenserfolg gefährden. Mittels eignungsdiagnostischer Auswahlverfahren versuchen Unternehmen folglich, die am besten geeigneten Kandidaten für zu besetzende Aufgabenfelder zu identifizieren. Nachvollziehbar erscheint dementsprechend die Forderung des Berufsverbandes Deutscher Psychologinnen und Psychologen (2009), dass Personalauswahlentscheidungen mit mindestens der gleichen Sorgfalt getroffen werden sollten, wie sie Unternehmen zum Beispiel bei der Investition von Maschinen an den Tag legen.

In den letzten zehn Jahren ist zu erkennen, dass nicht nur die Häufigkeit fundierter Eignungsdiagnostik zunimmt, sondern auch ihre Bedeutung. Infolge des demografischen Wandels sowie misslungener Bildungsplanung und -politik wird zunehmend ein Fachkräftemangel erkennbar, der den Erfolg zahlreicher Organisationen infrage stellt

(Kersting, 2011a). Die steigende Nachfrage nach qualifizierten Mitarbeitern kann nur ungenügend abgedeckt werden. Aufgrund der geringen Grundquote geeigneter Kandidaten wird demzufolge eine fundierte Eignungsdiagnostik immer notwendiger. Während Organisationen vor einigen Jahren noch in der Situation waren, sich ausschließlich auf hoch qualifizierte und einschlägig fachlich ausgebildete Bewerber (sogenannte „High Potentials“) zu konzentrieren, können sie es sich nun nicht mehr leisten, Potenziale zu übersehen. Angenommen, es besteht ein Überangebot an hoch qualifizierten Bewerbern für eine zu besetzende Position, so nimmt die Diagnostik lediglich eine untergeordnete Rolle ein. Bewerber mit unzureichendem Wunschprofil (z.B.: zu alt, zu schlechte Abschlussnoten) werden vorselektiert, sodass im besten Fall ausschließlich geeignete Bewerber zur Auswahl stehen. In diesem Szenario gelingt es mit jedem Auswahlverfahren, eine exzellente Trefferquote zu erreichen. Selbst das Übersehen eines geeigneten Kandidaten wäre verzeihbar, da die Stelle in jedem Fall mit einem anderen geeigneten Bewerber besetzt werden könnte (Kersting, 2011b). In Zeiten des Fachkräftemangels müssen Organisationen allerdings ihre Auswahlstrategien den Gegebenheiten anpassen. Die zum Auswahlverfahren zugelassenen Bewerber werden sich mehr denn je in ihren Fähigkeitsprofilen unterscheiden, was Auswirkungen auf die Grundquote haben wird und valide Auswahlverfahren zur Identifizierung geeigneter Bewerber bzw. der entsprechenden Bewerberpotenziale notwendig macht.

In der umfassenden Metaanalyse von Schmidt und Hunter (1998) wurde die prognostische Validität von insgesamt 19 verschiedenen Auswahlverfahren auf unterschiedliche berufliche Erfolgsindikatoren hin untersucht. Es zeigte sich, dass psychologische Testverfahren (Intelligenz- und Persönlichkeitstests) zu den validesten Auswahlverfahren avancierten und weitaus besser beruflichen Erfolg vorherzusagen vermochten als das deutlich kosten- und ressourcenintensivere Assessment Center (AC). Betrachtet man im Gegensatz dazu Studien, die die Verwendungshäufigkeit von Auswahlverfahren untersuchten (vgl. Schuler et al., 2007; Hell, Boramir, Schaar & Schuler, 2006), wird deut-

lich, dass die validesten Auswahlverfahren nicht gleichzeitig die am häufigsten verwendeten sind. Es stellt sich somit die Frage, welche Faktoren die Entscheidungsfindung für oder gegen ein Auswahlverfahren auf Organisationsseite beeinflussen. Basierend auf der *institutional theory* untersuchten König et al. (2010) den Stellenwert der sechs Einflussfaktoren (1) Verbreitungsgrad des Verfahrens, (2) juristische Probleme im Zusammenhang mit dem Verfahren, (3) Nutzen für die organisatorische Eigenwerbung, (4) Akzeptanz, (5) prognostische Validität sowie (6) Kosten des Verfahrens im Hinblick auf die unternehmerische Entscheidungsfindung zur Implementierung eines Auswahlverfahrens. Die Ergebnisse unterstreichen, dass insbesondere die Akzeptanz eines Verfahrens bestimmt, ob es eingesetzt wird. Die prognostische Validität wurde in dieser Studie von den 506 befragten Personalverantwortlichen diesbezüglich als weniger relevant bewertet. Aus Organisationssicht erscheinen akzeptierte Auswahlverfahren besonders lohnenswert, da eine wenig akzeptierende Haltung und Wahrnehmung der Bewerber gegenüber dem Auswahlverfahren dazu führen kann, dass Bewerber das Organisationsansehen abwerten oder ein Stellenangebot ablehnen (Hausknecht et al., 2004). Auch negative Auswirkungen auf das Unternehmensimage (Ryan & Ployhart, 2000) oder auf das Arbeitsverhalten der eingestellten Bewerber (Gilliland, 1993) werden angeführt.

Die bisherigen Ausführungen veranschaulichen, dass eignungsdiagnostische Entscheidungen aufgrund der aktuellen gesellschaftlichen sowie weltwirtschaftlichen Situation immer bedeutsamer werden. Allerdings genügt es nicht nur valide (also treffsichere) Auswahlverfahren den Unternehmen zur Verfügung zu stellen. Insbesondere eine augenscheinvalide Passung zwischen Auswahlverfahren und künftiger Arbeitstätigkeit beeinflusst die Verfahrensakzeptanz (Hausknecht et al., 2004) und bestimmt maßgeblich, ob ein Auswahlverfahren tatsächlich in der Unternehmenspraxis verwendet wird (König et al., 2010). Die Vereinbarkeit von Validität und Akzeptanz innerhalb eines Verfahrens erscheint demgemäß erstrebenswert. Während das AC aufgrund der simulationsorientierten Gestaltung im Gegensatz zu den meist berufunspezifisch gestalteten Testverfah-

ren als akzeptiertes Auswahlverfahren bezeichnet werden kann (Hausknecht et al., 2004; Schuler et al., 2007), erwies es sich im Vergleich zu Intelligenz- und Persönlichkeitstests in entsprechenden Metaanalysen als weniger prognostisch valide (Schmidt & Hunter, 1998). Damit die Verfahrensklassen *AC* und *psychologische Testverfahren* hinsichtlich der beiden erstrebenswerten Dimensionen *Validität* und *Akzeptanz* gleichermaßen überzeugen können, bedarf es einer verfahrensspezifischen Optimierung, die im Rahmen der vorliegenden Arbeit herausgearbeitet werden soll.

Ziel ist es, differenzierte Analysen hinsichtlich der Akzeptanz bzw. Validität der Auswahlverfahren *Assessment Center* sowie *psychologische Testverfahren* vorzunehmen, um die Vereinbarkeit von Akzeptanz und Validität eignungsdiagnostischer Instrumente zu prüfen. Im ersten Schritt wird die prognostische Validität von ACs in deutschen Unternehmen metaanalytisch untersucht. Hierbei stellt sich die Frage, ob auch im deutschsprachigen Raum eine abnehmende Validität der AC-Methode erkennbar wird, wie sie für den anglo-amerikanischen Raum identifiziert wurde (Hermelin, Lievens & Robertson, 2007). Erst bei empirischer Bestätigung einer zufriedenstellenden Validität von ACs, lässt sich bewerten, ob die steigende Beliebtheit und Verwendungshäufigkeit von ACs in Deutschland (Höft & Obermann, 2010; Schuler et al., 2007) auch unter dem Aspekt der prognostischen Validität gerechtfertigt erscheinen. Um sowohl die prognostische Validität als auch die Wirtschaftlichkeit des ACs zu optimieren, ist der Einsatz von validen Vorauswahlmethoden besonders entscheidend (Schuler, 2007a; Warmke, 1985). Deshalb soll im zweiten Schritt die Verwendungshäufigkeit verschiedener Vorauswahlmethoden zur Nominierung potenzieller AC-Kandidaten detailliert analysiert werden. Hierbei wird untersucht, welche Vorauswahlmethoden in deutschen Unternehmen eingesetzt und wie diese von den befragten Personalverantwortlichen hinsichtlich ihrer Validität eingeschätzt werden.

Im dritten Schritt steht die Frage im Vordergrund, ob deutsche Unternehmen weiterhin im internationalen Vergleich Intelligenz- und Persönlichkeitstests selten einsetzen

(Ryan et al., 1999; Schuler et al., 2007). Es werden mehrere Gründe für die Zurückhaltung deutscher Unternehmen diskutiert. Eine geringe Augenscheinvalidität, juristische Probleme, Intransparenz der Testqualität und Verfügbarkeit, Verfälschbarkeit der Testergebnisse sowie Betriebsratsentscheidungen stellen nur einige Aspekte dar (vgl. hierzu u.a. Hausknecht et al., 2004; Hülsheger & Maier, 2008; Kersting, 2008; Marcus, 2003b; Muck & Stumpp, 2007). Bislang fehlen allerdings Studien, die eine Relevanzeinschätzung dieser Beweggründe vorgenommen haben und folglich eine Priorisierung aus Organisationsperspektive leisten können. Dementsprechend wird zunächst überprüft, ob psychologische Testverfahren weiterhin in deutschen Unternehmen vergleichsweise selten verwendet werden (Ryan et al., 1999). Dabei wird auch der Frage nachgegangen, welche potenziellen Beweggründe den Einsatz einschränken.

Generell wird aus wissenschaftlicher Sicht bemängelt, dass Forschungsergebnisse nicht ausreichend von Praktikern umgesetzt würden (*scientist-practitioner gap*: vgl. hierzu Anderson, 2007; Anderson, Herriot & Hodgkinson, 2001; Anderson, Lievens, van Dam & Ryan, 2004; Cohen, 2007; Rynes, 2007). Empirische Untersuchungen, welche die Gründe für diese Mängel erforschen, gibt es hingegen international äußerst selten (Highhouse, 2008; Pfeffer & Sutton, 2006) und auf den deutschen Sprachraum bezogen keine, die über eine theoretische Diskussion hinausgehen (Albrecht & Deller, 2011). Daher wird am Beispiel der Prognostizität psychologischer Testverfahren untersucht, wie der Transfer wissenschaftlicher Empfehlungen in die deutsche Unternehmenspraxis gelingt. Untersuchungen aus dem nordamerikanischen Raum zeigten, dass in Praktikermagazinen (als häufig genutzte Informationsquelle für Unternehmen), nur selten die Güte psychologischer Testverfahren thematisiert wurde (Rynes et al., 2007). Daher ist von Interesse, ob auch in Deutschland der Transfer von wissenschaftlich generiertem Wissen zur Güte psychologischer Testverfahren durch das Medium *Zeitschrift* nur unzureichend stattfindet, was als weiterer Erklärungsgrund für die geringe Verwendungshäufigkeit angeführt werden könnte. Da psychologische Testverfahren aufgrund ihrer häufig kontext-

unspezifischen Gestaltung eine eher geringe Akzeptanz erfahren (Hausknecht et al., 2004; Kramer, 2009; Kersting, 2008) und u.a. deshalb selten in der Eignungsdiagnostik eingesetzt werden (Schuler et al., 2007), wird viertens eine differenzierte Akzeptanzeinschätzung zwischen allgemeinen und berufsbezogenen Testvarianten vorgenommen. Es soll explizit untersucht werden, ob sich bereits eine kontext- bzw. berufsspezifische Symbolgestaltung in Matrizentests positiv auf die Akzeptanzeinschätzung der Testkandidaten auswirkt und dessen ungeachtet der Test eine zufriedenstellende psychometrische Güte aufweist. Darüber hinaus wurde im Rahmen des kongruenten Forschungsprogramms dieser Dissertation ein berufsbezogener kognitiver Leistungstest (ANAKONDA) konstruiert, der sich in ersten Validierungsstudien als reliabler und valider Test erwies und eine höhere Akzeptanz als ein herkömmlicher Intelligenztest (IST-2000R: Amthauer, Brocke, Liepmann & Beauducel, 2001) erlangte.

Die vorliegende Arbeit gliedert sich wie folgt: Nach der Einleitung (Kapitel 1) wird im zweiten Kapitel der aktuelle Forschungshintergrund dargestellt. Dabei wird eingangs das AC als multimethodales Auswahlverfahren vorgestellt und die bisherige empirische Befundlage zu deren Validität skizziert. Anschließend erfolgt eine Darstellung der psychologischen Testverfahren (Intelligenz- und Persönlichkeitstest) hinsichtlich ihrer Konzeption und Validität in Bezug auf berufsbezogene Erfolgsindikatoren. Abschließend wird in diesem Kapitel in das Konzept der Akzeptanz von Auswahlverfahren eingeführt sowie ein tragfähiges Akzeptanzmodell (Hausknecht et al., 2004) skizziert; zudem werden Möglichkeiten der Akzeptanzmessung dargestellt. Im dritten Kapitel werden die eigenen Analysen vorgestellt und offene Fragestellungen herausgearbeitet. Das vierte Kapitel enthält die Diskussion der Ergebnisse der Studien, wobei diese miteinander in Beziehung gesetzt und künftige lohnenswerte Forschungsbemühungen abgeleitet werden. Abschließend werden weiterführende Implikationen der Ergebnisse diskutiert. Mit einem Fazit bzgl. der wesentlichen Untersuchungsschritte und Ergebnisse schließt diese Arbeit.

2 Forschungshintergrund

2.1 Das Assessment Center

Das Assessment Center stellt ein multimethodales Auswahlverfahren dar, welches sich weltweit steigender Beliebtheit und Verwendungshäufigkeit erfreut (Höft & Obermann, 2010; Shackleton & Newell, 1994; Schuler et al., 2007). Es werden grundsätzlich a.) mehrere Mitarbeiter oder Bewerber b.) von mehreren geschulten Beobachtern c.) in mehreren Übungen (Multimethodalität) und d.) in einer Vielzahl von Beurteilungssituationen im Hinblick auf bestehende oder zukünftige Arbeitssituationen und Aufgabenfelder e.) nach festgelegten Regeln und Kriterien beurteilt (vgl. u.a. Lievens & Klimoski, 2001; Obermann, 2006; Sarges, 2001). Aufgrund der Vielfalt der eingesetzten Übungen handelt es sich beim AC um ein aufwendiges sowie kostenintensives Verfahren, welches anderen Auswahlverfahren (z.B.: Interview, Testverfahren) hinsichtlich Kosten-Nutzen-Aspekten unterlegen ist (Goffin, Rothstein & Johnston, 1996). Gleichwohl resultiert aus der möglichst genauen Passung dieser Übungen mit den Anforderungen der beruflichen Praxis eine hohe Augenscheinvalidität (Arthur, Woehr & Maldegen, 2000; Spychalski, Quiñones, Gaugler & Pohley, 1997), die einerseits in eine hohe Verfahrensakzeptanz mündet und andererseits große Erwartungen der Anwender hinsichtlich der Verfahrensvalidität weckt (Höft & Obermann, 2010).

Eine der ersten und berühmtesten Studien zur Validität eines ACs wurde in der *American Telephone and Telegraph Company* (AT & T) durchgeführt (Bray & Grant, 1966). In dieser Einzelstudie ergab sich für das untersuchte AC ein zufriedenstellender Validitätskoeffizient von $r = .46$. Erst in den 1980er Jahren legten Thornton, Gaugler, Rosenthal und Bentson (1987) die erste Metaanalyse zur Validität von ACs vor. Die Autoren berücksichtigten insgesamt 50 Validierungsstudien mit 107 Validitätskoeffizienten und analysierten eine mittlere korrigierte Validität von $q = .37$. Auch in der Metaanalyse

von Schmidt und Hunter (1998) ergab sich für das AC eine mittlere korrigierte Validität von $\rho = .37$. Aktuellere Folgestudien konstatieren sogar eine abnehmende Validität. So berichten Hardison und Sackett (2007) in ihrer Metaanalyse, in die 40 Validitätskoeffizienten eingeflossen sind, von einer korrigierten mittleren Vorhersagekraft von $\rho = .26$. Hermelin et al. (2007) ermittelten auf Basis von 27 Einzelbefunden eine vergleichbar niedrige Validität von $\rho = .28$. Der Validitätsverlust über die Zeit hinweg, der sich ebenfalls in einer negativen Korrelation zwischen dem jeweiligen Publikationsjahr und der entsprechenden Validitätshöhe ($r = -.25$) finden lässt (Schuler, 2007b), ist beunruhigend. 2011 legten Becker, Höft, Holzenkamp und Spinath (bzw. Holzenkamp, Spinath & Höft, 2010)² die erste Metaanalyse für ACs aus dem deutschsprachigen Raum vor. Die mittlere korrigierte Validität der untersuchten Verfahren betrug im Ergebnis $\rho = .40$. Deutschsprachige ACs erwiesen sich damit als valide und aussagekräftige Verfahren, die allerdings psychologischen Testverfahren weiterhin hinsichtlich der prognostischen Validität unterlegen waren.

Allerdings verweist die Literatur nicht nur auf die Notwendigkeit für eine Optimierung der Prognostizität von ACs, ebenso wird die Konstruktvalidität dieser Verfahrensklasse bemängelt (Brannick, Michaels & Baker, 1989; Campell & Fiske, 1959; Sackett & Dreher, 1982; Silverman, Dalwssio, Woods & Johnson, 1986). Forschungsarbeiten zeigten, dass insbesondere Faktoren wie Intransparenz der Beurteilungskriterien, die Beobachterqualität, die Methode der Urteilsbildung und ein inkonsistentes Verhalten der Teilnehmer über verschiedene Aufgaben hinweg die Konstruktvalidität schwächen (Guion, 1998; Kleinmann, 1993; Kleinmann, Kuptsch & Köller, 1996; Kleinmann & Köller, 1997). Es konnte allerdings ebenso belegt werden, dass nach professionellen Richtlinien (Task Force on Assessment Center Guidelines, 1989) gestaltete sowie durchgeführte ACs eine höhere Konstruktvalidität aufweisen (Arthur et al., 2000) und sich diese richtlinien-

² Die Metaanalyse wurde erstmalig 2010 in der *Wirtschaftspsychologie-aktuell* in deutscher Sprache und anschließend 2011 im *Journal of Personnel Psychology* in englischer Sprache veröffentlicht.

konforme Gestaltung ebenfalls positiv auf die prognostische Validität auswirkt (Becker et al., 2011).

Aufgrund der uneinheitlichen Konstruktion und Durchführung ist eine Bewertung der Qualität von ACs schwer generalisierbar, da kaum ein AC dem anderen gleicht, sondern ACs lediglich „eine relativ heterogene Klasse von Kombinationsverfahren mit der beschriebenen AC-Grundstruktur“ (Höft, 2007, S. 479) darstellen. Trotz mangelnder Konstruktvalidität, lediglich mittlerer Vorhersagekraft und hoher monetärer Kosten scheint aber vor allem die multimethodale und augenscheinvalide Ausrichtung dieses Verfahrens die Anwender zu überzeugen.

2.2 Psychologische Testverfahren

Eine tragfähige Definition für Testverfahren liefern Lienert und Raatz (1998). Sie beschreiben ein Testverfahren als „wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer empirisch abgrenzbarer Persönlichkeitsmerkmale mit dem Ziel einer möglichst quantitativen Aussage über den relativen Grad der individuellen Merkmalsausprägung“ (Lienert & Raatz, 1998, S. 1).

Die Definition verweist auf die drei existenziellen Hauptgütekriterien, die an Testverfahren gestellt werden. So muss ein Testverfahren objektiv, zuverlässig und gültig das Zielkonstrukt erfassen können. Die Gütekriterien stehen dabei in einer Abhängigkeitsbeziehung zueinander und können folglich nicht isoliert voneinander betrachtet werden. Objektivität gilt als gegeben, wenn das Testergebnis unabhängig von den Personen ist, die das Testverfahren durchführen, auswerten und interpretieren. Ist ein Test nicht objektiv, kann dieser ferner nicht reliabel und valide ein Konstrukt erfassen. Im Vergleich zu aufwendigeren Erhebungsmethoden wie dem AC weisen Testverfahren eine hohe Objektivität auf. Selbst wenn gut geschulten Beobachtern ein aufgezeichnetes Interview oder eine AC-Übung präsentiert wird, gelangen diese oftmals bei der Beurtei-

lung der gleichen Leistung zu unterschiedlichen Bewertungen (Kleinmann, 2003). Werden hingegen Testverfahren nach eindeutigen Durchführungs-, Auswertungs- und Interpretationsrichtlinien vorgenommen, sollten alle Beobachter zu dem gleichen Ergebnis gelangen. Ohne Objektivität werden demzufolge auch die Reliabilität und Validität eines Verfahrens massiv eingeschränkt (Kersting, 2006, 2010b). Dabei bestimmt die Reliabilität den Messfehlergehalt eines Verfahrens und kennzeichnet den Grad der Genauigkeit, mit dem das geprüfte Merkmal gemessen wird (Bortz & Döring, 2006). Jede Messung – insbesondere die Messung latenter psychologischer Konstrukte – ist durch systematische, unsystematische oder zufällige Störeinflüsse fehlerbehaftet (Bühner, 2006). Je kleiner allerdings der Messfehler ausfällt, desto reliabler ist das Testergebnis. Die Reliabilität ist wiederum eine notwendige Voraussetzung für eine valide Messung.

