



Misst gut, ist gut? Vergleich eines abstrakten und eines berufsbezogenen Matrizen-tests

Nils Benit & Renate Soellner

Universität Hildesheim

ZUSAMMENFASSUNG

Obwohl Intelligenztests berufliche Erfolgsindikatoren valide vorherzusagen vermögen (Kramer, 2009; Schmidt & Hunter, 1998), werden sie in deutschen Unternehmen u.a. aufgrund geringer Akzeptanz selten eingesetzt (Schuler, Hell, Trapmann, Schaar & Boramir, 2007). Die vorliegende Studie untersucht, ob die Akzeptanz eines Matrizen-tests allein durch eine berufsbezogene Symbolgestaltung erhöht werden kann und der Test trotzdem testtheoretischen Gütekriterien genügt. Der BOMAT (Hossiep, Turck & Hasella, 2001), ein valider aber abstrakter Matrizen-test, wird mit dem BBM, einem für Forschungszwecke entwickelten Matrizen-test mit berufsbezogenen Symbolen, kontrastiert. Beide Verfahren erwiesen sich in der vorliegenden Stichprobe ($N = 360$) als reliabel und valide, insbesondere hinsichtlich hoher retrospektiver Validitäten mit den Kriterien Abiturnote ($r_{\text{BOMAT}} = -.45$; $r_{\text{BBM}} = -.43$) und Bachelornote ($r_{\text{BOMAT}} = -.41$; $r_{\text{BBM}} = -.39$). Hohe Korrelationen beider Verfahren untereinander sowie mit dem IST-2000R und niedrige Korrelationen mit dem NEO-FFI sprechen für Konstruktvalidität. Die berufsbezogene Gestaltung erwies sich der abstrakten Gestaltung bei gleichwertiger Messqualität, Kontrollierbarkeit und Belastungsfreiheit (AKZEPT!-L: Kersting, 2008) hinsichtlich der Augenscheinvalidität als überlegen.

Schlüsselbegriffe: Akzeptanz, Matrizen-test, Intelligenztest, berufliche Augenscheinvalidität, Validität

1 Intelligenztests in der Personalauswahl

Intelligenztests stellen in US-amerikanischen und anderen europäischen Unternehmen etablierte Instrumente zur Personalauswahl dar. Allerdings werden sie in deutschen Unternehmen vergleichsweise selten eingesetzt (Ryan, McFarland, Baron & Page, 1999). Seit Jahrzehnten liegt die Einsatzhäufigkeit bei ca. 30% (Göhs & Dick, 2001; Sackmann & Elbe, 2000; Schuler, Frier & Kauffmann, 1993; Schuler, Hell, Trapmann, Schaar & Boramir, 2007). Dieser stabile Befund verwundert, da mittlerweile in Metaanalysen übereinstimmend gezeigt werden konnte, dass Intelligenz einen überaus validen Prädiktor für Erfolgsindikatoren von Ausbildung und Beruf darstellt.

Metaanalytische Befunde hierzu stammten in den letzten drei Dekaden überwiegend aus den USA (Hunter & Hunter, 1984; Schmidt & Hunter, 1998). Seit diesem Jahrtausend ist jedoch eine zunehmende Forschungsaktivität auch in Deutschland und anderen europäischen Ländern erkennbar (Hülshager, Maier, Stumpp & Muck, 2006; Kramer, 2009; Salgado & Anderson, 2003). Somit kann die geringe Anwendungshäufigkeit von Intelligenztests in deutschen Unternehmen grundsätzlich nicht auf eine unzureichende oder länderunspezifische Befundlage zurückgeführt werden (vgl. Ackermann, 1994; Borman, Penner, Allen & Motowidlo, 2001; Salgado, Anderson, Moscoso, Bertua & de Fruyt, 2003). Ebenfalls existieren in Deutschland diverse Testverfahren zur Messung der Intelligenz, die im Rahmen von Berufseignungsprozessen erprobt wurden (vgl. Hülshager, Maier, Stumpp & Muck, 2006; Sarges & Wottawa, 2004). Ein quantitativer Mangel an Intelligenztests scheint demnach auch nicht für die ver-

gleichsweise geringe Einsatzhäufigkeit verantwortlich zu sein. Allerdings unterscheiden sich die Verfahren hinsichtlich der zugrunde liegenden Theorie von Intelligenz, ihrer Ökonomie, Qualität, beruflicher Augenscheinvalidität und Akzeptanz (Hell, 2003; Kersting, 2008).