Die Validität thematisiert die Aussagekraft der aus dem Testverfahren abgeleiteten Schlussfolgerungen und gibt an, wie gut der Test in der Lage ist, genau das Konstrukt zu messen, das zu erfassen er vorgibt (Bortz & Döring, 2006). Dabei ist es von grundlegender Bedeutung, wie gut die Aufgaben und Fragen das interessierende Konstrukt repräsentieren. Die sogenannte Inhaltsvalidität wäre bei eignungsdiagnostischen Testverfahren gegeben, wenn die Testaufgaben einen berufsspezifischen Kontextbezug aufweisen bzw. die Relevanz der Items für das Berufsleben plausibel erscheint. Die Konstruktvalidität prüft weitergehend, inwiefern der Test das zu erfassende Konstrukt misst und inwieweit die Testergebnisse im Sinne einer überzeugenden Theorie erklärt werden. So ist ein Test als ungeeignet zu bewerten, wenn er einerseits nicht das Merkmal erfasst, welches zu erfassen er vorgibt, und andererseits hinsichtlich der ihm zugrunde liegenden Theorie nicht (mehr) dem aktuellen Forschungsstand genügt (vgl. Kersting, 2010). Obwohl sogenannte Typentests (z.B.: D-I-S-G® oder MBTI®) sich in der Praxis hoher Beliebtheit erfreuen, basieren sie allzu oft „auf hoffnungslos veraltete[n] Theorien (z.B.: Carl Gustav Jung und William Moulton Marston) [...], für deren Gültigkeit bis heute keine empirischen Belege erbracht werden konnten“ (Kersting, 2010b, S. 22). Aktuelle

Theorien wie das Fünf-Faktoren-Modell der Persönlichkeit nach Costa und McCrae (1992) fußen auf theoriegeleiteten Annahmen (lexikalischer Ansatz) und ermöglichen eine Differenzierung der Persönlichkeitsbeschreibung nach verschiedenen, weitestgehend unabhängigen Persönlichkeitsdimensionen, deren Zusammenhang mit berufsbezogenen Erfolgsindikatoren mehrfach empirisch belegt werden konnte (Barrick & Mount, 1991, 2005). Besonders relevant im Kontext der Personalauswahl erscheint allerdings die Kriteriumsvalidität, welche empirisch eine Vorhersage erfolgsrelevanten Verhaltens in der künftigen Realität durch die Testleistung prüft. Demnach sollten im Rahmen der Eignungsdiagnostik Kandidaten, die im Testverfahren besser abgeschnitten haben, auch später erfolgreicher im Beruf agieren. Kann ein solcher Zusammenhang nicht nachgewiesen werden, bedeutet dieses inhaltlich, dass Kandidaten mit guten Testergebnissen keine höhere Leistung im Arbeitsalltag als Kandidaten mit schlechten Testergebnissen erbringen. Das Testverfahren hätte in einem solchen Fall seinen ursprünglichen Zweck, tatsächlich geeignete Kandidaten zu identifizieren, verfehlt und würde als nicht prognostisch valide eingestuft werden müssen.

Psychologische Testverfahren erweisen sich daher gegenüber vielen anderen Erhebungsmethoden aufgrund ihrer standardisierten Durchführung, Auswertung und Ergebnisinterpretation (Objektivität) als überlegen und dieses stellt einen Grundpfeiler für anschließende messgenaue (Reliabilität) und gültige (Validität) Testergebnisse dar. Die theoretische und empirische Evidenz im Hinblick auf die Geeignetheit von Intelligenz wie auch Persönlichkeitstests zu eignungsdiagnostischen Zwecken wird in den folgenden Kapiteln dargestellt.

2.2.1 Intelligenz

Die *allgemeine kognitive Fähigkeit* gehört wahrscheinlich zu den am besten untersuchten menschlichen Persönlichkeitsmerkmalen (Gottfredson, 2004). Gleichwohl existiert bislang noch keine allgemeingültige Definition, die das Konstrukt *Intelligenz* kontextunspe-

zifisch beschreiben kann. Es besteht allerdings Einigkeit darüber, dass die menschliche Intelligenz eine hierarchische Struktur aufweist, wobei der *g-Faktor* (allgemeine kognitive Fähigkeit) an der Spitze dieser Hierarchie steht. Während in Konzepten der *allgemeinen Intelligenz* davon ausgegangen wird, dass der *g-Faktor* (Spearman, 1904, 1927) an allen intellektuellen Leistungen beteiligt ist und die allgemeine kognitive Fähigkeit repräsentiert, beziehen sich Konzepte *spezifischer kognitiver Fähigkeiten* (Thurstone, 1938; Thurstone & Thurstone, 1941; Vernon, 1950, 1965; Cattell, 1941, 1971) auf enge, explizite Fähigkeitsbereiche wie verbales oder numerisches Verständnis, räumliches Vorstellungsvermögen, Gedächtnisleistungen oder schlussfolgernde Denkprozesse. Dabei unterliegen alle *spezifischen kognitiven Fähigkeiten* der übergeordneten, allgemeinen kognitiven Fähigkeit mit unterschiedlicher *g-Sättigung* (Carroll, 1982, 1993). In einer umfangreichen Arbeit überprüfte Carroll (1993) die Existenz dieser hierarchischen Struktur, indem er die Befunde aus über 400 Studien mit Hilfe einer Faktorenanalyse reanalysierte. Das daraus resultierende Drei-Ebenen-Modell beinhaltet auf der obersten Ebene den allgemeinen Intelligenzfaktor *g*, dem wiederum acht Sekundärfaktoren auf zweiter Ebene zugeordnet wurden.

Aber nicht nur konzeptionell wurde die *Intelligenz* ausführlich untersucht. Es existieren mittlerweile auch zahlreiche empirische Studien, denen substanzielle Zusammenhänge zwischen *allgemeiner kognitiver Fähigkeit* und Variablen aus unterschiedlichen Lebensbereichen zu entnehmen sind. So konnten korrelative Beziehungen zwischen *Intelligenz* und Schulleistungen, Ausbildungsniveau, Sozialstatus, Einkommen sowie Gesundheits- und Risikoverhalten gefunden werden (Brand, 1987; Gottfredson, 2004; Neisser et al., 1996), um nur einige Forschungsfelder zu nennen. Aber auch im Bereich der Eignungsdiagnostik erwies sich *Intelligenz* als besonders geeignet zur Vorhersage von Erfolgsindikatoren für Ausbildung und Beruf. In Metaanalysen wurde übereinstimmend gezeigt, dass *Intelligenz* den herausragenden Einzelprädiktor für berufliche Leistungskriterien (z.B.: *job performance* oder *career success*) darstellt. Aber warum vermag die *Intelli-*

genz Berufs- und Ausbildungserfolg valide vorherzusagen? Oder wie Gottfredson (2002) fragen würde: *Where and Why g matters?*

Einerseits zeichnen sich erfolgreiche Mitarbeiter durch eine hohe berufliche Lernleistung aus. Hierfür müssen sie nicht nur in der Lage sein, sich neues berufsrelevantes Wissen anzueignen. Vielmehr müssen sie dieses Wissen aktiv aufnehmen, Bedeutsames herausfiltern, um es dann abschließend in einer Form zu strukturieren sowie zu speichern, die dem späteren Abruf im Berufsalltag dienlich ist (Hunter, 1986). Intelligenten Personen gelingt dieser Prozess leichter, sodass sie in der Lage sind, mehr und schneller zu lernen als weniger intelligente Personen (Kramer, 2009). Hierdurch können sie innovativer auf neue Herausforderungen reagieren, bei denen Erfahrungswissen allein nicht ausreichen würde. Andererseits weisen erfolgreiche Mitarbeiter eine hohe Arbeitsleistung auf. Dabei gelten die kognitiven Fähigkeiten einer Person als essenziell für die Arbeitsleistung, da von ihnen abhängt, ob eine Aufgabe überhaupt kognitiv bewältigt und gelerntes (Fach-)Wissen in erfolgreiches Verhalten am Arbeitsplatz umgesetzt werden kann (Gottfredson, 2002). Es konnte gezeigt werden, dass intelligente Mitarbeiter kognitiv anspruchsvolle Aufgaben schneller erledigen, besser Prioritäten setzen und sich schneller an neue Situationen anpassen können. Demzufolge profitieren Personen mit ausgeprägten kognitiven Fähigkeiten im Vergleich zu Personen mit geringeren Kompetenzen bei steigender Komplexität der zu lösenden Probleme und Aufgabenstellungen, da *g* die Fähigkeit zur komplexen Informationsverarbeitung darstellt (Gottfredson, 1997; Schmidt, Hunter & Outerbridge, 1986).

Da Personen im Arbeitsleben mit komplexen Aufgaben und Herausforderungen konfrontiert werden, die ihnen bestimmte Fähigkeiten (wie zum Beispiel Probleme lösen, abstrakt denken, logische Schlussfolgerungen ziehen, sich Wissen aneignen, Wissen organisieren und Wissen anwenden) abverlangen, vermag *Intelligenz* berufliche Leistung im Allgemeinen und für hoch komplexe Berufe im Besonderen gut vorherzusagen (Gottfredson, 1997; Hunter, 1986; Schmidt, 2002; Schmidt & Hunter, 1993). Es ist daher

nicht verwunderlich, dass der positive Zusammenhang zwischen *Intelligenz* und beruflichen Erfolgskriterien vom Komplexitätsgrad des Berufs moderiert wird. Der Komplexitätsgrad lässt sich zum Beispiel mittels des DOT (Dictionary of Occupational Titles; U.S. Department of Labor, 1991) operationalisieren und klassifizieren. Generell wird Berufen wie Manager oder Ingenieur eine hohe Komplexität zugesprochen.

Lange Zeit stammten die metaanalytischen Studien zur Überprüfung des Zusammenhangs zwischen *Intelligenz* und beruflichen Erfolgsindikatoren überwiegend aus den USA (Hunter & Hunter, 1984; Schmidt & Hunter, 1998). Seit mehreren Jahren ist eine zunehmende Forschungsaktivität zu dieser Thematik auch im deutschsprachigen und gesamteuropäischen Raum erkennbar (vgl. hierzu Hülshager, Maier, Stumpp & Muck, 2006; Kramer, 2009; Salgado & Anderson, 2003). Insbesondere sind länderspezifische Metaanalysen in diesem Zusammenhang gewinnbringend. Denn es besteht zwischen den Ländern nicht nur eine Variation in Bezug auf die Anwendungshäufigkeiten von Intelligenztests und die Art der verwendeten Testverfahren, sondern auch hinsichtlich der vorherzusagenden Kriterien. Ebenso unterscheiden sich das organisationale Umfeld und das Verständnis von Personalauswahl zwischen Ländern beträchtlich (Ackermann, 1994; Bertua, Anderson & Salgado, 2005; Salgado, Anderson, Moscoso, Bertua & de Fruyt, 2003a, 2003b; Borman, Penner, Allen & Motowidlo, 2001; Ryan et al., 1999).

So identifizierten Schmidt und Hunter (1998) für den amerikanischen Raum eine prognostische Validität der *Intelligenz* von $\rho = .56$ für die berufliche Lernleistung, die mit den von Salgado und Anderson (2003) ermittelten Werten von $\rho = .54$ für den europäischen Bereich gut übereinstimmt. Für den deutschsprachigen Raum existieren zwei Studien, welche sich hinsichtlich der Ausprägung der gefundenen Zusammenhänge jedoch unterscheiden. So ermittelten Salgado und Anderson (2003) auf Basis von 26 Studien ($k = 26$) eine Validität von $\rho = .63$, während Hülshager et al. (2006) auf der Grundlage von 90 Studien ($k = 90$) von einer Validität von lediglich $\rho = .47$ berichten. Auch für das Kriterium Arbeitsleistung ließ sich berufsübergreifend ein positiver Zusammenhang mit *Intelli-*

genz finden, der für den amerikanischen Sprachraum allerdings etwas geringer ausfiel ($\rho = .51$; Schmidt & Hunter, 1998; für den europäischen Raum $\rho = .62$; Salgado & Anderson, 2003). Hülshager et al. (2006) identifizierten für Deutschland eine Validität von $\rho = .53$ ($k = 9$). Kramer (2009) spezifizierte metaanalytisch anhand von Studien aus Deutschland die Zusammenhänge von *Intelligenz* mit den Karriereerfolgsindikatoren Einkommen und berufliche Entwicklung sowie die Validität von *Intelligenz* in Bezug auf subjektiv bewertete Arbeitsleistung und berufsbezogene Lernleistung. Die Ergebnisse unterstreichen die Vorhersagegüte von *Intelligenz* in Bezug auf die subjektiv bewertete Arbeitsleistung ($\rho = .66$, $k = 18$, $N = 2\,739$), die berufsbezogene Lernleistung ($\rho = .62$, $k = 210$, $N = 30\,451$), das Einkommen ($\rho = .35$, $k = 7$, $N = 1\,980$) und die berufliche Entwicklung ($\rho = .33$, $k = 9$, $N = 1\,817$) in Deutschland. Die gefundenen Zusammenhänge mit *Intelligenz* und Indikatoren für Karriereerfolg im Sinne von *career success* (vgl. Judge, Kammeyer-Mueller, & Bretz, 2004) stehen im Einklang mit Befunden aus dem anglo-amerikanischen Raum (Ng, Eby, Sorensen & Feldman, 2005; Schmitt, Gooding, Noe & Kirsch, 1984; Vinchur, Schippmann, Schwitzer & Roth, 1998).

2.2.2 Persönlichkeit

In den 1960er bis 1980er Jahren galt der Nutzen von Persönlichkeitstests für die Vorhersage beruflicher Leistungen noch als ungenügend (Guion & Gottier, 1965). Erst durch zunehmende internationale Forschungsaktivitäten, die einen konsistenten bedeutsamen Zusammenhang zwischen dispositionellen Persönlichkeitsmerkmalen und berufsrelevanten Erfolgskriterien belegten (Barrick & Mount, 1991, 2005; Barrick, Mount & Judge, 2001; Salgado, 1997), konnte die Bereitschaft für den Einsatz von Persönlichkeitsinventaren im Rahmen der Eignungsdiagnostik international erhöht werden. Ein Großteil der empirischen Studien zur Überprüfung des Zusammenhangs zwischen Persönlichkeitseigenschaften und beruflichen Erfolgsindikatoren beziehen sich auf die fünf Persönlichkeitsdimensionen des Fünf-Faktoren-Modells (FFM: Goldberg, 1999; Norman, 1963;

Tupes & Christal, 1961). Auf Basis des lexikalischen Ansatzes wurden fünf weitestgehend unabhängige Persönlichkeitsdimensionen identifiziert, die faktorenanalytisch bestätigt werden konnten (McCrae & John, 1992). Die auch als *Big Five* bezeichneten Persönlichkeitsdimensionen konnten kulturübergreifend empirisch bestimmt werden (McCrae & Costa, 1997) und erwiesen sich als weitestgehend zeitlich stabil (Costa & McCrae, 1988; Digman, 1989). Das FFM umfasst die fünf Dimensionen *Neurotizismus*, *Extraversion*, *Offenheit für Erfahrungen*, *Verträglichkeit* und *Gewissenhaftigkeit*. Forschungsarbeiten zeigen, dass die *Big Five* mit unterschiedlichen Lebensaspekten, wie zum Beispiel subjektives Wohlbefinden (DeNeve & Cooper, 1998) oder sogar Langlebigkeit (Friedman et al., 1995), in positiver Beziehung stehen. Zahlreiche Studien untersuchten ebenso den Zusammenhang der FFM-Dimensionen mit beruflichen Erfolgskriterien (vgl. Barrick & Mount, 1991; Hurtz & Donovan, 2000; Robertson & Kinder, 1993; Salgado, 1997; Tett, Jackson & Rothstein, 1991).

Ergebnisse unterschiedlicher Metaanalysen deuten darauf hin, dass nur einige Persönlichkeitsdimensionen einen generalisierbaren Zusammenhang über verschiedene Berufe und Erfolgskriterien hinweg aufweisen. Während beispielsweise *Verträglichkeit* im positiven Zusammenhang mit Hilfsbereitschaft, Dienstleistungsorientierung und Teamwork steht (Barrick, Stewart, Neubert & Mount, 1998; Mount, Barrick & Stewart, 1998), korreliert die Dimension *Offenheit für Erfahrungen* positiv mit der Trainingsleistung (Barrick & Mount, 1991; Salgado, 1997). In bestimmten Tätigkeitsfeldern wie Verkauf und Management erwies sich eine ausgeprägte *Extraversion* als besonders erfolgsgünstig. Dieser Befund ist inhaltlich nachvollziehbar, da Eigenschaften wie Kontaktfreudigkeit, Geselligkeit oder Durchsetzungsfähigkeit sich insbesondere in Tätigkeitsbereichen, die sich durch eine hohe Interaktionsfrequenz auszeichnen, erfolgsförderlich auswirken können (Barrick & Mount, 1991; Hough, 1992; Mount et al., 1998). Hingegen konnten für *Gewissenhaftigkeit*, aber auch für *emotionale Stabilität* (Gegenpol zu

Neurotizismus) generalisierbare Zusammenhänge in Bezug auf unterschiedliche Erfolgsindikatoren gefunden werden (Barrick & Mount, 1991; Salgado, 1997; Tett et al., 1991).

In der Tat erscheint insbesondere eine hohe Ausprägung der *Gewissenhaftigkeit* für fast alle Arbeitsbereiche erfolgsgünstig, da eine schwache Ausprägung dieser Dimension eine Person als fahrlässig, wenig verantwortungsbewusst und gering erfolgsinteressiert umschreiben würde (Hough, 1992; Mount et al., 1998). So integrierten Barrick et al. (2001) in einer Metaanalyse zweiter Ordnung die Befunde von 15 primären Metaanalysen und analysierten eine korrigierte Vorhersagekraft von $q = .27$ für *Gewissenhaftigkeit* und $q = .13$ für *emotionale Stabilität* über verschiedene Berufsgruppen und Leistungskriterien hinweg. Die übrigen drei Dimensionen des FFM (*Extraversion*, *Verträglichkeit* und *Offenheit für Erfahrungen*) erlangten in dieser Studie keine generalisierbare Validität über verschiedene Kriterien hinweg. Schmidt und Hunter (1998) benennen eine prognostische Validität von $q = .31$ für *Gewissenhaftigkeit* in Bezug auf das Kriterium berufliche Leistung, wobei diese Persönlichkeitsdimension mit zusätzlichen 18% inkrementeller Varianzaufklärung über den Prädiktor *Intelligenz* hinaus zur Vorhersage von beruflicher Leistung beitragen konnte. In einer Längsschnittstudie untersuchten Judge, Higgins, Thoresen und Barrick (1999) den Einfluss von Persönlichkeitsmerkmalen auf den späteren Berufserfolg. Insbesondere im Kindesalter erfasste *Gewissenhaftigkeit* stellte den herausragenden Prädiktor für einen späteren intrinsischen ($r = .40$) und extrinsischen ($r = .41$) Berufserfolg dar. Unter Berücksichtigung aller im Kindesalter erfassten Dimensionen des FFM konnten Korrelationen von $r = .42$ und $r = .54$ für intrinsischen bzw. extrinsischen Berufserfolg gefunden werden. Auch hinsichtlich des Kriteriums Organizational Citizenship Behavior (OCB) wurden in den Metaanalysen von Organ und Ryan (1995) sowie Borman, Penner, Allen und Motowidlo (2001) bedeutsame Zusammenhänge von $q = .21$ und $q = .24$ mit der Dimension *Gewissenhaftigkeit* konstatiert.

2.3 **Verwendungshäufigkeiten von eignungsdiagnostischen Auswahlverfahren**

In den vergangenen Dekaden wurde zunehmend Wissen über die Verwendungshäufigkeit eignungsdiagnostischer Auswahlverfahren in Deutschland generiert. Bereits Mitte der 1980er Jahre fanden Untersuchungen sowohl zum Einsatz eignungsdiagnostischer Instrumente in deutschen Unternehmen (Schulz, Schuler & Stehle, 1985) als auch zur Personalauswahlstrategie im europäischen Vergleich (Schuler, Frier & Kauffmann, 1993) statt. Die aktuellsten Studien für den deutschsprachigen Raum legten Schuler et al. (2007) für externe Auswahlprozesse bzw. Hell et al. (2006) für die interne Rekrutierung sowie Personalentwicklung vor. Sackmann und Elbe (2000) fassten mehrere Befunde aus den 1990er Jahren zusammen und stellten fest, dass Interviews und Bewerbungsunterlagen die am häufigsten genutzten Verfahren zur Personalauswahl darstellen; damit bestätigen sie vergangene ebenso wie aktuellere Befunde (Schuler et al., 1993, 2007; Schulz et al., 1985).

Gleichwohl sind über die vergangenen Jahre unterschiedliche Trends der Verwendungshäufigkeit verschiedener Auswahlverfahren zu registrieren. Während Schulz et al. (1985) noch von einer Verwendungshäufigkeit von unter 20% für ACs berichteten, fanden Schuler et al. (1993) in ihrer Untersuchung eine Einsatzhäufigkeit von fast 40% für ACs in deutschen Unternehmen. Einen weiteren Anstieg stellten Krause, Meyer zu Kniendorf und Gebert (2001) sowie Krause und Gebert (2003) fest und berichteten, dass die Hälfte der deutschen, österreichischen und Schweizer Unternehmen ACs verwendeten. Schuler et al. (2007) identifizierten einen weiteren Anstieg der Verwendungshäufigkeit für ACs auf insgesamt 57.6% in einer repräsentativen Stichprobe deutscher Unternehmen unterschiedlicher Branchen und Bewerbergruppen im Bereich externer Auswahlprozesse. Die steigende Beliebtheit und Verwendungshäufigkeit der AC-Methode lässt sich allerdings auch international wiederfinden (Newell & Shackleton, 1994;

Scholarios & Lockyer, 1999) und stellt kein deutsches Phänomen der Personalauswahlstrategie dar.

Eine andere Entwicklung ist beim Einsatz psychologischer Testverfahren in deutschen Unternehmen erkennbar. Göhs und Dick (2001) untersuchten die Verwendungshäufigkeit psychologischer Testverfahren zur Personalauswahl und fanden ähnlich wie Sackmann und Elbe (2000) eine Prävalenzrate von ca. 30% bei deutschen Unternehmen. Schuler et al. (2007) identifizierten eine Verwendungshäufigkeit für Intelligenztests von 30,4% im Bereich externe Auswahlprozesse. Dabei sank die Einsatzhäufigkeit von Intelligenztests zwischen den Jahren 1993 und 2007 um knapp 4% (Schuler et al., 2007; Schuler et al., 1993). Bezogen auf die interne Rekrutierung verschiedener Zielgruppen stellten Hell et al. (2006) eine besonders niedrige Anwendungshäufigkeit von Intelligenztests (3,2%) fest. Auch Persönlichkeitstests finden nur wenig Einzug in die deutsche Auswahlpolitik (20% externe Auswahlprozesse: Schuler et al., 2007; 7% interne Auswahlprozesse: Hell et al., 2006). Im Gegensatz zur international einheitlichen Befundlage zur Validität psychologischer Testverfahren ergibt sich in Bezug auf die Einsatzhäufigkeit ein divergentes Bild. Während Testverfahren im Kulturvergleich in Ländern wie Frankreich, Belgien und Großbritannien regelmäßig bzw. in Spanien oder den Vereinigten Staaten von Amerika sogar häufig eingesetzt werden, erfährt diese Verfahrensklasse in deutschen Unternehmen nur selten Zuspruch (Ryan et al., 1999).

2.4 **Akzeptanz von eignungsdiagnostischen Auswahlverfahren**

Forschungsarbeiten zum Themenbereich der Personalauswahl konzentrierten sich traditionsgemäß auf die Unternehmensperspektive. Zahlreiche Studien zum Thema Validität und Nützlichkeit von Auswahlverfahren untersuchten beispielsweise, von welchen Auswahlverfahren die Unternehmen am meisten profitieren können (Schmidt & Hunter,

1998). Verstärkt rückte in den letzten Dekaden die Bewerberperspektive in den Fokus der Forschung, da sich nicht nur Unternehmen ihre Mitarbeiter aussuchen. Auch Bewerber stellen zunehmend höhere Ansprüche an ihre zukünftigen Arbeitgeber (Rynes, 1993). Die nähere Betrachtung und Untersuchung der *Bewerberreaktionen* erscheinen aus mehreren Gründen bedeutsam. Bewerber, die ein Auswahlverfahren als Eingriff in ihre Privatsphäre empfinden, werten das Unternehmensimage ab. Stellenangebote werden dann aufgrund des inakzeptablen Auswahlverfahrens von den ausgewählten Bewerbern nicht angenommen (Macan, Avedon, Paese & Smith, 1994; Ryan & Ployhart, 2000). Da der Verlust eines Top-Kandidaten zu beträchtlichen Kosten führen kann, ist die Erhaltung eines positiven Unternehmensimage während des Auswahlverfahrens von zentraler Bedeutung (Murphy, 1986). Ein negativ wahrgenommener Personalauswahlprozess kann ebenso dazu beitragen, dass Bewerber anderen potenziellen Bewerbern das Unternehmen nicht weiterempfehlen (Smither, Reilly, Millsap, Pearlman & Stoffey, 1993). Als unzumutbar wahrgenommene Personalauswahlprozesse können sich auch negativ auf den Absatz von Dienstleistungen und/oder Produkten der jeweiligen Unternehmen auswirken (Hausknecht et al., 2004). Darüber hinaus kann ein Bewerber, der sich im Auswahlverfahren ungerecht oder diskriminierend behandelt gefühlt hat, arbeits- bzw. zivilrechtlich vorgehen (Smither et al., 1993) oder Medien sowie Presse involvieren (Anderson, 2011).

Aufgrund der weitreichenden Folgen, welche die *Akzeptanz* des Auswahlverfahrens aus Bewerbersicht (*Bewerberreaktionen*) hat, wurde die Bewerberwahrnehmung Gegenstand zahlreicher Untersuchungen. U.a. wurden die Verfahrensgerechtigkeit (procedural justice; Bauer, Truxillo, Sanchez, Craig, Ferrara & Campion, 2001; Ployhart & Ryan, 1998), die Verteilungsgerechtigkeit (distributive justice; Smither et al., 1993), die zwischenmenschliche Gerechtigkeit (interpersonal justice; Ryan & Chan, 1999), die informationelle Gerechtigkeit (informational justice; Bauer, Maertz, Dolen & Campion, 1998), motivationale (Sanchez, Truxillo & Bauer, 2000) sowie Angst verursachende Faktoren

(Ryan, Ployhart, Greguras & Schmit, 1998) und die allgemeine Einstellung gegenüber unterschiedlichen Auswahlverfahren (Chan, Schmitt, Jennings, Clause & Delbridge, 1997; Macan et al., 1994) bislang untersucht.

Zur theoretischen Einbettung dieser Befunde stellten die Forscher zahlreiche Theorien vor, die sich entweder vorrangig den Einstellungen und Motiven der Bewerber widmeten (u.a. Arvey, Strickland, Drauden & Martin, 1990) oder aber den Fokus auf die Wahrnehmung des gesamten Auswahlprozesses richteten (u.a. Gilliland, 1993). So fußt zum Beispiel das von Gilliland präsentierte Modell auf der Gerechtigkeitstheorie und postuliert, dass die Bewerberwahrnehmung unmittelbar die Einstellungen und Verhaltensweisen der Bewerber sowohl in der Auswahlsituation als auch in der Auswahlentscheidung beeinflusst. Ryan und Ployhart (2000) integrierten in ihrer Übersichtsarbeit beide Forschungslinien und erarbeiteten ein integratives Modell, welches von Hausknecht et al. (2004) weiter spezifiziert wurde. In diesem Modell wird das bewerberseitige Akzeptanzurteil über ein diagnostisches Verfahren durch die Faktoren a.) Persönlichkeit des Bewerbers (z.B.: Alter, Arbeits-, Auswahlerfahrungen), b.) Verfahrenswahrnehmung (z.B.: Transparenz, Augenscheinvalidität), c.) Jobcharakteristika (z.B.: Anforderungen, Attraktivität) und d.) Organisationsmerkmale (z.B.: Selektionsrate, Historie) beeinflusst. Dabei interagieren die vier Faktoren miteinander, wobei der Zusammenhang zwischen ihnen und dem Akzeptanzurteil wiederum durch weitere Variablen (z.B.: Jobattraktivität, -alternativen) moderiert wird (Hausknecht et al., 2004). Auf Basis dieser Faktoren bilden Bewerber unter dem Gesichtspunkt der wahrgenommenen Gerechtigkeit anschließend ein Aufwand-Nutzen-Verhältnis (Kersting, 2010a). Das daraus resultierende Akzeptanzurteil der Bewerber kann sich wiederum auf die eigene Leistung im Auswahlverfahren, auf das Organisationsansehen oder die spätere Arbeitsleistung sowie Berufszufriedenheit auswirken (Hausknecht et al., 2004). Zahlreiche postulierte Zusammenhänge der Faktoren innerhalb des Modells konnten in Studien belegt und sogar metaanalytisch bestätigt werden (Chapman, Uggerslev, Carroll, Piasentin & Jones, 2004;

Hausknecht et al., 2004). Insbesondere ließen sich bedeutsame positive Zusammenhänge zwischen der Augenscheinvalidität eines Verfahrens sowie der wahrgenommenen Verfahrensvalidität mit dem Akzeptanzurteil nachweisen. Wiederum steht das Akzeptanzurteil positiv in korrelativer Beziehung zu Aspekten wie der Organisationsattraktivität oder der Jobangebotsannahme.

Trotz tragfähiger Konzepte zur *Akzeptanz* von Auswahlverfahren und zahlreicher Forschungsarbeiten in den vergangenen Jahrzehnten (vgl. Anderson, 2004; Ryan & Huth, 2008) bereitet der Forschung vor allem die psychometrische Qualität der Akzeptanzmessungen sowie die Repräsentativität der untersuchten Stichproben Sorgen. Das Konstrukt der *Akzeptanz* wird in vielen Untersuchungen nur unzureichend operationalisiert und nicht multidimensional abgebildet bzw. erfasst. In zahlreichen Studien wird die *Akzeptanz* eines Verfahrens uneinheitlich und lediglich durch *single-item-Messungen* erhoben und ausgewertet. Darüber hinaus können die Aussagekraft und Generalisierbarkeit von Akzeptanzuntersuchungen eignungsdiagnostischer Instrumente, in denen ausschließlich oder überwiegend Studierende befragt wurden, bezweifelt werden (Kersting, 2008). Es stehen bereits mehrere englischsprachige Messinstrumente zur Akzeptanzerfassung eignungsdiagnostischer Verfahren zur Verfügung (u.a. von Arvey et al., 1990; Bauer et al., 2001; Lounsbury, Bobrow & Jensen, 1989; Nevo & Sfez, 1986), die allerdings aufgrund mangelnder Psychometrik nur wenig Aufmerksamkeit finden (Kersting, 2008).

Eine der prominentesten und wegweisendsten Studien zur Messung der *Bewerberreaktion* wurde von Steiner und Gilliland (1996) durchgeführt. Ihr dafür entwickelter Fragebogen umfasste zwei Items zur Messung der *process-favorability* sowie sieben Items zur Erfassung *prozeduraler Dimensionen*. Die befragten französischen und US-amerikanischen Studenten erhielten eine kurze Darstellung und Beschreibung von zehn etablierten Auswahlverfahren (Interviews, Persönlichkeitstests, Lebensläufe, Arbeitsproben, biografische Daten, schriftliche Leistungstests, persönliche Referenzen, Ehrlichkeits-Tests, per-

sönliche Kontakte und Grafologie) und mussten anschließend jedes Verfahren hinsichtlich der *process-favorability* und der *prozeduralen Dimensionen* auf einer 5-Punkt-Likert-Skala (1 = stimme überhaupt nicht zu bis 5 = stimme voll und ganz zu) bewerten. Der Fragebogen erwies sich als reliables Messinstrument, sodass die Studie mit übersetzten Fragebogenversionen in zahlreichen Ländern an Stichproben mit Studenten und berufstätigen Personen repliziert wurde (Anderson & Witvliet, 2008; Bertolino & Steiner, 2007; Ispas, Ilie, Iliescu, Johnson & Harris, 2010; Moscoso & Salgado, 2004; Nikolaou & Judge, 2007).

Für den deutschsprachigen Raum präsentierte Kersting (2008) den Akzeptanzfragebogen AKZEPT!, der in der AKZEPT!-L-Version für Intelligenz- und Leistungstests, in der AKZEPT!-P-Version für Persönlichkeitsfragebögen und in der AKZEPT!-AC-Version für ACs vorliegt. Alle drei AKZEPT!-Versionen erfassen mit jeweils vier Items die Skalen *Messqualität*, *Augenscheinvalidität*, *Kontrollierbarkeit* und *Belastungsfreiheit*, wobei sich die Varianten verfahrensspezifisch in den Itemformulierungen unterscheiden (Kersting, 2010b). In mehreren Validierungsstudien konnten die Messgenauigkeit sowie die Konstruktvalidität des mehrdimensionalen Fragebogens zur Erfassung der *Akzeptanz* eignungsdiagnostischer Verfahren belegt werden (Kersting, 2008).

Aus der uneinheitlichen Erfassung der *Akzeptanz* über unterschiedliche Studien hinweg resultiert eine für die Wissenschaft geringe konstitutive Wissensakkumulation. Gleichwohl findet sich studienübergreifend der Befund, dass sowohl Intelligenz- als auch Persönlichkeitstests im Gegensatz zu eignungsdiagnostischen Verfahren, wie Einstellungsinterviews oder ACs, eine vergleichbar geringe *Akzeptanz* erfahren. Anderson, Salgado und Hülshager (2010; vgl. auch Steiner & Gilliland, 2001) bestimmten die *Akzeptanz* von zehn etablierten Auswahlverfahren, indem sie metaanalytisch 38 unabhängige Studien aus 26 Veröffentlichungen untersuchten, in denen zur Erfassung der *Bewerberreaktion* der Fragebogen von Steiner und Gilliland (1996) verwendet wurde. Länderübergreifend wurde Arbeitsproben (stellvertretend für AC-Aufgaben) sowie Interviews aus

Bewerbersicht eine hohe Akzeptanz attestiert, während Intelligenz- und Persönlichkeitstests lediglich einen mittleren Rang einnahmen. Auch in der Metaanalyse von Hausknecht et al. (2004), in die 86 Studien zur *Bewerberreaktion* integriert wurden, ergab sich ein vergleichbares Bild. Wiederum wurden Arbeitsproben und Interviews von den Bewerbern als im höheren Maße akzeptiert bewertet im Vergleich zu Intelligenz- sowie Persönlichkeitstests.

Explizit für den deutschsprachigen Raum konnten Schuler et al. (2007) sowie Hell et al. (2006) zeigen, dass unabhängig vom Einsatzgebiet sowohl Intelligenz- als auch Persönlichkeitstests nur wenig Akzeptanz erfahren. Im Rahmen der externen Personalauswahl erlangten Intelligenztests (Persönlichkeitstests) einen Akzeptanzwert von 1.9 (1.7) für Anwender bzw. 1.4 (1.5) für Nichtanwender auf einer dreistufigen Skala, wobei 3 die höchste Akzeptanzausprägung darstellte, wohingegen ACs, Arbeitsproben und Interviews zu den akzeptiertesten Verfahrensklassen avancierten. Ein vergleichbares Bild zeichnet sich ebenso im Rahmen interner Personalauswahlprozesse ab (Hell et al., 2006). Allerdings basieren die Akzeptanzurteile bei diesen beiden Studien auf Einschätzungen befragter Personalverantwortlicher. Diese sollten bewerten, wie akzeptiert potenzielle Bewerber die jeweiligen Auswahlverfahren beurteilen würden. Marcus (2003a) konnte indes an einer studentischen Stichprobe aus Deutschland zeigen, dass diese vergleichbar mit den Befunden von Steiner und Gilliland (1996) Interviews und Arbeitsproben hinsichtlich der *process-favorability* positiver bewerten als Intelligenztests. Empirische Belege aus dem deutschsprachigen Raum, die auf eine notwendige Differenzierung zwischen allgemeinen und berufsbezogenen Testverfahren hinweisen, liefern u.a. Klingner und Schuler (2004) sowie Krumm, Hüffmeier, Dietz, Findeisen und Dries (2011). Den Studien kann entnommen werden, dass eine berufsbezogene Testgestaltung zum Anstieg der wahrgenommenen Augenscheinvalidität und Akzeptanz im Vergleich zu abstrakten, nicht berufsbezogenen Testverfahren beitragen kann.

3 Offene Forschungsfragen

In den vorherigen Kapiteln wurde ein Überblick über die Themen Validität und Akzeptanz der Auswahlverfahren *Assessment Center* und *psychologische Testverfahren* gegeben. Dabei wurde deutlich, dass beide Auswahlverfahren in unterschiedlichen Dimensionen einerseits überzeugen können und andererseits optimierungswürdig sind. Während ACs sich steigender Beliebtheit erfreuen (Höft & Obermann, 2010) und als akzeptiertes Auswahlverfahren bezeichnet werden können (Schuler et al., 2007), verweisen entsprechende Metaanalysen auf eine eher moderate Validität. Für den primär anglo-amerikanischen Raum hat sich indes sogar ein Validitätsverlust über die vergangenen zwanzig Jahre abgezeichnet (Hermelin et al., 2007). Im Gegensatz dazu überzeugen psychologische Testverfahren aufgrund ihrer prognostischen Güte (Bertua et al., 2005; Salgado & Anderson, 2003; Schmidt & Hunter, 1998), erfahren allerdings eine nur geringe Akzeptanz u.a. wegen ihrer wenig berufsbezogenen Gestaltung und werden demzufolge nur selten eingesetzt (Ryan et al., 1999; Schuler et al., 2007).

Die Befunde zur Validität von ACs aus dem englischsprachigen Raum werfen zwei bedeutsame Forschungsfragen auf, denen sich Holzenkamp et al. (2010; Becker et al., 2011) widmeten. Die Autoren untersuchten, „inwieweit die primär anglo-amerikanischen Befunde auf den hiesigen [deutschsprachigen] (Unternehmens-) Kulturkreis übertragen werden können [und ob] ein ähnlicher negativer Trend in der Validitätsentwicklung auch hier zu beobachten [ist]“ (Holzenkamp et al., 2010, S. 19). Da bislang keine weitere Metaanalyse aus dem deutschsprachigen Raum vorgenommen wurde, waren Rückschlüsse auf eine Validitätszunahme bzw. -abnahme für den deutschsprachigen Raum nicht möglich. Berücksichtigt man neben der negativen Validitätsentwicklung zusätzlich die hohen zeitlichen, menschlichen und finanziellen Ressourcen, welche in die Konstruktion und Durchführung eines ACs investiert werden (Gierschmann, 2005), stellt sich zudem die Frage, ob die steigende Beliebtheit und Verwendungshäufigkeit von ACs in Deutschland (Höft & Obermann, 2010) auch unter dem

Aspekt der prognostischen Validität gerechtfertigt erscheinen. Aus diesem Grund wurde in Studie 1 die Validitätsgeneralisierung von ACs aus deutschen Unternehmen für beruflichen Erfolg geprüft. Im Gegensatz zu der von Holzenkamp und Kollegen im Jahr 2010 vorgelegten Metaanalyse über ACs aus dem deutschsprachigen Raum basieren die Daten der vorliegenden metaanalytischen Befundintegration auf Selbstauskünften von Personalverantwortlichen deutscher Unternehmen, die via Fragebogen erfasst wurden. Insgesamt wurden von den befragten Personalverantwortlichen 23 verwertbare und unabhängige Validitätskoeffizienten ($k = 23$) berichtet, die von $r = .18$ bis $r = .68$ reichten. Die Gesamtstichprobe umfasste 3977 AC-Kandidaten, wobei die angegebenen Stichprobengrößen der jeweiligen ACs beträchtlich variierten ($N = 30$ bis $N = 700$). Die fast identische Anzahl der eingeflossenen Koeffizienten aus der vorliegenden Studie im Vergleich mit der Metaanalyse von Holzenkamp et al. (2010) ($k = 23$ vs. $k = 24$) sowie die ähnlich großen Stichprobenumfänge ($N = 3977$ vs. $N = 3556$) begünstigen einen direkten Vergleich der beiden Studien und erlauben eine erste Aussage über eine Validitätsentwicklung für ACs im deutschsprachigen Raum. Zugleich kann ein Vergleich der Datenerhebungsmethoden (Literaturrecherche vs. Selbstauskünfte) vorgenommen werden. Der vorliegende Beitrag bildet somit eine wichtige Erweiterung bisheriger Untersuchungen, wenn erstmals im deutschsprachigen Raum die Validitätsentwicklung von ACs aus deutschen Unternehmen über die Zeit hinweg betrachtet werden kann und ein Methodenvergleich möglich wird. Folgende Fragen sollen beantwortet werden:

1. Ist auch für Assessment Center aus dem deutschsprachigen Raum ein negativer Trend in der Validitätsentwicklung zu beobachten?
2. Welche Vor- bzw. Nachteile bringen die unterschiedlichen Erhebungsmethoden (Literaturrecherche vs. Selbstauskünfte) im Rahmen einer metaanalytischen Überprüfung der Validitätsgeneralisierung von Assessment Centern mit sich?