In der wissenschaftlichen Debatte wird u. a. die Akzeptanz als mögliche Ursache für die geringe Anwendungshäufigkeit von Intelligenztests in der Unternehmenspraxis angeführt (König, Klehe, Berchtold & Kleinmann, 2010; Schuler et al., 2007). Während Einstellungsinterviews oder Assessment Center von Organisationen und Bewerbern als hoch akzeptiert bewertet und häufig als Auswahlverfahren eingesetzt werden, wird Intelligenztests lediglich eine geringe Akzeptanz attestiert (Anderson & Witvliet, 2008; Hausknecht, Day & Thomas, 2004; Hell, Boramir, Schaar & Schuler, 2006; König et al., 2010; Schuler et al., 2007; Steiner & Gilliland, 1996).

Die Thematik *Akzeptanz von Auswahlverfahren* und damit verbundene *Bewerberreaktionen* erfuhr in den vergangenen Jahren ein zunehmendes Forschungsinteresse (Anderson, 2004; Ryan & Huth, 2008), welches in zahlreichen Forschungsarbeiten mündete (vgl. Hülshager & Anderson, 2009). Bereits in den 1990er Jahren wurden tragfähige Akzeptanz-Konzepte im internationalen und deutschsprachigen Raum vorgestellt (Gilliland, 1993; Schuler, 1990). Gilliland (1993) differenziert in seinem „Model of Applicants' Reaction to Employment Selection Systems“ zwischen der Wahrnehmung der Auswahl-situation und der Wahrnehmung der Auswahlentscheidung und erarbeitete zehn Regeln, die bei Missachtung zur Reduzierung der Akzeptanz gegenüber dem Auswahlverfahren führen kön-

nen. Auch Schuler (1990; Schuler & Stehle, 1983) akzentuiert im „Konzept der sozialen Validität“ insbesondere die Interaktion zwischen Bewerbern und Organisationsvertretern. Ryan und Ployhart (2000) präsentieren ein heuristisches Modell in Anlehnung an Gilliland, differenzieren zwischen vier Bewerber-Wahrnehmung-Prozessen (applicant perceptions of the procedure/process itself, of one's affective and cognitive state during the process, of the procedure's outcome, of the selection process in general) und berücksichtigten zusätzlich die Vorerfahrung der Bewerber sowie potentielle Moderatoren zwischen den Vorerfahrungen und der Bewerberwahrnehmung. In jüngerer Vergangenheit findet insbesondere das erweiterte Modell von Hausknecht et al. (2004) starke Beachtung. Im Rahmen ihrer umfassenden Metaanalyse integrierten sie zusätzlich in die bestehenden Ansätze die zwei Faktoren Prüfungsangst (test anxiety) und Prüfungsmotivation (test motivation).

Aus Organisationssicht erscheinen akzeptierte Auswahlverfahren besonders lohnenswert, da eine wenig akzeptierende Haltung und Wahrnehmung der Bewerber gegenüber dem Testverfahren dazu führen kann, dass Bewerber das Organisationsansehen abwerten oder ein Stellenangebot ablehnen (Hausknecht et al., 2004). Neben negativen Auswirkungen auf das Unternehmensimage (Ryan & Ployhart, 2000) oder auf das Arbeitsverhalten der eingestellten Bewerber (Gilliland, 1993) werden Widerstände von Betriebsräten sowie die Debatte um die Rechtmäßigkeit des Einsatzes von Intelligenztests zu eignungsdiagnostischen Zwecken (Hülshager & Maier, 2008) als bedeutsame Gründe für den geringen Einsatz von Intelligenztests in deutschen Unternehmen angeführt. Weitere Befürchtungen sind, dass Bewerber, die sich im Auswahlprozess benachteiligt oder diskriminiert behandelt fühlen, rechtliche Schritte einleiten oder schlimmstenfalls die Medien involvieren können (Anderson, 2011).

Die Akzeptanz eines Verfahrens wird zudem stark durch seine berufliche Augenscheinvalidität beeinflusst. Schuler (2000) akzentuiert diesbezüglich, dass die berufliche Augenscheinvalidität lediglich ein Aspekt sozialer Validität sei und somit nicht den Hauptgütekriterien angehöre, allerdings in ihrer Bedeutung nicht zu unterschätzen sei. „Augenscheinvalidität ist somit ein relationaler Begriff, 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).