Der Artikel zur Studie 1 ist dem **Anhang A** entnehmbar.

Benit, N. & Soellner, R. (2012). Validität von Assessment Centern in deutschen Unternehmen: Vergleich von Unternehmensdaten mit einer bestehenden Metaanalyse. *PERSONALFÜHRUNG*, 11, 32–39.

Status:

- *veröffentlicht*

Eigenanteil:

- *Entwicklung der Konzeption (vollständig)*
- *Literaturrecherche (vollständig)*
- *Methodenentwicklung (vollständig)*
- *Versuchsdesign (vollständig)*
- *Datenerhebung (vollständig)*
- *Datenauswertung (vollständig)*
- *Ergebnisdiskussion (überwiegend)*
- *Erstellung des Manuskripts (überwiegend)*
- *Revision/Überarbeitung des Manuskripts (vollständig)*

Um sowohl die prognostische Validität als auch die Wirtschaftlichkeit des ACs zu optimieren, ist der Einsatz von validen Vorauswahlmethoden besonders entscheidend (Schuler, 2007a; Warmke, 1985). Als Resultat eines in den Gesamtauswahlprozess eingebetteten Vorauswahlprozesses ist mit einer erhöhten prognostischen Vorhersagekraft zu rechnen (Sarges, 2001), welche sich auf den Ebenen monetärer, zeitlicher und menschlicher Ressourcen niederschlägt (Bögl & Schmidt, 2005). Zur Optimierung der Trefferquote bzgl. geeigneter Bewerber ist daher neben der Validität der eingesetzten Vorauswahlmethoden auch die Selektionsquote bei steigender Grundquote förderlich (Görlich & Schuler, 2006; Taylor & Russel, 1939). Mittels einer angemessenen Vorauswahl lässt sich die Basisrate der potenziell geeigneten Kandidaten in der Bewerberstichprobe bereits zu einem frühen Zeitpunkt deutlich steigern, was wiederum mit einer erhöhten

Trefferwahrscheinlichkeit für „positiv Richtige“ im weiteren Auswahlprozess einhergeht. Da bislang weder internationale noch nationale Daten über die Verwendung von Vorauswahlmethoden zur Nominierung potenzieller AC-Kandidaten vorliegen, leistet Studie 2 eine erste Bestandsaufnahme zur Einsatzhäufigkeit verschiedener Vorauswahlmethoden für die Nominierung potenzieller AC-Kandidaten in deutschen Unternehmen. Insgesamt nahmen 109 Personalverantwortliche deutscher Unternehmen an der Untersuchung teil und gaben Auskunft über die in ihrem Unternehmen verwendeten Vorauswahlmethoden, schätzten die jeweiligen Vorauswahlmethoden hinsichtlich ihrer Validität ein und benannten sowie bewerteten die Zuständigkeit und diagnostische Qualität der die Vorauswahl und das AC durchführenden Personen. Somit leistet Studie 2 einen bedeutsamen Beitrag für die Generierung potenzieller Optimierungsmöglichkeiten zur Validitätserhöhung von ACs und widmet sich folgenden Fragen:

3. Welche Vorauswahlmethoden werden zur Nominierung potenzieller AC-Kandidaten in deutschen Unternehmen verwendet und wie werden diese hinsichtlich ihrer Validität bewertet?
4. Unterscheiden sich die Zuständigkeit und diagnostische Qualität der durchführenden Personen in der Vorauswahl von den AC durchführenden Personen?

Der Artikel zur Studie 2 ist dem **Anhang B** entnehmbar.

Benit, N., Mojzisch, A. & Soellner, R. (resubmitted). Pre-Selection Methods Prior to the Internal Assessment Center. *Journal of Personnel Psychology*.

Status:

- *Revision eingereicht*

Eigenanteil:

- *Entwicklung der Konzeption (vollständig)*
- *Literaturrecherche (vollständig)*
- *Methodenentwicklung (vollständig)*
- *Versuchsdesign (vollständig)*
- *Datenerhebung (vollständig)*
- *Datenauswertung (vollständig)*
- *Ergebnisdiskussion (überwiegend)*
- *Erstellung des Manuskripts (überwiegend)*
- *Revision/Überarbeitung des Manuskripts (mehrheitlich)*

Obwohl sich bereits vor fast 15 Jahren eine Kombination aus Intelligenz- und Persönlichkeitstests zur Vorhersage von beruflichen Erfolgskriterien als geeignet herausgestellt hat (Schmidt & Hunter, 1998), fristen psychologische Testverfahren in der Personalpraxis ein Schattendasein. Die aktuellsten Daten zur Verwendungshäufigkeit psychologischer Testverfahren in deutschen Unternehmen sind den beiden Studien von Schuler et al. (2007) sowie Hell et al. (2006) bezogen auf externe und interne Auswahlprozesse entnehmbar. Eine Aktualisierung dieser Befunde erscheint sinnvoll, um zu überprüfen, ob sich die berichtete Abnahme der Einsatzhäufigkeit von Intelligenz- und Persönlichkeitstests weiter fortgesetzt hat (Schuler et al., 1993, 2007). Darüber hinaus zeigten bisherige Studien, dass die befragten Personalverantwortlichen nur teilweise über die Validität von Auswahlverfahren informiert waren, da sie ACs deutlich über- und Intelligenztests deutlich unterschätzten (Schuler et al., 2007). Neben der Fehleinschätzung der Praktiker bezüglich der Prognostizität psychologischer Testverfahren werden in der Literatur weitere Beweggründe diskutiert, die den Verfahrenseinsatz in deutschen Unternehmen limitieren. Eine geringe Augenscheinvalidität (Hausknecht et al., 2004), Widerstände von den Betriebsräten (Muck & Stumpp, 2007), juristisch-rechtliche Bedenken (Hülshager &

Maier, 2008) sowie Intransparenz der Güte und Beziehbarkeit solcher Testverfahren (Kersting, 2008) sind einige Beispiele. Eine empirische Prüfung der Relevanz solcher Beweggründe gegen den Einsatz psychologischer Testverfahren fehlte jedoch bislang und wird im Rahmen dieser Studie vorgenommen.

Als weitere Erklärung für die geringe Verbreitung psychologischer Testverfahren wird ebenso der unzureichende Transfer zwischen wissenschaftlich generiertem Wissen und dessen Umsetzung in die Praxis vermutet (vgl. Anderson, 2007). Für den englischsprachigen Raum kann entsprechenden Studien entnommen werden, dass gut dokumentierte, wissenschaftliche Forschungsbefunde den befragten Praktikern größtenteils unbekannt waren oder die Praktiker deren wissenschaftliche Evidenz sogar bezweifelten (Rynes, Brown & Colbert, 2002). Als Grund für die weit verbreitete Uneinigkeit bzw. Unkenntnis bezüglich der wissenschaftlichen Forschungsbefunde in der Praxis wurde u.a. die geringe Thematisierungsquote der Güte psychologischer Testverfahren in englischsprachigen Zeitschriften für Human Resource Management angeführt (Rynes et al., 2007). Die Autoren konnten zeigen, dass die Güte von Intelligenz und Persönlichkeit als Prädiktoren für berufliche Leistungskriterien sehr selten in den drei etablierten Zeitschriften *HR Magazine*, *Human Resource Management* und *Harvard Business Review* thematisiert wurde und wenn, dann nicht der wissenschaftlichen Evidenz entsprach. Für den deutschsprachigen Raum existieren bislang keine empirischen Studien, die den Transfer wissenschaftlicher Erkenntnisse durch entsprechende Zeitschriften für Human Resource Management in die Praxis untersuchten (Albrecht & Deller, 2011; Deller & Albrecht, 2007; Kanning, Rosenstiel, Schuler, Petermann, Nerdinger & Batinic, 2007; Maier & Antoni, 2009).

So nimmt Studie 3 zunächst eine aktuelle Bestandsaufnahme der Verwendungshäufigkeit psychologischer Testverfahren in deutschen Unternehmen vor, differenziert nach unterschiedlichen Zielgruppen und Einsatzzwecken und versucht weiterführend am Beispiel der Prognostizität psychologischer Testverfahren Erkenntnisse zu gewinnen

bezüglich der Frage, ob wissenschaftliche Empfehlungen auch in die deutsche Unternehmenspraxis nur ungenügend transferiert werden. Mittels einer Aufgabe zur Wissensüberprüfung wurde untersucht, ob Forschungsbefunde über die Güte psychologischer Testverfahren die Praxis erreicht haben. Die befragten Personalverantwortlichen sollten hierfür sechs etablierte Auswahlverfahren in eine absteigende Rangreihe nach deren Vorhersagekraft für die Auswahl tatsächlich geeigneter Bewerber sortieren. Zur empirischen Prüfung, welche bedeutsamen Beweggründe deutsche Unternehmen vom Einsatz psychologischer Testverfahren abhalten könnten, wurde zunächst in der Literatur nach entsprechenden Aspekten recherchiert und anschließend Expertenbefragungen (in Wissenschaft und Praxis) durchgeführt. Daraus resultierten zehn bedeutsame Aspekte, die von den Personalverantwortlichen nach deren Relevanz eingeschätzt wurden. Weitergehend wurde in Studie 3 untersucht, ob auch in Deutschland der Transfer von wissenschaftlich generiertem Wissen durch das Medium Zeitschrift nur unzureichend stattfindet. Mittels einer Datenbankrecherche erfolgte die Prüfung, wie verbreitet die Befunde zur Güte von Intelligenz und Persönlichkeit als Prädiktoren beruflicher Leistungskriterien in deutschsprachigen, praxisorientierten Zeitschriften für Human Resources sind.

Der vorliegende Beitrag bietet somit eine Aktualisierung und detaillierte Erweiterung bisheriger Untersuchungen in diesem Themenfeld und soll zudem eine Forschungslücke schließen, indem untersucht wird, ob auch im deutschsprachigen Raum wissenschaftliche Befunde (am Beispiel der Prognostizität psychologischer Testverfahren) durch das Medium Zeitschrift nur unzureichend in die Unternehmenspraxis transferiert werden. Folgende Fragen werden untersucht:

5. Setzt sich die von Schuler und Kollegen (2007) berichtete Häufigkeitsabnahme des Einsatzes von Intelligenz- und Persönlichkeitstests in deutschen Unternehmen fort?

6. Sind Personalverantwortliche über die prognostische Güte von Auswahlverfahren informiert?
7. Welche Beweggründe erachten die Personalverantwortlichen als besonders relevant für den geringen Einsatz psychologischer Testverfahren in der Praxis?
8. Findet auch in deutschsprachigen Praktikermagazinen ein ungenügender Transfer von wissenschaftlich generiertem Wissen (über die Güte psychologischer Testverfahren) statt?

Der Artikel zur Studie 3 ist dem **Anhang C** entnehmbar.

Benit, N. & Soellner, R. (submitted). Scientist-practitioner gap in Deutschland: Eine empirische Studie am Beispiel der Prognostizität psychologischer Testverfahren. *Zeitschrift für Arbeits- und Organisationspsychologie*.

Status:

- *eingereicht*

Eigenanteil:

- *Entwicklung der Konzeption (vollständig)*
- *Literaturrecherche (vollständig)*
- *Methodenentwicklung (vollständig)*
- *Versuchsdesign (vollständig)*
- *Datenerhebung (vollständig)*
- *Datenauswertung (vollständig)*
- *Ergebnisdiskussion (überwiegend)*
- *Erstellung des Manuskripts (überwiegend)*

Das Ausmaß der Akzeptanz eines Auswahlverfahren auf Bewerberseite hängt stark von der beruflichen Augenscheinvalidität des Verfahrens ab (Hausknecht et al., 2004). Dabei ist zu beachten, dass die Augenscheinvalidität lediglich als ein relationaler Begriff zu verstehen ist, „der die Wahrnehmung des Verfahrens mit der Wahrnehmung der diagnostischen Fragestellung verbindet. Ein Test hat nicht eine Augenscheinvalidität, sondern für verschiedene (wahrgenommene) Anwendungszwecke verschiedene Augenscheinvaliditäten“ (Kersting, 2008, S. 424). In einer Reihe von Studien konnte gezeigt werden, dass aus einer geringen beruflichen Augenscheinvalidität eine wenig akzeptierende Haltung und Wahrnehmung gegenüber dem Testverfahren resultieren können (vgl. Hausknecht et al., 2004). Testverfahren, die eine abstrakte Aufgabengestaltung ohne einen erkennbaren Bezug zu den Arbeitsanforderungen des Arbeitsplatzes aufweisen, erscheinen sowohl für Bewerber als auch für Organisationsvertreter kontextfern und erfahren daher eine geringe Akzeptanz (Eleftheriou & Robertson, 1999). So konnte u.a. belegt werden, dass die wahrgenommene Augenscheinvalidität der Aufgaben positiv mit der Testleistung korrelierte (Smither et al., 1993). In einer weiteren Studie konnte ein positiver Zusammenhang zwischen beruflicher Augenscheinvalidität eines Intelligenztests und der Testmotivation identifiziert werden, wobei sich die Motivation wiederum positiv auf die Testleistung auswirkte (Chan et al., 1997).

Positive Effekte einer augenscheinvaliden Testgestaltung beschränken sich allerdings nicht nur auf das Akzeptanzurteil der Testteilnehmer. So unterstreichen die Befunde von Schmit, Ryan, Stierwalt und Powell (1995; vgl. auch Bing, Whanger, Davison & Van Hook, 2004; Hunthausen, Truxillo, Bauer & Hammer, 2003; Lievens, De Corte & Schollaert, 2008), dass eine Testmaterialdarbietung mit Herstellung eines Berufskontextes zu einem Validitätsanstieg im Vergleich zu den Standardinstruktionen bzw. -aufgaben beitragen kann. Aufgrund der bisherigen Forschungslage besteht seit längerer Zeit die Forderung nach mehr wissenschaftlich fundierten Testverfahren mit hoher beruflicher Augenscheinvalidität und direktem Anforderungsbezug (vgl. Hülshager &

Maier, 2008; Kramer, 2009), wobei die eigentliche Intention des Verfahrens – geeignete Bewerber zu identifizieren – keinesfalls vernachlässigt werden sollte.

Um die Auswirkungen einer berufsbezogenen Testgestaltung zu konkretisieren, sind „Studien mit systematischen Variationen von Merkmalen [...] wünschenswert. So könnte beispielsweise geprüft werden, ob die Augenscheinvalidität durch Informationen und/oder durch eine bestimmte Gestaltung der Aufgaben [...] beeinflusst werden kann, sodass Erkenntnisse zu einer akzeptanzfördernden Gestaltung von Tests gewonnen werden, die nicht zulasten der eigentlichen Testgüte geht“ (Kersting, 2008, S. 432). Dieser Forderung wurde im Rahmen der vierten Studie nachgekommen, indem geprüft wurde, ob ein bewährtes diagnostisches Testverfahren (BOMAT: Hossiep et al., 2001) durch eine sehr einfache Oberflächenanpassung an Akzeptanz für die Eignungsdiagnostik gewinnen kann. Hierfür wurden die Akzeptanz und psychometrische Güte des BOMAT im Vergleich zu einem eigens konstruierten Matrizentest (BBM: berufsbezogener Matrizentest) überprüft, der auf der Grundlage der Testaufgaben des BOMAT neu entwickelt wurde. Im Mittelpunkt dieser Studie steht somit nicht nur die Prüfung einer Akzeptanz-erhöhung aufgrund der Nutzung berufsbezogener Matrizensymbole, sondern ebenso wird die psychometrische Güte der beiden Testverfahren kontrastiert. Eine differenziell psychologische Beeinflussung der Akzeptanz und Testleistung durch Persönlichkeitsmerkmale (FFM) sowie durch Alter und Geschlecht wurde ebenfalls betrachtet. Diese Kontrolle erscheint umso wichtiger, wenn man berücksichtigt, dass einige dieser Merkmale wiederum mit Außenkriterien kovariieren, wie zum Beispiel Berufserfolg (Hülshager & Maier, 2008).

Studie 4 soll zur Beantwortung der folgenden Forschungsfragen beitragen:

9. Führt die berufsbezogene Symbolgestaltung des BBM zu einer Erhöhung der Akzeptanz im Vergleich zum BOMAT?
10. Kann der BBM trotz berufsbezogener Symboladaption eine zufriedenstellende psychometrische Güte aufweisen?

Der Artikel zur Studie 4 ist dem **Anhang D** entnehmbar.

Benit, N. & Soellner, R. (2012). Misst gut, ist gut? Vergleich eines abstrakten und eines berufsbezogenen Matrizentests. *Journal of Media & Business Psychology*, 3 (1), 22–29.

Status:

- *veröffentlicht*

Eigenanteil:

- *Entwicklung der Konzeption (vollständig)*
- *Literaturrecherche (vollständig)*
- *Methodenentwicklung (vollständig)*
- *Versuchsdesign (vollständig)*
- *Datenerhebung (vollständig)*
- *Datenauswertung (vollständig)*
- *Ergebnisdiskussion (überwiegend)*
- *Erstellung des Manuskripts (überwiegend)*
- *Revision/Überarbeitung des Manuskripts (vollständig)*

4 Abschlussdiskussion

Im Folgenden werden die Hauptbefunde der beschriebenen Studien dargestellt und die Ergebnisse im Hinblick auf Erweiterungen des gegenwärtigen Forschungsstandes diskutiert. Die Abschlussdiskussion bezieht sich dabei auf die Gesamtheit aller Manuskripte und führt ihre Einzelergebnisse zusammen. Ebenso werden die verwendeten Methoden übergreifend diskutiert sowie Implikationen für weitere Forschung und Praxis abgeleitet.

Bisherige Studien zur Verwendungshäufigkeit psychologischer Auswahlverfahren zeigten, dass Verfahren mit besonders hoher prognostischer Validität nicht gleichzeitig die am häufigsten verwendeten sind (Hell et al., 2006; Schuler et al., 2007). Studie 1 überprüfte daher, ob der hohe Verbreitungsgrad von ACs in deutschen Unternehmen auch unter dem Aspekt der prognostischen Validität gerechtfertigt ist. Der hinsichtlich der Stichprobengröße sowie Kriteriumsreliabilität korrigierte und gemittelte Validitätskoeffizient betrug $\rho = .42$. Die Ergebnisse weisen auf eine zufriedenstellende Vorhersagegüte von ACs im deutschsprachigen Raum hin und geben einen weiteren Hinweis für die Generalisierbarkeit der prognostischen Validität dieses Auswahlverfahrens (Becker et al., 2011; Holzenkamp et al., 2010). Trotz einer zufriedenstellenden Validität der untersuchten ACs bleibt die prognostische Güte jedoch hinter der von psychologischen Testverfahren zurück (Salgado & Anderson, 2003; Salgado et al., 2003a). Allerdings erwies sich der Zertifizierungsprozess nach DIN 33430 als bedeutsamer Moderator. ACs aus bereits zertifizierten oder sich noch im Zertifizierungsprozess befindenden Unternehmen erlangten eine höhere Validität als Verfahren aus Unternehmen, die sich noch nicht im Zertifizierungsprozess befanden ($\rho = .47$ vs. $\rho = .38$). Dieser Befund unterstreicht die Notwendigkeit einer nach wissenschaftlichen Empfehlungen und Kriterien durchgeführten Eignungsdiagnostik und verdeutlicht, dass die Durchführung von ACs keine ‚Spielwiese für Laien‘ sein sollte (vgl. Schuler, 2007b). Insbesondere ACs unterliegen aufgrund ihrer simulationsorientierten Aufgabengestaltung einer höheren Fehleranfäll-

lichkeit als psychologische Testverfahren. Während Intelligenz- und Persönlichkeitstests standardisierte Durchführungs-, Auswertungs- sowie Interpretationsrichtlinien aufweisen, die der Qualitäts- und Validitätssicherung dienlich sind, unterliegen der Beobachtungs- sowie Bewertungsprozess in ACs zahlreichen Störeinflüssen, welche sich negativ auf die Verfahrensvalidität auswirken können (Hough & Oswald, 2000; Lievens, 1998; Woehr & Arthur, 2003). So konnte u.a. gezeigt werden, dass die Qualität des Beobachtertrainings moderierend auf die Validität der ACs Einfluss nimmt (Becker et al., 2011).

Während für den anglo-amerikanischen Raum bereits zahlreiche Metaanalysen zur Validität der AC-Methode vorliegen und eine abnehmende Validitätsentwicklung über die letzten zwanzig Jahre hinweg abbilden (Hardison & Sackett, 2007; Hermelin et al., 2007; Schmidt & Hunter, 1998; Thornton et al., 1987), besteht nun erstmals mit Studie 1 dieser Arbeit die Möglichkeit, einen Trend auch für den deutschsprachigen Raum zu identifizieren. Die Befunde der ersten Metaanalyse für deutschsprachige ACs (Becker et al., 2011; Holzenkamp et al., 2010) stehen im Einklang mit den Befunden der vorliegenden Studie. Auch die fast identische Anzahl der eingeflossenen Koeffizienten ($k = 23$ vs. $k = 24$) sowie die ähnlich großen Stichprobenumfänge ($N = 3977$ vs. $N = 3556$) der beiden Studien begünstigen einen direkten Vergleich und erlauben die Skizzierung einer Validitätsentwicklung. Nach dem bisherigen Forschungsstand erweisen sich ACs aus dem deutschsprachigen Raum als aussagekräftige Verfahren. Der Validitätsverlust über die Zeit hinweg, der sich für den anglo-amerikanischen Raum zeigte, konnte für den hiesigen (Unternehmens-)Kulturkreis nicht bestätigt werden. Aufgrund der Überschneidung des nahen Zeitbezugs zur Metaanalyse von Holzenkamp et al. (2010) bzw. Becker et al. (2011) erscheinen weitere Untersuchungen im Themenfeld notwendig, um eine langfristige Validitätsentwicklung von ACs in deutschen Unternehmen abbilden zu können.

Ebenso kann noch nicht abschließend beantwortet werden, ob die Erhebungsmethode tatsächlich zu einem *bias* hinsichtlich der Validitätsgeneralisierung geführt hat. Die Wahl der Erhebungsmethode ermöglicht im Rahmen einer Metaanalyse unterschiedliche Akzentuierungen: Eine literaturbasierte Metaanalyse überzeugt aufgrund der Berücksichtigung von vorher festgelegten Kriterien der Primärstudien, da verfahrens- sowie auswertungsbezogene Hintergrundinformationen den entsprechenden Originalbeiträgen entnehmbar sind. Resultat ist ein Höchstmaß an Transparenz, Objektivität und Replizierbarkeit. Eine auf Selbstauskünften basierende metaanalytische Befundintegration scheint insbesondere dann geeignet zu sein, wenn ein möglichst aktueller Stand abgebildet werden soll. Dennoch unterliegen auf Primärstudien basierende Metaanalysen insbesondere der Kritik, dass sie nicht den tatsächlichen Forschungsstand repräsentieren (Rustenbach, 2003), da vorrangig signifikante Ergebnisse publiziert und in die Datenintegration aufgenommen werden (*publication bias*). Im Hinblick auf diesen Aspekt besteht ein Vorteil der Erhebungsmethode Befragung. So können von den Unternehmen auch nicht signifikante oder noch nicht veröffentlichte Koeffizienten genannt werden und zudem werden durch diese Erhebungsmethode im Vergleich zu aus der Literatur gesammelten Studien aktuellere Daten erfasst. Beispielsweise flossen in die Metaanalyse von Holzenkamp et al. (2010) Studien ein, die im Zeitraum von 1987 bis 2007 durchgeführt wurden. Es stellt sich die Frage, ob 10 Jahre alte oder noch ältere Studien überhaupt zur generalisierbaren Schätzung der Güte aktueller ACs geeignet sind.

Ebenfalls wird bei Metaanalysen kritisiert, dass gemäß dem Uniformitätsprinzip die Vergleichbarkeit der in die Metaanalysen einbezogenen Studien nicht immer gegeben sei (*apples and oranges*) und eine hohe Variabilität der methodischen Güte zwischen den Primärstudien bestehe (*garbage in – garbage out*). Diese beiden Kritikpunkte betreffen ebenfalls die vorliegende Studie. Gleichwohl wurde bereits während der intensiven Recherche bzw. Akquise der potenziellen Unternehmen darauf geachtet, dass diese ACs für die Zielgruppe der Führungskräfte durchführen und mindestens 150 Mitarbeiter beschäftigen. Darüber hinaus wurde eine möglichst große Vielfalt an Unternehmens-

branchen in die Untersuchung einbezogen. Somit sollte dem Uniformitätsprinzip Rechnung getragen und eine bestmögliche Vergleichbarkeit der ACs erreicht worden sein. Eine methodische Qualitätskontrolle der berichteten Koeffizienten aus den ACs kann allerdings aufgrund der fehlenden verfahrens- sowie auswertungsbezogenen Hintergrundinformationen nicht gewährleistet werden. Ein weiterer Kritikpunkt an Metaanalysen liegt in der Tatsache, dass oftmals mehrere Koeffizienten einer Primärstudie in die Analyse einbezogen werden und diese somit abhängig voneinander sind, aber wie unabhängige Daten behandelt werden (*nonindependent effects*). In der vorliegenden Studie wurde jedem Unternehmen lediglich ein Fragebogen zugesandt, sodass ausschließlich unabhängige Daten für die Befundintegration vorlagen. Zusammenfassend lässt sich – trotz der aufgezeigten methodischen Probleme einer Metaanalyse – festhalten, dass die Vereinbarkeit von Validität und Akzeptanz bei ACs realisiert werden kann, wenn diese aufgrund ihrer augenscheinlichen Passung zwischen Aufgabenformat und künftiger Arbeitstätigkeit als akzeptiert geltenden Verfahren nach wissenschaftlichen Kriterien konstruiert und durchgeführt werden.

Trotz der zufriedenstellenden Güte der untersuchten ACs aus deutschen Unternehmen in Studie 1 bleibt weiterhin die Forderung nach einer Validitätserhöhung dieser Verfahrensklasse bestehen; insbesondere aufgrund hoher monetärer sowie zeitlicher Kosten (Goffin et al., 1996). Sollten Folgeuntersuchungen für den deutschsprachigen Raum einen bedeutsamen Validitätsabfall der ACs bestätigen, wie sie bereits für den anglo-amerikanischen Raum vorliegen, würde dieses langfristig zu einer geringeren Verfahrensakzeptanz führen können. Eine realitätsnahe Aufgabengestaltung allein erscheint nicht ausreichend, um die psychische sowie physische Belastung der Bewerber in den simulationsorientierten AC-Übungen zu rechtfertigen (Fletcher, 1991; Noe & Steffy, 1987; van Emmerik, Bakker & Euwema, 2008). Berücksichtigt man zusätzlich, dass im AC mehrere Konstrukte multimethodal erfasst werden, erscheint die generell erreichte Validitätshöhe insbesondere im Vergleich zu Verfahren (Persönlichkeits- oder Intelligenztest), die lediglich ein Konstrukt erfassen (z.B.: *Gewissenhaftigkeit*, *g-Faktor*), proble-

matisch (Hough, 2001). Eine Möglichkeit, die prognostische Validität sowie Wirtschaftlichkeit des AC zu optimieren, liegt demnach im Einsatz geeigneter Vorauswahlmethoden (Schuler, 2007a). Als Resultat einer validen Vorauswahl kann mit einer erhöhten prognostischen Validität des nachgeschalteten ACs gerechnet werden (Görlich & Schuler, 2006; Sarges, 2001), welche sich auf den Ebenen monetärer, zeitlicher und menschlicher Ressourcen niederschlagen sollte (Bögl & Schmidt, 2005).

Daher wurde in Studie 2 erstmalig die Verwendungshäufigkeit von Vorauswahlmethoden zur Nominierung potenzieller AC-Kandidaten in deutschen Unternehmen untersucht. Die Ergebnisse zeigten, dass in der Vorauswahl deutscher Unternehmen *Bewerbungsunterlagen* und *Vorgesetztereinschätzungen* zur Entscheidungsfindung dominieren. Hingegen werden valide Auswahlverfahren wie das *strukturierte Interview* oder *psychologische Testverfahren* nur selten zur Vorauswahl eingesetzt, obwohl beide Verfahrensklassen zur Vorselektion geeignet erscheinen. Interviews erfahren von Bewerberseite hohe Akzeptanz (Hausknecht et al., 2004), sind stellenspezifisch informierend und ermöglichen frühzeitig ein Stärken betonendes sowie Entwicklungsfelder aufzeigendes Feedback (Köchling & Körner, 1996; Schuler, 2002). Hierdurch könnte Bewerbern, welche bestimmte stellenspezifische Kompetenzen noch unzureichend besitzen oder sich aufgrund der kommunizierten Stellen- und Tätigkeitsinformationen gegen die Position entscheiden, ein aufwendiges Auswahlverfahren erspart werden. Auch für psychologische Testverfahren könnte in der Vorauswahl ein gewinnbringender Einsatzbereich liegen, denn diese sind prädiktiv valide (Kramer, 2009; Schmidt & Hunter, 1998), in Anbetracht ihrer geringen zeitlichen Aufwendungen hoch ökonomisch und nach fachlicher Einarbeitung auch von nicht psychologisch-diagnostisch geschultem Personal objektiv durchführbar. So könnte dem Befund Rechnung getragen werden, dass nach Einschätzung der befragten Personalabteilungsleiter in der Vorauswahl selten psychologisch-diagnostisch qualifiziertes Personal beschäftigt ist. Die Befunde der Studie weisen ebenfalls darauf hin, dass die Methodenvielfalt im Rahmen der Vorauswahl mit der Unternehmensgröße, der psychologisch-diagnostischen Kompetenz der für die Vorauswahl zuständigen Per-

sonen und dem Zertifizierungsstatus des ACs in Beziehung steht. Weitergehende Analysen zeigten, dass mit einer Methodenvielfalt im Rahmen der Vorauswahl die Wahrscheinlichkeit ansteigt, dass neben *Bewerbungsunterlagen* auch Vorauswahlmethoden höherer Güte wie *Interviews* oder *Testverfahren* eingesetzt werden.

Analog zur ersten Studie unterstreichen die Ergebnisse aus Studie 2 insbesondere den hohen Stellenwert einer nach wissenschaftlichen Erkenntnissen ausgerichteten Eignungsdiagnostik. Unternehmen, die in die Qualitätssicherung ihrer Auswahlverfahren und in die Qualifikation ihrer Personalverantwortlichen investieren, verwenden aussagekräftigere Vorauswahlmethoden und werden auf lange Sicht von dieser Personalstrategie profitieren. Allerdings könnte der generell eher geringe Einsatz valider Vorauswahlmethoden (z.B.: Interview oder psychologische Testverfahren) bereits das Resultat einer durch zunehmende Globalisierung sich verändernden Personalauswahlstrategie der befragten Unternehmen darstellen. Aufgrund der stetig sinkenden Verfügbarkeit von hoch qualifizierten Fachkräften wird möglicherweise im Rahmen der Vorauswahl eine weniger strenge Auslese getroffen (Kersting, 2011a). Die zunehmende Globalisierung begünstigt zudem den Anstieg einer heterogeneren Grundquote, bestehend aus unterschiedlich qualifizierten Bewerbern. Die daraus resultierende Heterogenität der Bewerber macht jedoch im anschließenden Hauptauswahlprozess die Nutzung valider Auswahlverfahren unverzichtbar. Es bleibt abzuwarten, ob diese veränderte Bewerbersituation langfristig Auswirkungen auf die Validität dieser Verfahrensklasse nimmt.

Um das als akzeptiert geltende AC hinsichtlich der Validitätsaspekte weiter zu optimieren, kann demnach dem Einsatz valider Vorauswahlmethoden eine große Bedeutung zugesprochen werden (Warmke, 1985). Psychologische Testverfahren erscheinen aufgrund ihrer Validität sowie Ökonomie geeignet, die Qualität der eignungsdiagnostischen Entscheidung bereits im Rahmen der Vorauswahl zu erhöhen, da zum Beispiel eindeutig ungeeigneten AC-Kandidaten ein aufwendiges Auswahlverfahren erspart bleibt und die Basisrate der potenziell geeigneten Kandidaten ansteigt. Schon in der Me-

taanalyse von Becker et al. (2011) konnte gezeigt werden, dass ACs mit integrierten Intelligenztests eine höhere Validität erlangten als Verfahren, die keine Intelligenztests eingesetzt hatten.

Trotz der aufgezeigten Vorteile, psychologische Testverfahren bereits im Vorauswahlprozess einzusetzen, verweisen die Befunde aus Studie 3 gleichwohl auf eine generell eher anhaltend ablehnende Haltung deutscher Unternehmen gegenüber dieser Verfahrensklasse. So veränderte sich die im internationalen Vergleich geringe Einsatzhäufigkeit psychologischer Testverfahren im Rahmen eignungsdiagnostischer Auswahlprozesse kaum (Ryan et al., 1999). Verglichen mit den Studien von Schuler et al. (2007) und Hell et al. (2006) aus dem deutschsprachigen Raum lassen sich einige positive Trends bei der Nutzung dieser Verfahren beobachten. Während psychologische Testverfahren weiterhin überwiegend in der externen Personalauswahl eingesetzt werden und die Einsatzhäufigkeiten hierbei nur marginal im Vergleich zu vorangegangenen Studien anstiegen, zeichnet sich im Rahmen der internen Personalauswahl und Personalentwicklung eine zunehmende Verwendung ab (vgl. Hell et al., 2006).

Allerdings weisen die Befunde darauf hin, dass weiterhin sowohl Intelligenz- als auch Persönlichkeitstests von den befragten Personalverantwortlichen als wenig valide eingeschätzt werden. Die Ergebnisse von Studie 3 unterstreichen, dass auch in deutschen Unternehmen wissenschaftlich generiertes Wissen in nur unzureichendem Maße die Praxis erreicht (Albrecht & Deller, 2011; Deller & Albrecht, 2007; Kanning, Rosenstiel, Schuler, Petermann, Nerdinger & Batinic, 2007; Maier & Antoni, 2009). Die geringe Verwendung sowie die fehlerhafte Validitätseinschätzung bzgl. psychologischer Testverfahren in deutschen Unternehmen als Beispiel der Vernachlässigung wissenschaftlicher Erkenntnisse in der Praxis können dabei entlang mehreren Argumentationslinien erklärt werden.

So kann durchaus die geringe Thematisierungsquote wissenschaftlicher Befunde über die Güte psychologischer Testverfahren in Praktikermagazinen als ein bedeutsamer

Faktor angeführt werden. Die Ergebnisse der Studie 3 lassen darauf schließen, dass auch in Deutschland nur unzureichend wissenschaftlich generiertes Wissen über das Medium *Zeitschrift* transportiert wird. Wenig beruhigen kann hierbei die Tatsache, dass es sich um kein rein deutsches Phänomen handelt (vgl. Rynes et al., 2007). Im Vergleich zur Studie von Rynes und Kollegen (2007) wurde weitergehend geprüft, ob in personalpsychologischen Fachzeitschriften die Prognostizität psychologischer Testverfahren ausreichend thematisiert wird. Für diese Zeitschriftenformate zeigte sich erwartungskonform, dass entsprechende Befunde ausreichend häufig dargestellt werden; deren Erkenntnis-transfer in Praktikermagazine schlussendlich unterbleibt. Generell unterliegen hingegen Beiträge aus personalpsychologischen Fachzeitschriften der berechtigten Kritik, dass diese oftmals eine unzureichende Praxisrelevanz aufweisen (Maier & Antoni, 2009). Auch wenn zahlreiche Beiträge unter der Überschrift *praktische Implikationen* die aus den Studienergebnissen abgeleiteten Handlungsempfehlungen darstellen, ist fraglich, inwieweit die Ideen und Vorstellungen der Wissenschaftler (mit meistens geringer Erfahrung in der Organisationspraxis außerhalb von Universitäten) in der Unternehmenspraxis in praktisch-umsetzbare Resultate münden (Albrecht & Deller, 2011; Bartunek & Rynes, 2010). Maier und Antoni (2009) schlagen daher vor, die praktische Relevanz von Forschungsartikeln von Praktikern beurteilen zu lassen.

Zur weiteren Reduzierung der Lücke zwischen Forschung und Praxis wurden bereits umfassende Handlungsempfehlungen für Akademiker und Praktiker aufgestellt (Anderson, 2007; Rynes et al., 2002, 2007). Neben methodischen Empfehlungen, welche die Relevanz und Nützlichkeit von akademischer Forschung für Praktiker erhöhen sollen, werden insbesondere die Ausweitung und Modifikation von Publikationskanälen und -formaten gefordert. So könnten *lunch lectures* oder die stärkere Anbindung von Alumni-Organisationen attraktive Möglichkeiten darstellen, um einen ungezwungenen und wertvollen Transfer realisieren zu können (Albrecht & Deller, 2011). Studien, die eine Bestimmung der Nachhaltigkeit und Effektivität solcher Interventionen untersuchen, liegen allerdings noch nicht vor.

Der unzureichende Transfer wissenschaftlicher Erkenntnisse zeigt sich ebenfalls in dem Befund, dass Informationen über die Qualität angebotener Testverfahren von den befragten Personalverantwortlichen als intransparent bewertet wurden und sich die *intransparente Testqualität* als statistisch bedeutsamer Beweggrund für die Verwendung von Intelligenz- oder Persönlichkeitstests in der Unternehmenspraxis herausstellte. Der Befund aus Studie 3, dass weniger als ein Viertel der befragten Unternehmen die verwendeten Testverfahren von psychologischen Testverlagen beziehen und zusätzlich fast die Hälfte der Unternehmen, die psychologische Testverfahren einsetzt, sich individualisierte Verfahren konstruieren lassen, unterstreicht die Feststellung des unzureichenden Informationsaustausches zwischen Wissenschaft und Praxis. Durch geeignete Kommunikationsplattformen, die eine transparente und verständliche Qualitätsbeurteilung von verfügbaren Testverfahren zur Verfügung stellen, könnte der Zugang zu wissenschaftlichen Testverfahren, die auf Basis der Gütekriterien überprüft wurden, für Praktiker erleichtert werden. In diesem Fall besteht Nachholbedarf auf Seiten der Wissenschaft, die Praxis über entsprechende Angebote adäquat zu informieren.

Ein weiterer bedeutsamer Faktor für den geringen Einsatz psychologischer Testverfahren in deutschen Unternehmen könnte in der personalrechtlichen Ungewissheit der Arbeitgeber vermutet werden. Bedeutsame Güter wie Schutz der Privatsphäre, Gleichstellung oder Datenschutz unterliegen gesetzlichen Bestimmungen (BetrVG, BGB, BDSG). Rechtlich problematisch wird es insbesondere, wenn die eignungsdiagnostischen Situationen keinen erkennbaren Bezug zu dem zu besetzenden Arbeitsplatz aufweisen und die Erfassung eines Merkmals kein berechtigtes Interesse an dieser erkennen lässt (Weiler & Knipp, 2012). Während das AC aufgrund seiner realitätsnahen Aufgabengestaltung dieser Forderung augenscheinlich nachkommt, besteht eine noch un abgeschlossene Diskussion über die Rechtmäßigkeit psychologischer Testverfahren im Rahmen der Eignungsdiagnostik. Obwohl Intelligenz und Persönlichkeitsdimensionen sich länder-, kultur- sowie berufsübergreifend als valide Prädiktoren für berufsbezogene Erfolgskriterien herausgestellt haben (Hülshager et al., 2006; Kramer, 2009; Salgado & Anderson,

2003; Schmidt & Hunter, 1998) und demnach ein berechtigtes Interesse an der Messung dieser Merkmale besteht, wird die Ermittlung eines individuellen Intelligenzquotienten gemäß deutscher Rechtsprechung bislang als Eingriff in die Persönlichkeitssphäre bewertet und gilt daher grundsätzlich als unzulässig, da „... i.d.R. der Bezug zum konkreten Arbeitsplatz fehlt ...“ (Preis, 2005, S. 1447; vgl. auch Franken & Melzer, 2007; Kaehler, 2006). Daher werden vornehmlich, wenn Intelligenztests überhaupt eingesetzt werden, Subtests aus Testbatterien verwendet, die basierend auf einer Anforderungsanalyse eine arbeitsplatzrelevante kognitive Kompetenz abbilden sollen (vgl. Hülsheger & Maier, 2008). Da Subtests aus etablierten und theoretisch fundierten Intelligenztests (z.B.: IST-2000-R: Amthauer et al., 2001; BIS: Jäger, Süß & Beauducel, 1997) fast durchgängig einer berufsunspezifischen Konstruktion unterliegen, erfahren sie (bzw. aus ihnen verwendete Subtests) aufgrund der wenig erkennbaren Berufsnähe nur wenig Akzeptanz bei den Bewerbern (Hausknecht et al., 2004; König et al., 2010).