Eine geringe berufliche Augenscheinvalidität kann zum einen aus einer abstrakten und kontextfernen Gestaltung der Testverfahren sowie zum anderen aus dem für Bewerber und Organisationsgremien nicht eindeutig erkennbaren Bezug zu den Arbeitsanforderungen des Arbeitsplatzes resultieren (Eleftheriou & Robertson, 1999). Damit einher geht häufig eine wenig akzeptierende Haltung und Wahrnehmung gegenüber dem Testverfahren (Hausknecht et al., 2004). Smither, Reilly, Millsap, Pearlman und Stoffey (1993) konnten zeigen, dass die wahrgenommene Augenscheinvalidität der Aufgaben positiv mit der Testleistung korrelierte. Auch Chan, Schmitt, Deshon, Clause und Delbridge (1997) unterstrichen in einem längsschnittlichen

Design den Zusammenhang zwischen beruflicher Augenscheinvalidität eines Intelligenztests und der Testmotivation, die sich wiederum positiv auf die Testleistung auswirkte. Zusätzlich konnte gezeigt werden, dass eine arbeits-spezifische Testmaterialdarbietung auch zu einem Validitätsanstieg beitragen kann (Schmit, Ryan, Stierwalt & Powell, 1995; vgl. auch Bing, Whanger, Davison & Van Hook, 2004; Hunthausen, Truxillo, Bauer & Hammer, 2003; Lievens, De Corte & Schollaert, 2008). In diesem Zusammenhang ist seit längerer Zeit die Forderung nach mehr wissenschaftlich fundierten Testverfahren mit hoher beruflicher Augenscheinvalidität und direktem Anforderungsbezug zu beobachten (vgl. Hülshager & Maier, 2008; Kramer, 2009), verbunden mit der Hoffnung, dass diese dann auch in deutschen Unternehmen vermehrt Einzug finden.

Intelligenztests mit hoher beruflicher Augenscheinvalidität und direktem Anforderungsbezug stellen nach Kersting (2006, 2008) eher die Ausnahme als die Regel auf dem intransparenten Testmarkt dar. Jüngere Publikationen unterstreichen vermehrt die wissenschaftlichen Bemühungen, fundierte Intelligenztests mit hoher Augenscheinvalidität zu konstruieren und validieren (vgl. Analytischer Test: Ostapczuk, Musch & Lieberei, 2011; Reasoning Ability at Work Test: Krumm, Hüffmeier, Dietz, Findeisen & Dries, 2011). Ob diese Verfahren aber auch tatsächlich ihren Weg in die Unternehmenspraxis finden werden, bleibt abzuwarten.

2 Untersuchungsvorhaben

Ein Intelligenztest, der explizit für den Einsatzbereich der Eignungsdiagnostik und Personalentwicklung konstruiert wurde, ist der BOMAT – advanced – short version (Hossiep, Turck & Hasella, 2001). Der BOMAT entspricht in seiner Form einem Matrizen-test und zielt auf die Erfassung der Allgemeinintelligenz ab. Der BOMAT wurde als sprachfreies Instrument entwickelt, dient der Erfassung der Intelligenzkapazität im höheren Leistungsbereich im Sinne des Spearman'schen „g-Faktors“ (Spearman, 1904) und erwies sich als hinreichend objektives, reliables sowie valides Instrument (vgl. Hossiep, Turck & Hasella, 2004). Die berufliche Augenscheinvalidität des BOMAT ist allerdings aufgrund seiner abstrakten und kontextfernen Gestaltung als gering zu bezeichnen. Seine Akzeptanz wurde bisher lediglich über Mittelwertsunterschiede nach verschiedenen Kriterien wie Geschlecht und Muttersprache empirisch überprüft.

Ziel der vorliegenden Studie war es zu untersuchen, ob sich bereits eine kontext- bzw. berufsspezifische Symbolgestaltung in Matrizen-tests positiv auf die Akzeptanz-schätzung auswirkt. Hierfür wurde die Akzeptanz und psychometrische Güte des BOMAT im Vergleich zu einem eigens entwickelten berufsbezogenen Matrizen-test (BBM) überprüft, der auf der Grundlage der Testaufgaben des BOMAT neu entwickelt wurde. Während der BOMAT sich aus schwarz-weißen geometrischen Figuren zusammensetzt, besteht der BBM aus berufsbezogenen Symbolen aus der Automobil-Branche mit unterschiedlichen Farbausprägungen. Eine Beispielmatrix aus dem BBM, die in Anlehnung an den BOMAT konstruiert wurde, kann Abbildung 1 entnommen werden. Für die Konstruktion des BBM

wurde eine repräsentative Auswahl von Items aus beiden Parallelversionen des BOMAT getroffen, die sowohl das Spektrum unterschiedlicher Schwierigkeitsindizes abdeckte als auch eine zufriedenstellende Trennschärfe laut Testmanual aufwies. Einige Aufgabenformate aus dem BOMAT (z.B.: „Sukzessives Folgen“) konnten nicht adäquat in ein berufsbezogenes Design überführt werden. Da hohe Akzeptanzwerte alleine noch keinen reliablen und validen Test ausmachen, wurden beide Instrumente sowohl testtheoretisch (Reliabilität, Validität) als auch hinsichtlich ihrer Akzeptanz überprüft.