Demzufolge kann die geringe Augenscheinvalidität als weiterer bedeutsamer Faktor für den geringen Verbreitungsgrad psychologischer Testverfahren in deutschen Unternehmen angeführt werden. Die befragten Personalverantwortlichen aus Studie 3 bewerteten die geringe Augenscheinvalidität als relevantesten Beweggrund für den geringen Praxiseinsatz dieser Verfahrensklasse. Positive Auswirkungen der Augenscheinvalidität auf Auswahlverfahren generell (Hausknecht et al., 2004) sowie bei psychologischen Testverfahren im Speziellen (Bing et al., 2004; Chan et al., 1997; Hunthausen et al., 2003; Lievens et al., 2008; Smither et al., 1993; Eleftheriou & Robertson, 1999) sind international erforscht. In bisherigen Untersuchungen aus dem deutschsprachigen Raum wurde hingegen kaum eine Differenzierung bei der Akzeptanzeinschätzung zwischen allgemeinen und berufsbezogenen Varianten vorgenommen. Zudem wird in zahlreichen Studien die Bewerberakzeptanz lediglich subjektiv von befragten Personalverantwortlichen eingeschätzt (Hell et al., 2006; Schuler et al., 2007), so dass eine Generalisierung auf die tatsächliche Bewerberreaktionen problematisch erscheint. Um die Vereinbarkeit der Validität und Akzeptanz bei psychologischen Testverfahren realisieren zu können, sollte

daher insbesondere die Augenscheinvalidität dieser Verfahrensklasse aus der Bewerberperspektive untersucht und mögliche Modifikationsmaßnahmen zur Testgestaltung generiert werden.

Dementsprechend wurde im Rahmen der Studie 4 die Akzeptanz für einen abstrakt (BOMAT: Hossiep et al., 2001) sowie einen berufsbezogen gestalteten Matrizen-test (BBM) an potenziellen Bewerbern überprüft. „Es kommt [allerdings] nicht allein darauf an, dass die Verfahren bei den Teilnehmern gut ankommen, sondern es kommt darauf an, dass diagnostische Verfahren valide sind und zusätzlich gut ankommen“ (Kersting, 2008, S. 432). Folglich wurde neben der Akzeptanzbewertung auch die psychometrische Güte der beiden Matrizen-tests überprüft. Beide Testverfahren erwiesen sich als hinreichend reliabel sowie konstrukt- und kriteriumsvalide. Die Testgestaltung beeinflusste die Akzeptanzeinschätzung hinsichtlich der Augenscheinvalidität zugunsten des BBM mit berufsbezogenen Symbolen im Vergleich zum BOMAT. Zusätzlich konnte der positive Zusammenhang zwischen Testleistung und wahrgenommener Augenscheinvalidität, der bereits in vorherigen Studien an amerikanischen Stichproben gefunden wurde (Chan et al., 1997; Smither et al., 1993), für den deutschsprachigen Raum bestätigt werden. Studie 4 leistet vor allem deshalb eine wichtige Erweiterung zu bisherigen Untersuchungen in diesem Themenfeld, da systematisch lediglich ein Merkmal des etablierten BOMAT variiert wurde und die Ergebnisse auf eine positive Beeinflussung der Augenscheinvalidität durch die Manipulation dieses Merkmals (Symbolgestaltung) verweisen. Diese Erkenntnisse können zur Konstruktion künftiger Testverfahren genutzt werden, da die Symbolveränderung zur Akzeptanzerhöhung beigetragen hat und sich nicht zu Lasten der eigentlichen Testgüte auswirkte. Die Erkenntnisse der vorliegenden Studien unterstreichen die Forderung nach mehr wissenschaftlich fundierten Testverfahren mit hoher beruflicher Augenscheinvalidität und direktem Anforderungsbezug (vgl. Hülshager & Maier, 2008; Kramer, 2009) und zeigen auf, dass auch für diese Verfahrensklasse die Vereinbarkeit von Validität und Akzeptanz realisiert werden kann.

Die eindeutigen Befunde aus Studie 4, die eine Akzeptanzerhöhung allein durch die berufsbezogene Aufgabengestaltung aufzeigen konnten sowie der Befund aus Studie 3, demnach insbesondere eine geringe Augenscheinvalidität zur Nichtverwendung psychologischer Testverfahren in der Unternehmenspraxis führt, lassen die Konstruktion eines berufsbezogenen Intelligenzverfahrens äußerst sinnvoll erscheinen. Dessen Konstruktion und Validierung soll im Folgenden als ein Bestandteil des kongruenten Forschungsprogramms dieser Dissertation mit seinen ersten Ergebnissen dargestellt werden.

Aus den Studien resultierende Forschung

Im Rahmen dieser Arbeit wurde ein Testverfahren (ANAKONDA: Analogiebezogene Kognitiv Numerische DatenAnalyse) entwickelt, das einerseits zuverlässig die Intelligenz (kognitive Kompetenz) erfasst und andererseits einen explizit augenscheinlichen Berufsbezug aufweist. ANAKONDA wurde auf der Grundlage der klassischen Testtheorie entwickelt und zielt auf die Erfassung der Allgemeinintelligenz im Sinne des Spearman'schen g-Faktors ab. Das Testverfahren misst Intelligenz im höheren Leistungsbereich und überprüft neben dem analytischen und analogiebezogenen Denkvermögen auch die numerische Kompetenz. Die Kandidaten lösen dabei insgesamt 24 Aufgaben. Die Aufgabe der Testkandidaten besteht jeweils darin, Unternehmenskennziffern aus einem Diagramm und einer Tabelle zu analysieren, aus denen Informationen über fiktive Gewinn- und Verlustunternehmen entnehmbar sind. Die Kandidaten müssen identifizieren, welche Faktoren im Zusammenspiel den Erfolg der Gewinnunternehmen ermöglicht haben. Hieraus soll anschließend abgeleitet werden, welche Faktoren bei einem Verlustunternehmen den finanziellen Misserfolg „verursacht“ haben. Pro Aufgabe müssen sich die Bewerber jeweils für eine Antwort entscheiden, die sie aus fünf Antwortmöglichkeiten auswählen. Dem **Anhang E** sind die Instruktionen sowie Beispielaufgaben des Testverfahrens entnehmbar. Im Rahmen der Testkonstruktionsphase wurden unterschiedliche Aufgaben- sowie Antwortformate an 360 Studierenden verschiedener Studiengänge und Uni-

versitäten erprobt. Es werden nachfolgend die Validierungsergebnisse der aktuellen Testversion dargestellt.

ANAKONDA erwies sich im Rahmen dieser Validierungsstudie ($N = 224$) als hinreichend schwierig, trennscharf und reliabel. Die Schwierigkeitsindizes variierten von .35 bis .84 ($M = .63$, $SD = .46$) und deckten somit unterschiedliche Schwierigkeitsbereiche ab. Die Trennschärfen lagen nach part-whole-Korrektur zwischen .31 und .82 ($M = .64$, $SD = .11$). ANAKONDA umfasst 24 Items und erzielte eine hervorragende interne Konsistenz mit einem Cronbachs $\alpha = .94$. Die Testteilnehmer erlangten im Durchschnitt 14.95 von 24 möglichen Punkten ($SD = 7.75$). ANAKONDA korrelierte mit dem retrospektiven Kriterium *Abiturgesamtnote* moderat ($r = -.31$, $p < .01$, $N = 224$). Für eine Teilstichprobe ($N = 84$) lagen auch Daten zu dem Kriterium *Bachelor- bzw. Vordiplomgesamtnote* vor. Der Zusammenhang erwies sich mit $r = -.38$ ($p < .01$) als statistisch signifikant. Die Befunde zur Kriteriumsvalidität für ANAKONDA weisen auf eine akzeptable Validität hin und unterstreichen die Güte von Intelligenztests zur Vorhersage von Ausbildungserfolg in Deutschland (Hülshager et al., 2006). Zur Prüfung der konvergenten und diskriminanten Validität wurde ANAKONDA mit unterschiedlichen Konstrukten korrelativ in Beziehung gesetzt. Die statistisch signifikanten positiven Korrelationen mit den Subtests *Matrizen* ($r = .58$, $p < .01$), *Analogie* ($r = .55$, $p < .01$) und *Zahlenreihe* ($r = .68$, $p < .01$) des IST 2000-R (Amthauer et al., 2001) weisen auf eine konvergente Validität des Testverfahrens hin. Hinweise auf Divergenz leisten die geringen und nicht signifikanten Korrelationen mit den *Big Five* (Borkenau & Ostendorf, 2008: $r_{\text{Neurotizismus}} = .05$, $p = .66$; $r_{\text{Extraversion}} = .01$, $p = .95$; $r_{\text{Offenheit}} = -.07$, $p = .49$; $r_{\text{Verträglichkeit}} = -.11$, $p = .32$; $r_{\text{Gewissenhaftigkeit}} = .10$, $p = .33$).

Die Akzeptanz des Testverfahrens wurde über den AKZEPT!-L (Kersting, 2008) erfasst und mit den Akzeptanzwerten des IST 2000-R verglichen. Hinsichtlich der zwei Dimensionen *Augenscheinvalidität* ($M_{\text{ANAKONDA}} = 3.30$, $SD = 1.07$; $M_{\text{IST2000R}} = 2.21$, $SD = .86$; $t(131) = 11.09$, $p < .001$) und *Messqualität* ($M_{\text{ANAKONDA}} = 3.52$, $SD = .97$; $M_{\text{IST2000R}} = 3.19$, $SD = 1.03$; $t(131) = 2.97$, $p = .004$) konnten signifikante Unterschiede zugunsten von ANA-

KONDA gefunden werden. Der IST 2000-R wurde hingegen als belastungsfreier eingeschätzt (*Belastungsfreiheit*: $M_{\text{ANAKONDA}} = 3.34$, $SD = 1.06$; $M_{\text{IST2000R}} = 3.69$, $SD = 1.05$; $t(131) = -3.23$, $p = .002$). Die beiden Testverfahren unterschieden sich nicht statistisch bedeutsam in Hinsicht auf die Dimensionen *Kontrollierbarkeit* ($M_{\text{ANAKONDA}} = 4.38$, $SD = 1.18$; $M_{\text{IST2000R}} = 4.61$, $SD = 1.39$; $t(131) = -1.46$, $p = .15$). Zur Überprüfung der Testfairness wurde das Testergebnis des ANAKONDA hinsichtlich der Kontrollvariablen Geschlecht und Alter überprüft. Es ergaben sich für das Geschlecht ($M_{\text{weiblich}} = 15.37$, $SD = 7.29$; $M_{\text{männlich}} = 13.54$, $SD = 7.85$; $t(222) = 1.56$, $p = .12$) keine signifikanten Mittelwertsunterschiede. Das Alter stand allerdings in negativem Zusammenhang mit dem Testergebnis ($r = -.19$; $p = .003$).

Mit ANAKONDA wurde ein Testverfahren zur Messung der kognitiven Kompetenzen entwickelt, das im hohen Maße mit klassischen Intelligenztests korreliert ist. Die Aufgabeninhalte bewegen sich im anspruchsvollen Berufskontext (mit Aufgaben hohen Komplexitätsgrades) und überzeugen aufgrund ihrer Alltagsnähe. Dieses hat zur Folge, dass Testteilnehmer ANAKONDA im Vergleich zum klassischen Intelligenztest eine signifikant höhere Augenscheinvalidität attestieren (Benit, 2012a). Um den Gültigkeitsanspruch der bisherigen Befunde zu erweitern, sind eine Folgevalidierung und Normierung an einer Bewerberstichprobe in Arbeit. ANAKONDA wurde im U-Form Testverlag veröffentlicht (Benit, 2012b) und steht Unternehmen zur Auswahl ihrer Auszubildenden, Trainees und dualen Studenten zur Verfügung. Die dort gewonnenen Daten werden für weitere Forschungsvorhaben (Validierung, Normierung) vom Testverlag zur Verfügung gestellt.

Mit Blick auf alle vier Studien erscheinen kontextspezifische Testverfahren mit Berufsbezug (wie ANAKONDA), die Eignungsdiagnostik ein bedeutsames Stück verbessern zu können. So kann die Validität von ACs durch die Integration *berufsbezogener* Testverfahren bedeutsam erhöht werden (vgl. Becker et al., 2011). Zur weiteren Optimierung der kosten- und zeitintensiven ACs erscheint eine effektivere Vorauswahl geeignet, die ebenso durch psychologische Testverfahren ökonomisch und zielführend realisiert

werden könnte (Mussel, Frintrup, Pfeiffer & Schuler, 2007). Neben ACs sind Interviews etablierte und häufig verwendete Auswahlverfahren (Hell et al., 2006; Ryan et al., 1999; Schuler et al., 2007), die allerdings größtenteils nur biografie- sowie vergangenheitsbezogene Merkmale erfassen. Unternehmen, die auf ein kostenintensives AC verzichten möchten, könnten ihren Auswahlprozess durch den Einsatz *berufsbezogen gestalteter* Testverfahren optimieren, da sich durch diese Verfahrensklasse ergänzend zum Interview eigenschaftsorientierte Merkmale messen lassen.

Anregungen für weitere Forschungsarbeiten

Die dargestellten Ergebnisse der jeweiligen Studien werfen allerdings auch neue Fragen auf und regen zu weiterführenden Forschungsarbeiten an. So ist es bislang noch ungeklärt, ob die metaanalytischen Befunde aus Studie 1 einem Methodenartefakt unterliegen. Eine abschließende Prüfung bedarf einer langfristig angelegten Untersuchung, in der rückwirkend für den Erhebungszeitraum der Studie eine auf veröffentlichten Primärstudien basierende Metaanalyse vorgenommen wird. So könnte bestmöglich abgeschätzt werden, ob die unterschiedlichen Erhebungsmethoden zu vergleichbaren Validitätsschätzungen gelangen. Da die hier dargestellte metaanalytische Untersuchung in einem nahen Zeitbezug zur Metaanalyse von Becker et al. (2011) durchgeführt wurde, sollten in den kommenden Jahren regelmäßig Metaanalysen im deutschsprachigen Raum erfolgen, um Aussagen über die langfristige Validitätsentwicklung machen zu können.

Weiterhin stellt sich die Frage, inwieweit sich die in dieser Arbeit dargestellten Befunde zur Verwendungshäufigkeit der eingesetzten Vorauswahlmethoden (Studie 2) auch auf andere Bewerbergruppen (bzgl. der Zielposition) sowie Verfahrensklassen und Kulturkreise übertragen lassen. Da bislang auch international keinerlei Studien vorliegen, die einen systematischen Überblick über die Nutzung von Vorauswahlmethoden leisten, erscheint weiterführende Forschung anknüpfend an Studie 2 wünschenswert. Die Prüfung der von Kersting (2011a) aufgestellten Hypothese, dass infolge der geringe-

ren Verfügbarkeit von hoch qualifizierten Fachkräften die Vorauswahl in Zukunft weniger streng ausfallen werde, bedarf eines längsschnittlich angelegten Forschungsdesigns. Hierfür würde sich Studie 2 als Referenzpunkt eignen. Ebenso werfen die Erkenntnisse aus Studie 2 die weitergehende Frage auf, ob eine Konfundierung zwischen eingesetzten Vorauswahlmethoden und im AC verwendeten Übungen vorliegt. Da in der vorliegenden Untersuchung keine Informationen über die verwendeten Aufgaben der ACs erhoben wurden, kann nicht beurteilt werden, ob Unternehmen zum Beispiel keine Interviews im Rahmen der Vorauswahl einsetzen, da diese bereits Bestandteil des anschließenden ACs sind. Dieser Aspekt sollte in Folgestudien Berücksichtigung finden. Lohnenswert scheint darüber hinaus die Prüfung, welchen inkrementellen Validitätszuwachs Vorauswahlverfahren über AC-Übungen hinweg leisten können.

Studie 3 leistet einen substanziellen Beitrag zum besseren Verständnis des *scientist-practitioner gap* im deutschsprachigen Raum. Bedenkt man die Relevanz des Transfers wissenschaftlicher Erkenntnisse in die Unternehmenspraxis, wird ersichtlich, dass es dringend weiterer Forschungsbemühungen und vor allem fundierter Umsetzungsbemühungen bedarf, um Fortschritte für Wissenschaft und Praxis zu erreichen. Insbesondere wenn man berücksichtigt, dass „Studien, die darauf abzielen, die Beziehung zwischen Forschung und Praxis (über eine theoretische Diskussion hinausgehend) empirisch zu beleuchten, [...] bislang ausschließlich im nordamerikanischen Raum“ (Albrecht & Deller, 2011, S. 49) stattgefunden haben, sollten Folgestudien dieses Phänomen auch in Deutschland auf Basis empirischer Untersuchungen erschließen. Vorstellbar wäre in diesem Zusammenhang eine Replikation der Studie von Rynes et al., (2002), indem Personalverantwortlichen deutscher Unternehmen gut dokumentierte, wissenschaftliche Forschungsbefunde präsentiert werden und überprüft wird, inwieweit diese den Praktikern bekannt sind. Darüber hinaus sollte in zukünftigen Studien exploriert werden, über welche Themenbereiche bzw. Fragestellungen Praktiker eigentlich informiert werden möchten. Es wird häufig bemängelt, dass Forschungsergebnisse nicht von Praktikern aufgegriffen würden. Allerdings existieren bislang auch nur wenige Studien, die poten-

tielle Gründe hierfür untersuchen. An Studie 3 anknüpfende Forschung, die zur Optimierung des Transfers wissenschaftlicher Befunde sowie deren Umsetzung in die Praxis beitragen, erscheint notwendig und lohnenswert.

Um den Gültigkeitsanspruch der Befunde aus Studie 4 zu erweitern, sollte hingegen eine Validierung der beiden Testverfahren (BOMAT und BBM) in einem randomisierten Versuchsaufbau an einer berufstätigen Bewerberstichprobe realisiert werden. Lohnenswert erscheint eine weitere Variation der berufsspezifischen Symbole. So sollte sich die Symbolgestaltung nicht nur auf die Branche des Automobilbaus beschränken. Die Entwicklung von Matrizensymbolen für unterschiedliche Branchen (z.B.: Pharmazie, Banken, Versicherungen, Telekommunikation) oder die Verwendung firmenspezifischer Matrizensymbole erscheint relevant und sollte durch die Verwendung spezifischer Forschungsdesigns weiter fortgeführt werden. Die Befunde aus der präsentierten Validierungsstudie des neu entwickelten Testverfahrens ANAKONDA weisen darauf hin, dass ein messgenaues und augenscheinvalides Instrument konstruiert wurde. Eine weitere Validierung an einer beruflichen Bewerberstichprobe erscheint notwendig sowie vielversprechend und ist bereits in Planung.

5 Fazit

Zusammengefasst besteht der wissenschaftliche Wert der vorliegenden Arbeit in der stringenten Weiterführung des bisherigen Forschungsstands und der Beantwortung offener Forschungsfragen zur Validität von Assessment Centern und der Akzeptanzmessung psychologischer Testverfahren. Die gewonnenen Kenntnisse sind nicht nur für die Personalpsychologie als Wissenschaft, sondern auch für die Unternehmenspraxis von Interesse und Nutzen.

So ermöglichte Studie 1 eine erste Bewertung der Validitätsentwicklung von ACs im deutschsprachigen Raum, da zuvor lediglich eine entsprechende Metaanalyse aus diesem Themenfeld vorlag (Becker et al., 2011). Die Befunde zeigten, dass ACs aus dem deutschsprachigen Raum eine zufriedenstellende Validität aufweisen, insbesondere wenn sie nach wissenschaftlichen Kriterien konstruiert und durchgeführt wurden. Wenn das gegenwärtig als akzeptiert geltende AC auch weiterhin als akzeptiertes Verfahren gelten will, muss es in weiteren Metaanalysen der Validitätsprüfung standhalten. Mit Studie 2 wurde erstmalig eine Bestandsaufnahme von Vorauswahlmethoden vorgelegt, die zum besseren Verständnis dieses wichtigen Auswahlprozesses beiträgt. Eine Optimierung des Vorauswahlprozesses erscheint notwendig und könnte ebenso zur Akzeptanz- sowie Validitätserhöhung der AC-Methode beitragen, da zum Beispiel eindeutig ungeeigneten AC-Kandidaten ein aufwendiges Auswahlverfahren erspart bleibt und die Basisrate der potenziell geeigneten Kandidaten ansteigt (Obermann, 2006; Sarges, 2001; Warmke, 1985). Studie 3 kam der bestehenden Forderung nach, auch für den deutschsprachigen Raum empirisch den Transfer zwischen wissenschaftlich generiertem Wissen sowie dessen Umsetzung in die Praxis zu untersuchen (Albrecht & Deller, 2011). Am Beispiel der Prognostizität psychologischer Testverfahren konnte eine anhaltende Zurückhaltung deutscher Unternehmen, psychologische Testverfahren zur Eignungsdiagnostik einzusetzen, aufgezeigt werden, die u.a. auf einen unzureichenden Transfer wissenschaftlicher Erkenntnisse in die Unternehmenspraxis zurückgeführt werden kann.

Darüber hinaus wurden erstmalig in der Literatur als bedeutsam angeführte Beweggründe gegen den Einsatz psychologischer Testverfahren (vgl. hierzu u.a. Hausknecht et al., 2004; Hülshager & Maier, 2008; Kersting, 2008; Marcus, 2003b; Muck & Stumpp, 2007) von deutschen Personalverantwortlichen hinsichtlich deren Relevanz bewertet. Somit konnte eine Priorisierung dieser Beweggründe aus Organisationsperspektive abgebildet werden. Als bedeutsamster Grund avancierte eine geringe Augenscheinvalidität der Testverfahren. Die systematische Variation der Testgestaltung in Studie 4 lieferte Erkenntnisse zu einer akzeptanzförderlichen Gestaltung von Testverfahren, die nicht zulasten der eigentlichen Testgüte geht (Kersting, 2008). Mit dem neu konstruierten Test ANAKONDA wurde außerdem ein Verfahren entwickelt, welches für die Realisierbarkeit der Forderung nach Vereinbarkeit von Validität sowie Akzeptanz steht und der wissenschaftlichen Forderung nachkommt, valide Instrumente mit eindeutigem Berufsbezug der Praxis zur Verfügung zu stellen (Kramer, 2009).

Abschließend kann festgehalten werden, dass sowohl für *Assessment Center* als auch für *psychologische Testverfahren* die Vereinbarkeit von Verfahrensvalidität und –akzeptanz unter bestimmten Bedingungen realisiert werden kann. Für das als akzeptiert geltende AC kann ein Validitätszuwachs durch eine nach wissenschaftlichen Kriterien gestalteten Konstruktion und Durchführung sowie durch eine optimierte Vorauswahl erreicht werden. Valide Testverfahren können hingegen durch eine berufsbezogene Aufgabengestaltung an Akzeptanz gewinnen und so die angestrebte Vereinbarkeit von Validität und Akzeptanz erzielen. Wie zu Beginn dieser Arbeit angemerkt, wird die Bedeutung einer validen und akzeptierten Eignungsdiagnostik zukünftig weiter ansteigen. Umso entscheidender sind vor diesem Hintergrund die Befunde der vorliegenden Dissertation, dass berufsbezogene sowie nach wissenschaftlichen Empfehlungen konstruierte Auswahlverfahren das Potenzial besitzen, diesen beiden Anforderungen - Validität und Akzeptanz – gleichermaßen gerecht zu werden.

Literatur

- Ackermann, P. L. (1994). *Intelligence, attention and learning: Maximal and typical performance*. In D. K. Dettermann (Ed.), *Current topics in human intelligence; Theories of intelligence* (1–27). Norwood, NJ: Ablex.
- Albrecht, A.-G. & Deller, J. (2011). Arbeits- und organisationspsychologische Forschung und betriebliche Praxis – eine Replik zu Maier & Antoni. *Zeitschrift für Arbeits- und Organisationspsychologie*, *55*, 1–5.
- Amthauer, R., Brocke, B., Liepmann, D. & Beauducel, A. (2001). *I-S-T 2000 R – Intelligenz-Struktur-Test 2000 R*. Göttingen: Hogrefe.
- Anderson, N. (2004). Editorial – The Dark Side of the Moon: Applicant perspectives, negative psychological effects (NPEs), and candidate decision making in selection. *International Journal of Selection and Assessment*, *12*, 1–8.
- Anderson, N. (2007). The Practitioner–Researcher Divide Revisited: Strategic-level bridges and the roles of IWO psychologists. *Journal of Occupational and Organizational Psychology*, *80*, 175–183.
- Anderson, N. (2011). Perceived job discrimination: Toward a model of applicant propensity to case initiation in selection. *International Journal of Selection and Assessment*, *19*, 229–244.
- Anderson, N., Herriot, P. & Hodgkinson, G. (2001). The practitioner–researcher divide in Industrial, Work and Organizational (IWO) psychology: Where are we now, and where do we go from here? *Journal of Occupational and Organizational Psychology*, *74*, 391–411.
- Anderson, N., Lievens, F., van Dam, K. & Ryan, A. M. (2004). Future Perspectives on Employee Selection: Key Directions for Future Research and Practice. *Applied Psychology*, *53*, 487–501.
- Anderson, N., Salgado, J. F. & Hülsheger, U. R. (2010). Applicant Reactions in Selection: Comprehensive Meta-Analysis into Reaction Generalization Versus Situational Specificity. *International Journal of Selection and Assessment*, *18*, 291–304.
- Anderson, N. & Witvliet, C. (2008). Fairness Reactions to Personnel Selection Methods: An international comparison between the Netherlands, the United States, France, Spain, Portugal, and Singapore. *International Journal of Selection and Assessment*, *16*, 1–13.
- Arthur W., Woehr, D.J. & Maldegen, R. (2000). Convergent and discriminant validity of assessment center dimensions: a conceptual and empirical re-examination of the assessment center construct-related validity paradox. *Journal of Management*, *26*, 813–835.

- Arvey, R.D., Strickland, W., Drauden, G. & Martin, C. (1990). Motivational components of test taking. *Personnel Psychology*, 43, 695–716.
- Barrick, M. R. & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1–26.
- Barrick, M. R., Mount, M. K. & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, 9, 9–20.
- Barrick, M. R. & Mount M. K. (2005). Yes, personality matters: Moving on to more important matters. *Human Performance*, 18, 359–372.
- Barrick, M., Stewart, G., Neubert, M. & Mount, M. K. (1998). Relating member ability and personality to work-team processes and team effectiveness. *Journal of Applied Psychology*, 83, 337–391.
- Bartunek, J. M. & Rynes, S. L. (2010). The construction and contributions of "implications for practice": What's in them and what might they offer? *Academy of Management Learning & Education*, 9, 100–117.
- Bauer, T., Maertz, C., Dolen, M. & Campion, M. (1998). A longitudinal assessment of applicant reactions to an employment test. *Journal of Applied Psychology*, 83, 892–903.
- Bauer, T., Truxillo, D., Sanchez, R., Craig, J., Ferrara, P. & Campion, M. (2001). Development of the Selection Procedural Justice Scale (SPJS). *Personnel Psychology*, 54, 387–419.
- 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.
- Benit, N. (2012a). Intelligenztests in der Personalauswahl: Wieso, weshalb, warum??? *arecruiter*, 9, 34-37.
- Benit, N. (2012 b). *Analogiebezogene Kognitio Numerische DatenAnalyse (ANAKONDA)*. Solingen: U-Form Verlag.
- Bertolino, M. & Steiner D.D. (2007). Fairness reactions to selection methods: An Italian study. *International Journal of Selection and Assessment*, 15, 107–205.
- Bertua, C., Anderson, N. & Salgado, J. F. (2005). The predictive validity of cognitive ability tests: A UK meta-analysis. *Journal of Occupational and Organizational Psychology*, 78, 387–409.
- Berufsverband Deutscher Psychologinnen und Psychologen. (2009). *Informationen zur Norm 33430 [On-line]*. Zugriff am 02.08.2009. Verfügbar unter <http://www.bdp-verband.org/bdp/politik/din.shtml>.

- Bing, M. N., Whanger, J. C., Davison, H. K. & Van Hook, J. B. (2004). Incremental Validity of the Frame-of-Reference Effect in Personality Scale Scores: A Replication and Extension. *Journal of Applied Psychology*, 89, 150–157.
- Bögl, M. & Schmidt, R. (2005). Instrumente zur Vorauswahl von AC-Teilnehmern. In Arbeitskreis Assessment Center e.V. (Hrsg.), *Aus Konfusion, Kontroversen und Konzepten wächst Kompetenz. Dokumentation zum 6. Deutschen Assessment Center Kongress* (S. 57–64). Lengerich: Pabst Science Publishers.
- Borkenau, P. & Ostendorf, F. (Hrsg.). (2008). *NEO-Fünf-Faktoren Inventar nach Costa und McCrae (NEO-FFI)* (2., neu normierte und vollständig überarbeitete Auflage). Göttingen: Hogrefe.
- Borman, W. C., Penner, L. A., Allen, T. D. & Motowidlo, S. J. (2001). Personality predictors of citizenship performance. *International Journal of Selection and Assessment*, 9, 52–69.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. Aufl.). Heidelberg: Springer-Verlag.
- Brand, C. (1987). The importance of general intelligence. In S. Modgil & C. Modgil (Ed.), *Arthur Jensen: Consensus and controversy* (251–265). New York: Falmer.
- Brannick, M., Michaels, C. & Baker, D. (1989). Construct Validity of In-Basket Scores. *Journal of Applied Psychology*, 74, 957–963.
- Bray, D. & Grant, D. (1966). The assessment center in the measurement of potential for business management. *Psychological Monographs: General and Applied*, 80, 1–27.
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion* (2. Aufl.). München: Pearson Studium.
- Campbell, D. & Fiske, D. (1959). Convergent and discriminant validation by the multitrait- multimethod matrix. *Psychological Bulletin*, 56, 81–105.
- Carroll, J. B. (1982). The measurement of intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (29–120). New York: Academic Press.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. New York: Cambridge University Press.
- Cattell, R.B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin*. 40, 153-193.
- Cattell, R.B. (1971). *Abilities - their structure, growth and action*. Boston: Houghton Mifflin
- Chan, D., Schmitt, N., DeShon, R. P., Clause, C. S. & Delbridge, K. (1997). Reactions to cognitive ability tests: The relationships between race, test performance, face validity perceptions, and test-taking motivation. *Journal of Applied Psychology*, 82, 300–310.

- Chapman, D., Uggerslev, K., Carroll, S., Piasentin, K. & Jones, D. (2005). Applicant attraction to organizations and job choice: A meta-analytic review of the correlates of recruiting outcomes. *Journal of Applied Psychology, 90*, 928–944.
- Cohen, D. (2007). The very Separate Worlds of Academic and Practitioner Publications in Human Resource Management: Reasons for the Divide and Concrete Solutions for Bridging the Gap. *Academy of Management Journal, 50*, 1013–1019.
- Costa, P. & McCrae, R.R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five Factor Inventory. Professional Manual*. Odessa, Fla.: Psychological Assessment Resources.
- Costa, P. T. & McCrae, R.R. (1988). Personality in adulthood: A six-year longitudinal study of self-reports and spouse ratings on the NEO Personality Inventory. *Journal of Personality and Social Psychology, 54*, 853–863.
- Deller, J. & Albrecht, A.-G. (2007). Wie ein Fisch im Trockenen? Zur Praxisrelevanz publizierter Forschungsbeiträge. *Zeitschrift für Personalpsychologie, 6*, 150–159.
- DeNeve, K. M. & Cooper, H. (1998). The happy personality: A metaanalysis of 137 personality traits and subjective well-being. *Psychological Bulletin, 124*, 197–229.
- Digman, J. M. (1989). Five robust trait dimensions: Development, stability, and utility. *Journal of Personality, 57*, 195–214.
- DIN Deutsches Institut für Normung e.V. (2002). *Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen – DIN 33430*. Berlin: Beuth.
- Dürndorfer, M., Nink, M. & Wood, G. (2005). *Human Capital Management in deutschen Unternehmen: Eine Studie von Gallup und der Value Group*. Hamburg: Murmann Verlag.
- Eleftheriou, A. & Robertson, I. (1999). A Survey of Management Selection Practices in Greece. *International Journal of Selection and Assessment, 7*, 203–208.
- Fletcher, C. (1991). Candidates' reaction to assessment centres and their outcomes: A longitudinal study. *Journal of Occupational Psychology, 64*, 117–127.
- Franken, T. & Melzer, M. (2007). Rechtliche Grundlagen der Eignungsbeurteilung. In M. John & G. W. Maier (Hrsg.), *Eignungsdiagnostik in der Personalarbeit: Grundlagen, Methoden, Erfahrungen* (S. 73–100). Wiesbaden: Symposium.
- Friedman, H. S., Tucker, J. S., Schwartz, J. E., Martin, L. R., Tomlinson-Keasey, C., Wingard, D. L. & Criqui, M. H. (1995). Childhood conscientiousness and longevity: Health behaviors and cause of death. *Journal of Personality and Social Psychology, 68*, 696–703.

- Gierschmann, F. (2005). Evaluation von Auswahl- und Potenzial Assessment Centern – Beispiele der Deutschen Post. In K. Sünderhauf, S. Stumpf & S. Höft (Hrsg.), *Assessment Center. Von der Auftragsklärung bis zur Qualitätssicherung ; ein Handbuch von Praktikern für Praktiker* (S. 373–388). Lengerich: Pabst Science Publ.
- Gilliland, S. W. (1993). The Perceived Fairness of Selection Systems: An Organizational Justice Perspective. *The Academy of Management Review*, 18, 694–734.
- Goffin, R.D., Rothstein, M.G. & Johnston, N.G. (1996). Personality testing and the assessment center: incremental validity for managerial selection. *Journal of Applied Psychology*, 81, 746–756.
- Göhs, N. & Dick, J. (2001). Testverfahren bei der Personalauswahl Qualitätssuche im intransparenten Markt. *Personal - Zeitschrift für Human Resource Management*, 53, 46–48.
- Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. J. Deary, F. de Fruyt & F. Ostendorf (Ed.), *Personality Psychology in Europe* (7. Aufl., S. 7–28). Tilburg, the Netherlands: Tilburg University Press.
- Görlich, Y. & Schuler, H. (2006). Personalentscheidungen, Nutzen und Fairness. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (2. Aufl., S. 797–840). Göttingen: Hogrefe.
- Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, 24, 79–132.
- Gottfredson, L. S. (2002). Where and why g matters: Not a mystery. *Human Performance*, 15, 25–46.
- Gottfredson, L. S. (2004). Is it the epidemiologists' elusive „Fundamental cause“ of social class inequalities in health? *Journal of Personality and Social Psychology*, 86, 174–199.
- Guion, R. M. (1998). *Assessment, Measurement, and Prediction for Personnel Decisions*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Guion, R. M. & Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology*, 18, 135–164.
- Hardison, C. M. & Sackett, P. R. (2007). Kriteriumsbezogene Validität des Assessment Centers: lebendig und wohlauf? In H. Schuler (Hrsg.), *Assessment Center zur Potenzialanalyse* (S. 192–202). Göttingen: Hogrefe.
- Hausknecht, J. P., Day, D. V. & Thomas, S. C. (2004). Applicant Reactions to Selection Procedures: An Updated Model and Meta-Analysis. *Personnel Psychology*, 57, 639–683.
- Hell, B., Boramir, I., Schaar, H. & Schuler, H. (2006). Interne Personalauswahl und Personalentwicklung in deutschen Unternehmen. *Wirtschaftspsychologie aktuell*, 8, 2–22.

- Hermelin, E., Lievens, F. & Robertson, I. T. (2007). The validity of assessment centres for the prediction of supervisory performance ratings: A meta-analysis. *International Journal of Selection and Assessment*, 15, 405–411.
- Highhouse, S. (2008). Stubborn reliance on intuition and subjectivity in employee selection. *Industrial and Organizational Psychology. Perspectives on Science and Practice*, 1, 333–342.
- Höft, S. (2007). Assessment Center. In H. Schuler (Hrsg.), *Handbuch der Arbeits- und Organisationspsychologie* (S. 475–482). Göttingen: Hogrefe.
- Höft, S. & Obermann, C. (2010). Praxiseinsatz von Assessment Centern im deutschsprachigen Raum: Eine zeitliche Verlaufsanalyse basierend auf den Anwenderbefragungen des Arbeitskreises Assessment Center e.V. von 2001 und 2008. *Wirtschaftspsychologie aktuell*, 2, 5–16.
- Holzenkamp, M., Spinath, F. & Höft, S. (2010). Wie valide sind Assessment Center im deutschsprachigen Raum? Eine Überblicksstudie mit Empfehlungen für die AC-Praxis. *Wirtschaftspsychologie aktuell*, 2, 17–25.
- Hossiep, R., Turck D. & Hasella M. (2001). *Bochumer Matrizentest - advanced - short version (BOMAT - A)*. Göttingen: Hogrefe.
- Hough, L. (1992). The 'Big Five' personality variables - construct confusion: Description versus prediction. *Human Performance*, 5, 139–155.
- Hough, L. M. (2001). I/O owes its advances to personality. In W. Roberts & R. Hogan (Ed.), *Personality and Psychology in the Workplace* (19–44). Washington, DC: American Psychological Association.
- Hough, L. M. & Oswald, F. L. (2000). Personnel selection: Looking toward the future-remembering the past. *Annual Review of Psychology*, 57, 631–664.
- Hülshager, U. R. & Maier, G. W. (2008). Persönlichkeitseigenschaften, Intelligenz und Erfolg im Beruf: Eine Bestandsaufnahme internationaler und nationaler Forschung. *Psychologie Rundschau*, 59, 108–122.
- Hülshager, U. R., Maier, G. W., Stumpp, T. & Muck, P. M. (2006). Vergleich kriteriumsbezogener Validitäten verschiedener Intelligenztests zur Vorhersage von Ausbildungserfolg in Deutschland. *Zeitschrift für Personalpsychologie*, 5, 145–162.
- Hunter, J. E. (1986). Cognitive ability, cognitive aptitude, jobknowledge, and job performance. *Journal of Vocational Behavior*, 29, 340–362.
- Hunter, J. E. & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72–98.

- Hunthausen, J. M., Truxillo, D. M., Bauer, T. N. & Hammer, L. B. (2003). A field study of frame-of-reference effects on personality test validity. *Journal of Applied Psychology, 88*, 545–551.
- Hurtz, G. M. & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology, 85*, 869–879.
- Ispas, D., Ilie, A., Iliescu, D., Johnson, R. & Harris, M. (2010). Fairness reactions to selection methods: A Romanian study. *International Journal of Selection and Assessment, 18*, 102–110.
- Jäger, A., Süß, H.-M. & Beauducel, A. (1997). *BIS - Berliner Intelligenzstruktur-Test Form 4*. Göttingen: Hogrefe.
- Judge, T. A., Higgins, C. A., Thoresen, C. J. & Barrick, M. R. (1999). The Big Five personality traits, general mental ability, and career success across the life span. *Personnel Psychology, 52*, 621–652.
- Judge, T. A., Kammeyer-Mueller, J. & Bretz, R. D. (2004). A longitudinal model of sponsorship and career success: A study of industrial-organizational psychologists. *Personnel Psychology, 57*, 271–303.
- Kaehler, B. (2006). Individualrechtliche Zulässigkeit des Einsatzes psychologischer Testverfahren zu Zwecken der betrieblichen Bewerberauswahl. *Der Betrieb, 5*, 277–288.
- Kanning, U. P., von Rosenstiel, L., Schuler, H. & Petermann, F. (2007). Angewandte Psychologie im Spannungsfeld zwischen Grundlagenforschung und Praxis – Plädoyer für mehr Pluralismus. *Psychologische Rundschau, 58*, 238–248.
- Kersting, M. (2006). Zur Beurteilung der Qualität von Tests: Resümee und Neubeginn. *Psychologie Rundschau, 57*, 243–253.
- Kersting, M. (2008). Zur Akzeptanz von Intelligenz- und Leistungstests. *Report Psychologie, 33*, 420–433.
- Kersting, M. (2010a). Akzeptanz von Assessment Centern: Was kommt an und worauf kommt es an? *Wirtschaftspsychologie aktuell, 12*, 58–65.
- Kersting, M. (2010b). Tests und Persönlichkeitsfragebogen in der Personalarbeit. *Personalführung, 10*, 20–31.
- Kersting, M. (2011 a). Schlaraffenland ist abgebrannt – warum sich die Bewerberauswahl ändern muss. *Wirtschaftspsychologie aktuell, 4*, 23–26.
- Kersting, M. (2011 b). Personalgewinnung und Personalauswahl im öffentlichen Dienst: Potpourri der Methoden. *Personal, 1*, 10–12.
- Kleinmann, M. (1993). Are rating dimensions in assessment centers transparent for participants? Consequences for criterion and construct validity. *Journal of Applied Psychology, 78*, 988–993.