Beantwortet werden sollte, ob der erstmals zu Forschungszwecken eingesetzte BBM durch eine berufsbezogene Symbolgestaltung eine höhere Akzeptanz im Vergleich zum BOMAT erzielen und dessen ungeachtet eine ebenfalls zufriedenstellende psychometrische Güte aufweisen kann. Anders formuliert stellten wir uns die Fragen, ob der BBM *gut misst* (also ein reliables und valides Testverfahren darstellt) und aus Sicht der Probanden auch *gut ist* (also akzeptiert wird).

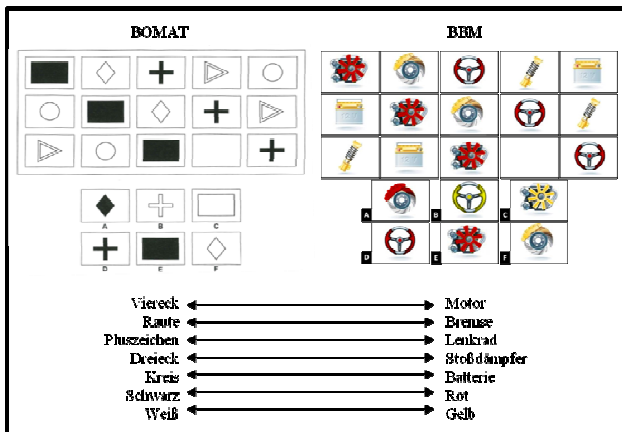


Abb. 1: Kontrastierung der Symbole im BBM und BOMAT

3 Methode

3.1 Material

3.1.1 Bochumer Matrizentest

Der Bochumer Matrizentest – advanced – short version (Hossiep et al., 2001) stellt eine gekürzte Fassung des BOMAT – advanced – aus dem Jahre 1999 dar. Jede der 28 Aufgaben setzt sich aus einer 15 Felder-Matrize zusammen, bei der ein Feld ausgespart ist. Im Auswahlprozess besteht die Aufgabe der Bewerberinnen und Bewerber darin, das freie Feld entsprechend des logischen Prinzips der restlichen 14 Felder zu ergänzen. Dafür stehen den Bewerbern sechs Antwortmöglichkeiten zur Auswahl, wovon lediglich eine korrekt ist. Da der BOMAT eindimensional das abstrakt-logische Denkvermögen erfassen soll, liegen keine Subtests vor.

3.1.2 Berufsbezogener Matrizentest

Der berufsbezogene Matrizentest (BBM) wurde ausschließlich zu Forschungszwecken konstruiert und soll wie sein Pendant, der BOMAT, eindimensional das abstrakt-logische Denkvermögen messen. Der BBM besteht aus 22 Items, die nach dem Aufgabenformat des BOMAT best-

möglich adaptiert worden sind und einen Vergleich der beiden Matrizenformate ermöglichen sollen.

3.1.3 Intelligenz-Struktur-Test 2000R

Zur Überprüfung der konvergenten Validität des BBM und BOMAT wurden zusätzlich die Subtests *Matrizen*, *Analogie* und *Zahlenreihe* aus dem IST-2000R (Liepmann, Beauducel, Brocke & Amthauer, 2007) eingesetzt. Die Subtests setzen sich jeweils aus 20 Aufgaben zusammen und erfassen figural-räumliche, verbale bzw. numerische Intelligenzfacetten.

3.1.4 NEO-Fünf-Faktoren-Inventar

Zur Überprüfung der divergenten Validität des BBM und BOMAT wurde die neu normierte und vollständig überarbeitete Version des NEO-FFI (Borkenau & Ostendorf, 2008) eingesetzt. Dieses Persönlichkeitsinventar erfasst mit jeweils 12 Items die fünf Dimensionen Neurotizismus, Extraversion, Offenheit für Erfahrungen, Verträglichkeit und Gewissenhaftigkeit. Die Beantwortung erfolgt auf einer 5-fach abgestuften Likert-Skala, die von *starke Ablehnung* bis *starke Zustimmung* reicht.

3.1.5 Erfolgsindikatoren

Zur Überprüfung der retrospektiven Kriteriumsvalidität wurden die *Abiturgesamtnote* sowie die *Bachelor- bzw. Vordiplomgesamtnote* als Erfolgsindikatoren herangezogen.

3.1.6 Akzeptanzmessung

Zur Akzeptanzüberprüfung sollten die Probanden nach dem Bearbeiten des BBM und BOMAT ein Item auf einer 5-stufigen Skala (1 = sehr berufsnah bis 5 = wenig berufsnah) zur beruflichen Augenscheinvalidität nach Smither et al. (1993) einschätzen. Ebenso wurde der multidimensionale Fragebogen AKZEP!-L (Kersting, 2008) eingesetzt, der die vier Aspekte Augenscheinvalidität, Kontrollierbarkeit, Messqualität und Belastungsfreiheit mit jeweils vier Items auf einer 6-fach abgestuften Likert-Skala (*trifft nicht zu bis trifft genau zu*) erfasst.

3.1.7 Sozio-demografische Variablen

Es wurden sozio-demografische Variablen (Alter, Geschlecht, Deutsch als Muttersprache, Universität, Studium) erhoben, die zur Stichprobenbeschreibung sowie im Rahmen der Akzeptanzprüfung als Kontrollvariablen verwendet wurden.

3.2 Untersuchungsvorgehen

Die Studie wurde als Onlinebefragung konzipiert, mit der Unipark Software (<http://www.unipark.info>) erstellt und entsprechend gängiger Empfehlungen für Onlinestudien umgesetzt, indem bspw. rechtliche Hinweise zur Datenspeicherung gegeben wurden (Birnbauer, 2004; Kraut, Olson, Banaji, Bruckman, Cohen & Couper, 2004). Die Teilnahme war freiwillig und bestand aus vier Modulen, die separat angewählt werden konnten. Alle Teilnehmerinnen und Teilnehmer mussten mit dem ersten Modul beginnen, in welchem sozio-demografische Daten und ein Zuordnungscode generiert sowie allgemeine Instruktionen und Informationen zur Testbearbeitung/-teilnahme dargestellt wurden. Zusätzlich musste in diesem Modul der BBM be-

arbeitet sowie dessen Akzeptanz eingeschätzt werden. Das Modul 2 umfasste die drei Subtests des IST-2000R. Im dritten Modul wurde der NEO-FFI erhoben, im vierten Modul wurde der BOMAT durchgeführt und dessen Akzeptanz von den Probanden bewertet. Um einen gemeinsamen beruflichen Kontext für die Bewertung der Aufgaben herzustellen, erhielten die Teilnehmer vor der Bearbeitung der zwei Matrizenaufgaben (BBM und BOMAT) die Instruktion, sich vorzustellen, in einem Unternehmen aus der Automobil-Branche als Fachreferent ihres jeweiligen Studienganges beschäftigt zu sein. Die Teilnehmer konnten nach dem ersten Modul frei bestimmen, welche und in welcher Reihenfolge sie die restlichen Module bearbeiten wollten. Auf eine Randomisierung der Module 2 bis 4 wurde verzichtet, da die Bearbeitungszeiten der Module stark differierten und den Probanden durch die individuelle Modulwahl eine flexible Durchführung ermöglicht wurde. Als Anreiz zur Teilnahme erhielten die Befragten eine individuelle Ergebnismeldung sowie die Möglichkeit, an einer Verlosung von 20 Gutscheinen teilzunehmen. Psychologiestudierenden wurden entsprechende Versuchspersonenstunden angerechnet.

3.3 Stichprobe

Die Stichprobe setzt sich aus 360 Studierenden verschiedener Studiengänge und Universitäten zusammen. Es nahmen insgesamt 176 Psychologiestudierende und 184 Studierende anderer Studiengänge (Sozial- und Erziehungswissenschaften ($n = 80$), Informationsmanagement ($n = 29$), Rechts- und Wirtschaftswissenschaften ($n = 33$), weitere Studiengänge ($n = 42$)) an der Studie teil. Im Durchschnitt waren die Teilnehmerinnen und Teilnehmer 24.78 Jahre alt ($SD = 6.50$). In der Stichprobe waren zu 82% Frauen vertreten. Alle 360 Probanden bearbeiteten das erste Modul vollständig mit einer durchschnittlichen Bearbeitungszeit von 53 Minuten. Zur Bestimmung der Retest-Reliabilität des BBM bearbeiteten 20 Personen den Test nach ca. drei Monaten ein zweites Mal. Modul 2 wurde von 233 Personen ($M = 48$ min) und Modul 3 von 229 Personen ($M = 16$ min) absolviert. Das vierte Modul hatte die höchste Abbrecherquote und wurde noch von 75 Probanden ($M = 83$ min) durchgeführt.