- Kleinmann, M. (2003). *Das Assessment Center*. Göttingen: Hogrefe Verlag.
- Kleinmann, M. & Köller, O. (1997). Construct validity of assessment centers: Appropriate use of confirmatory factor analysis and suitable construction principles. *Journal of Social Behavior and Personality*, 12, 65–84.
- Kleinmann, M., Kuptsch, C. & Köller, O. (1996). Transparency: A necessary requirement for the construct validity of assessment centres. *Applied Psychology: An international Review*, 45, 67–84.
- Klingner, Y. & Schuler, H. (2004). Improving participants' evaluations while maintaining validity by a work sample-intelligence test hybrid. *International Journal of Selection and Assessment*, 12, 120–134.
- Köchling, A. C. & Körner, S. (1996). Personalauswahl in deutschen Unternehmen. Zur bewerberorientierten Gestaltung von Beurteilungssituationen. *Zeitschrift für Arbeits- und Organisationspsychologie*, 40, 22–37.
- König, C. J., Klehe, U.-C., Berchtold, M. & Kleinmann, M. (2010). Reasons for being selective when choosing personnel selection procedures. *International Journal of Selection and Assessment*, 18, 17–27.
- Kramer, J. (2009). Allgemeine Intelligenz und beruflicher Erfolg in Deutschland: Vertiefende und weiterführende Metaanalysen. *Psychologie Rundschau*, 60, 82–98.
- Krause, D. E. & Gebert D. (2003). 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., Meyer zu Kniendorf, C. & Gebert, D. (2001). Aktuelle Trends in der Assessment-Center-Anwendung. *Wirtschaftspsychologie*, 3, 47–55.
- Krumm, S., Hüffmeier, J., Dietz, F., Findeisen, A. & Dries, C. (2011). Towards Positive Test Takers' Reactions to Cognitive Ability Assessments: Development and Initial Validation of the Reasoning Ability at Work Test. *Journal of Business and Media Psychology*, 2, 11–18.
- Leitl, M. (2007). *Was ist Humankapital?* [On-line]. Zugriff am 19.12.2012. Verfügbar unter <http://www.harvardbusinessmanager.de/heft/artikel/a-622148.html>.
- Lienert, G. & Raatz, U. (1998). *Testaufbau und Testanalyse*. Weinheim: Beltz.
- Lievens, F. (1998). Factors which improve the construct validity of assessment centers: A review. *International Journal of Selection and Assessment*, 6, 141–152.
- Lievens, F., De Corte, W. & Schollaert, E. (2008). A closer look at the frame-of-reference effect in personality scale scores and validity. *Journal of Applied Psychology*, 93, 268–279.

- Lievens, F. & Klimoski, R.J. (2001). Understanding the assessment center process: Where are we now? In C.L. Cooper & I.T. Robertson (Eds.) *International Review of Industrial and Organizational Psychology* vol. 16. (245-286). Chichester: John Wiley & Sons, Ltd.
- Lounsbury, J. W., Bobrow, W. & Jensen, J. B. (1989). Attitudes Toward Employment Testing: Scale Development, Correlates, and »Known-Group« Validation. *Professional Psychology: Research and Practice*, 20, 340–349.
- Macan, T., Avedon, M., Paese, M. & Smith, D. (1994). The effects of applicants' reactions to cognitive ability tests and an assessment center. *Personnel Psychology*, 47, 715–738.
- Maier, C. & Antoni, C. (2009). Arbeits- und organisationspsychologische Forschung und betriebliche Praxis – zwei ferne Königskinder? *Zeitschrift für Arbeits- und Organisationspsychologie*, 53, 22–32.
- Marcus, B. (2003a). Attitudes toward personnel selection methods: A partial replication and extension in a German sample. *Applied Psychology: An International Review*, 52, 515-532.
- Marcus, B. (2003b). Persönlichkeitstests in der Personalauswahl: Sind „sozial erwünschte“ Antworten wirklich nicht wünschenswert? *Zeitschrift für Psychologie*, 211, 138-148.
- McCrae, R. R. & Costa, P. T. (1997). Personality trait structure as a human universal. *American Psychologist*, 52, 509–516.
- McCrae, R. R. & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, 2, 175–215.
- Moscose, S. a. S. J. (2004). Fairness reactions to personnel selection techniques in Spain and Portugal. *International Journal of Selection and Assessment*, 12, 187–196.
- Mount, M. K., Barrick, M. R. & Stewart, G. (1998). Five-factor model of personality and performance in jobs involving interpersonal interactions. *Human Performance*, 11, 145–165.
- Muck, P. M. & Stumpp, T. (2007). Persönlichkeits- und Interessenstests. In M. John & G. W. Maier (Hrsg.), *Eignungsdiagnostik in der Personalarbeit: Grundlagen, Methoden, Erfahrungen* (S. 173–197). Wiesbaden: Symposium.
- Murphy, K. (1986). When your top choice turns you down: Effect of rejected job offers on the utility of selection tests. *Psychological Bulletin*, 99, 133–138.
- Mussel, P., Frintrup, A., Pfeiffer, K., & Schuler, H. (2007). Vorauswahlmethoden für Assessment Center: Referenzmodell und Anwendung. In H. Schuler (Hrsg.), *Assessment Center zur Potenzialanalyse* (S. 330–345). Göttingen: Hogrefe.
- Neisser, U., Boodoo, G., Bouchard, T. J. Jr., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J. & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51, 77–101.

- Nevo, B. & Sfez, J. (1986). Examinees' Feedback Questionnaire (EFeQ). In B. Nevo & J. Sfez (Ed.), *Psychological Testing: The examinee perspective* (21–47). Göttingen: Hogrefe.
- Newell, S. & Shackleton, V. (1994). Guest Editorial: International differences in selection methods. *International Journal of Selection and Assessment*, 2, 71–73.
- Ng, T. W. H., Eby L. T., Sorensen K. L. & Feldman D. C. (2005). Predictors of objective and subjective career success: A meta-analysis. *Personnel Psychology*, 58, 367–408.
- Nikolaou, I. & Judge, T. (2007). Fairness reactions to personnel selection techniques in Greece. *International Journal of Selection and Assessment*, 15, 206–219.
- Noe, R. A. & Steffy, B. D. (1987). The influence of individual characteristics and assessment centre evaluation on career exploration behavior and job involvement. *Journal of Vocational Behavior*, 30, 187–203.
- Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal Psychology*, 66, 574–583.
- Obermann, C. (2006). *Assessment Center: Entwicklung, Durchführung, Trends*. Wiesbaden: Gabler.
- Organ, D. W. & Ryan, K. (1995). A meta-analytic review of attitudinal and dispositional predictors of organizational citizenship behavior. *Personnel Psychology*, 48, 775–802.
- Pfeffer, J. & Sutton, R. I. (2006). *Hard facts, dangerous half-truths and total nonsense: Profiting from evidence-based management*. Boston, MA: Harvard Business School Press.
- Ployhart, R. & Ryan, A. (1998). Applicants' reactions to the fairness of selection procedures: The effects of positive rule violations and time of measurement. *Journal of Applied Psychology*, 83, 3–16.
- Preis, U. (2005). Abschnitt 8. Einzelne Schuldverhältnisse. In T. Dieterich, R. Müller-Glöge, U. Preis & G. Schaub (Hrsg.), *Erfurter Kommentar zum Arbeitsrecht* (5. Aufl., S. 1378–1530). München: Beck.
- Robertson, I. T. & Kinder, A. (1993). Personality and job competencies: The criterion-related validity of some personality variables. *Journal of Occupational and Organizational Psychology*, 66, 226–244.
- Rustenbach, S. J. (2003). *Metanalyse. Eine anwendungsorientierte Einführung*. Bern: Huber.
- Ryan, A. & Chan, D. (1999). Perceptions of the EPPP: How do licensure candidates view the process? *Professional Psychology: Research and Practice*, 30, 519–530.
- Ryan, A. & Huth, M. (2008). Not Much More than Platitudes? A critical look at the utility of applicant reactions research. *Human Resource Management Review*, 18, 119–132.

- Ryan, A. M., McFarland, L., Baron, H. & Page, R. (1999). An international look at selection practices: Nation and culture as explanations for variability in practice. *Personnel Psychology, 52*, 359–391.
- Ryan, A. & Ployhart, R. (2000). Applicant perceptions of selection procedures and decisions: A critical review and agenda for the future. *Journal of Management, 26*, 565–606.
- Ryan, A., Ployhart, R., Greguras, G. & Schmit, M. (1998). Test preparation programs in selection contexts: Self-selection and program effectiveness. *Personnel Psychology, 51*, 599–621.
- Rynes, S. (1993). Who's selecting whom? Effects of selection practices on applicant attitudes and behavior. In N. Schmitt, W. Bormann & Associates (Ed.), *Personnel Selection in Organizations* (240–274). San Francisco: Jossey-Bass.
- Rynes, S. L. (2007). Let's create a tipping point: What academics and practitioners can do, alone and together. *Academy of Management Journal, 50*, 1046–1054.
- Rynes, S. L., Brown, K. G. & Colbert, A. E. (2002). Seven common misconceptions about human resource practices: Research findings versus practitioners beliefs. *Academy of Management Executive, 16*, 92–103.
- Rynes, S., Giluk, T. & Brown, K. (2007). The Very Separate Worlds of Academic and Practitioner Periodicals in Human Resource Management: Implications for evidence-based management. *Academy of Management Journal, 50*, 987–1008.
- Sackett, P. & Dreher, G. F. (1982). Constructs and assessment center dimensions: Some troubling empirical findings. *Journal of Applied Psychology, 67*, 401–410.
- Sackmann, S. & Elbe, M. (2000). Tendenzen und Ergebnisse empirischer Personalforschung der 90er Jahre in West-Deutschland. *Zeitschrift für Personalforschung, 2*, 131–157.
- Salgado, J. F. (1997). The Five Factor Model of Personality and job performance in the European community. *Journal of Applied Psychology, 82*, 30–43.
- Salgado, J. F. & Anderson, N. (2003). Validity generalization of GMA tests across countries in the European Community. *European Journal of Work and Organizational Psychology, 12*, 1–17.
- Salgado, J.F., Anderson, N., Moscoso, S., Bertua, C. & de Fruyt, F. (2003a). International validity generalization of GMA and cognitive abilities: A european community meta-analysis. *Personnel Psychology, 56*, 573–605.
- Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., Fruyt, F. de & Rolland, J. P. (2003b). A meta-analytic study of generalmental ability validity for different occupations in the European community. *Journal of Applied Psychology, 88*, 1068–1081.

- Sanchez, R., Truxillo, D. & Bauer, T. (2000). Development and examination of an expectancy-based measure of test-taking motivation. *Journal of Applied Psychology, 85*, 739–750.
- Sarges, W. (2001). *Weiterentwicklung der Assessment Center-Methode* (2. Aufl.). Göttingen: Hogrefe.
- Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance, 15*, 187–211.
- Schmidt, F. L. & Hunter, J. E. (1993). Tacit knowledge, practical intelligence, general mental ability, and job knowledge. *Current Directions in Psychological Science, 2*, 8–9.
- 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, E. & Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology, 71*, 432–439.
- Schmit, M. J., Ryan, A. M., Stierwalt, S. L. & Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology, 80*, 607–620.
- Schmitt, N., Gooding, R. Z., Noe, R. A. & Kirsch, M. (1984). Metaanalyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology, 37*, 407–422.
- Scholarios, D. & Lockyer, C. (1999). Recruiting and Selecting Professionals: Context, qualities and methods. *International Journal of Selection and Assessment, 7*, 142–156.
- Schuler, H. (2002). *Wirtschaftspsychologie. Das Einstellungsinterview*. Göttingen: Hogrefe Verlag für Psychologie.
- Schuler, H. (2007a). *Assessment Center zur Potenzialanalyse*. Göttingen: Hogrefe.
- Schuler, H. (2007 b). Spielwiese für Laien? Weshalb das Assessment-Center seinem Ruf nicht mehr gerecht wird. *Wirtschaftspsychologie aktuell, 2*, 27–30.
- Schuler, H., Frier, D. & Kauffmann, M. (1993). Beiträge zur Organisationspsychologie: Personalauswahl im europäischen Vergleich. Göttingen: Verl. für Angewandte Psychologie.
- Schuler, H., Hell, B., Trapmann, S., Schaar, H. & Boramir, I. (2007). Die Nutzung psychologischer Verfahren der externen Personalauswahl in deutschen Unternehmen. Ein Vergleich über 20 Jahre. *Zeitschrift für Personalpsychologie, 6*, 60–70.

- Schulz, C., Schuler, H. & Stehle, W. (1985). Die Verwendung eignungsdiagnostischer Methoden in deutschen Unternehmen. In H. Schuler & W. Stehle (Hrsg.), *Organisationspsychologie und Unternehmenspraxis: Perspektiven der Kooperation* (S. 126-132). Göttingen: Hogrefe Verlag für angewandte Psychologie.
- Shackleton, V. & Newell, S. (1994). European Management Selection Methods: A Comparison of Five Countries. *International Journal of Selection and Assessment*, 2, 91–102.
- Silverman, W., Dalwssio, A., Woods, S. & Johnson, R. (1986). Influence of assessment center methods on assessor's ratings. *Personnel Psychology*, 39, 565–578.
- Smither, J. W., Reilly, R. R., Millsap, R. E., Pearlman, K. & Stoffey, R. W. (1993). Applicant Reactions to Selection Procedures. *Personnel Psychology*, 46, 49–76.
- Spearman, C. (1904). „General Intelligence“ objectively determined and measured. *American Journal of Psychology*, 15, 201–293.
- Spearman, C. (1927). *The abilities of man*. London: MacMillan
- Spychalski, A., Quiñones, M., Gaugler, B.B. & Pohley, K. (1997). A survey of assessment center practices in organizations in the United States. *Personnel Psychology*, 50, 71–90.
- Steiner, D. D. & Gilliland, S. W. (1996). Fairness reactions to personnel selection techniques in France and the United States. *Journal of Applied Psychology*, 81, 134–141.
- Steiner, D. & Gilliland, S. (2001). Procedural justice in personnel selection: International and cross-cultural perspectives. *International Journal of Selection and Assessment*, 9, 124–137.
- Task Force on Assessment Center Guidelines. (1989). Guidelines for ethical considerations. *Public Personnel Management*, 18, 457–470.
- Taylor, H. C. & Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. *Journal of Applied Psychology*, 23, 565–578.
- Tett, R. P., Jackson, D. N. & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology*, 44, 703–742.
- Thornton III, G. C., Gaugler, B., Rosenthal, D. B. & Bentson, C. (1987). Die prädiktive Validität des Assessment Centers – eine Metaanalyse. In H. Schuler & W. Stehle (Hrsg.), *Assessment Center als Methode der Personalentwicklung* (S. 36–60). Göttingen: Verlag für Angewandte Psychologie.
- Thurstone, L. (1938). *Primary Mental Abilities*. Chicago: University of Chicago Press.
- Thurstone, L. L., & Thurstone, T. G. (1941). *Factorial studies of intelligence*. Chicago: University of Chicago Press.
- Tupes, E. C. & Christal, R. E. (1961). *Recurrent personality factors based on trait ratings* (Technical Report ASD-TR-61-97). Lackland Air Force Base, Texas: U.S. Air Force.

- Van Emmerik, I., Bakker, A. & Euwema, M. (2008). What happens after the developmental assessment center? Employees' reactions to unfavorable performance feedback. *Journal of Management Development*, 27, 513–527.
- Vernon, P.E. (1950). *The structure of human abilities*. London: Methuen
- Vernon, P.E. (1965). *Ability factors and environmental influences*. London: Methuen
- Vinchur, A. J., Schippmann, J. S., Switzer, F. S. & Roth, P. L. (1998). A meta-analytic review of predictors of job performance for salespersons. *Journal of Applied Psychology*, 83, 586–597.
- Warmke, D. L. (1985). Preselection for assessment centers: Some choices and issues to consider. *Journal of Management Development*, 4, 28–39.
- Weiler & Knipp (2012). Welche Rolle spielt der Betriebsrat im Recruiting-Prozess? *arecruiter*, 9, 52-54.
- Woehr, D. J. & Arthur W. (2003). The construct related validity of assessment center ratings: A review and meta-analysis of the role of methodological factors. *Journal of Management*, 29, 231–258.

Danksagung

Mit der Fertigstellung dieser Dissertation ist es an der Zeit, nochmals all denjenigen herzlichst zu danken, die mich begleitet haben, und ohne deren Unterstützung diese Arbeit nicht zustande gekommen wäre. Besonderer Dank gilt dabei zunächst meiner Doktormutter Frau Prof. Dr. Renate Soellner, die nicht nur die Qualität der vorliegenden Arbeit durch kritische sowie inspirierende Fachdiskussionen erhöht hat, sondern mich auch im Hinblick auf meine fachliche, berufliche und persönliche Weiterentwicklung stets gefördert hat. Die bislang drei Jahre Mitarbeitertätigkeit in ihrer Arbeitsgruppe waren für mich eine lehrreiche und glückliche Zeit; insbesondere aufgrund der ausgesprochen angenehmen, vertrauensvollen und wertschätzenden Atmosphäre. Nicht weniger zu danken gilt es meinem Doktorvater, Herrn Prof. Dr. Andreas Mojzisch, für die engagierte, lehrreiche und inspirierende Begleitung meines Promotionsvorhabens. Beide haben mir mit fachlichem Rat und auch mit motivationaler sowie emotionaler Unterstützung stets zur Seite gestanden. Unter ihrer Anleitung habe ich viel gelernt und konnte mich stets weiterentwickeln. Hierfür danke ich ihnen vom ganzen Herzen.

Mein Dank gilt ebenso meinen lieben KollegInnen vom Institut für Psychologie der Universität Hildesheim; insbesondere bedanke ich mich bei meiner Arbeitsgruppe: Astrid Bräker, Dr. Monika Ludwig, Maren Reder, Anna Machmer und Dr. Godehard Henze für die freundschaftliche, hilfsbereite und unterstützende Zusammenarbeit. Ich möchte darüber hinaus einerseits Frau Dr. Ann-Katrin Bockmann für ihre emotionale und fachliche Unterstützung in unterschiedlichen Phasen während der Fertigstellung dieser Arbeit danken. Andererseits gilt mein Dank dem Dekan Herrn Prof. Dr. Werner Greve für sein unermüdliches Engagement das Institut für Psychologie an der Universität Hildesheim nachhaltig zu gestalten sowie auszubauen und ich freue mich, dass ich einen Teil dieses Prozesses begleiten konnte.

Darüber hinaus bedanke ich mich herzlichst bei allen Personen, die an den unterschiedlichen Studien teilgenommenen haben. Mein Dank richtet sich somit an über 200 Personalverantwortliche und an über 350 Studierende, ohne deren freiwillige Teilnahme und Unterstützung diese Forschungsarbeit nicht möglich gewesen wäre.

Ebenso danke ich meiner Schwester, meinem Schwager, Anna-Lena, Ben-Ole sowie all meinen Freunden für die langjährige Begleitung, Aufmunterung und Unterstützung in guten wie schlechteren Zeiten.

Nicht zuletzt bin ich meinen Eltern für ihre unentwegte emotionale wie tatkräftige Unterstützung in jeder Lebenssituation und ihr uneingeschränktes Vertrauen zu tiefstem Dank verpflichtet. Ihnen sei die vorliegende Arbeit von ganzem Herzen gewidmet.

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Dissertation selbstständig, ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel verfasst habe. Aus anderen Quellen wörtlich oder inhaltlich entnommene Stellen sind als solche kenntlich gemacht. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Hannover, den 30. Dezember 2012



Nils Benit

Anhang

Anhang A:

Benit, N. & Soellner, R. (2012). Validität von Assessment Centern in deutschen Unternehmen: Vergleich von Unternehmensdaten mit einer bestehenden Metaanalyse. *PERSONALFÜHRUNG*, 11, 32–39.

Anhang B:

Benit, N., Mojzisch, A. & Soellner, R. (resubmitted). Pre-Selection Methods Prior to the Internal Assessment Center. *Journal of Personnel Psychology*.

Anhang C:

Benit, N. & Soellner, R. (submitted). Scientist-practitioner gap in Deutschland: Eine empirische Studie am Beispiel der Prognostizität psychologischer Testverfahren. *Zeitschrift für Arbeits- und Organisationspsychologie*.

Anhang D:

Benit, N. & Soellner, R. (2012). Misst gut, ist gut? Vergleich eines abstrakten und eines berufsbezogenen Matrizentests. *Journal of Media & Business Psychology*, 3 (1), 22–29.

Anhang E:

Instruktionen und Beispielaufgaben aus: Benit, N. (2012). *Analogiebezogene Kognitiv Numerische DatenAnalyse (ANAKONDA)*. Solingen: U-Form Verlag.

Hinweis:

Aufgrund der Veröffentlichung der Manuskripte in den entsprechenden Fachzeitschriften sind die rechtlichen Bestimmungen der Verlage zu wahren, so dass die Anhänge nicht Bestandteil dieses Dokumentes sind. Die Manuskripte sind über die entsprechenden Verlage beziehbar.