4 Ergebnisse

Da die Datenerhebung computergestützt stattfand, wurde die Auswertungsobjektivität als gegeben angesehen und auf eine Quantifizierung verzichtet (Shrout & Fleiss, 1979; Wirtz & Caspar, 2002).

4.1 Itemkennwerte und Reliabilität

Der BBM erwies sich in der vorliegenden Studie als hinreichend schwierig, trennscharf und reliabel. Die Schwierigkeitsindizes variierten von .09 bis .88 ($M = 0.44$, $SD = 0.42$) und deckten somit unterschiedliche Schwierigkeitsbereiche ab. Die Trennschärfen lagen nach part-whole Korrektur zwischen .05 und .52 ($M = 0.26$, $SD = 0.17$). Drei Items erwiesen sich mit Trennschärfen $< .10$ als wenig zwischen den Probanden differenzierend. Der BBM bestand aus 22 Items und erzielte für die Gesamtstichprobe eine zufriedenstellende interne Konsistenz mit ei-

nem Cronbachs $\alpha = .71$. Die Retestreliabilität betrug in einer Teilstichprobe ($n = 20$) $r_{tt} = .79$.

Die Itemkennwerte und Reliabilität des BOMAT bewegten sich in einem ähnlichen Bereich. So streuten die Itemschwierigkeiten von .23 bis .85 ($M = 0.60$, $SD = 0.44$) bzw. die korrigierten Trennschärfen von .09 bis .64 ($M = 0.28$, $SD = 0.19$) bei einer internen Konsistenz von Cronbachs $\alpha = .75$ bei 28 Items.

4.2 Kriteriumsvalidität

Der BBM korrelierte mit dem retrospektiven Kriterium *Abiturgesamtnote* moderat ($r = -.43$, $p < .01$, $n = 232$). Für eine Teilstichprobe ($n = 93$) lagen auch Daten zu dem Kriterium *Bachelor- bzw. Vordiplomgesamtnote* vor. Die Korrelation mit dem BBM lag hierbei etwas niedriger als für die *Abiturgesamtnote*. Der Zusammenhang erwies sich mit $r = -.39$ ($p < .01$) als statistisch signifikant. Die Korrelationen zwischen BOMAT und *Abiturgesamtnote* ($r = -.45$, $p < .01$, $n = 66$) sowie *Bachelor- bzw. Vordiplomgesamtnote* ($r = -.41$, $p < .05$, $n = 28$) waren in Richtung und Höhe vergleichbar.

4.3 Konstruktvalidität

Zur Prüfung der konvergenten und diskriminanten Validität wurde der BBM mit unterschiedlichen Konstrukten korrelativ in Beziehung gesetzt (vgl. Tabelle 1). Die Korrelation zwischen dem BBM und seinem Pendant BOMAT von $r = .75$ ($p < .01$) deutet darauf hin, dass beide Testverfahren substantiell dasselbe Intelligenzkonstrukt erfassen. Ebenso weisen die statistisch signifikanten Korrelationen mit den Subtests *Matrizen* ($r = .51$, $p < .01$), *Analogie* ($r = .41$, $p < .01$) und *Zahlenreihe* ($r = .44$, $p < .01$) des IST 2000-R auf eine konvergente Validität des BBM hin. Hinweise auf Divergenz leisten die geringen Korrelationen des BBM mit den Big Five. Lediglich Offenheit ($r = .20$, $p < .05$) und Gewissenhaftigkeit ($r = .31$, $p < .01$) korrelieren mit dem BBM statistisch bedeutsam. Auch der BOMAT korreliert moderat mit den Subtests des IST-2000R ($r = .36$ bis $r = .41$; jeweils $p < .01$) und lediglich mit einer Dimension aus dem NEO-FFI signifikant (Offenheit: $r = .31$, $p < .05$).

4.4 Soziale Akzeptanz

64 Probanden, die sowohl den BBM als auch den BOMAT bearbeitet und deren Akzeptanz mittels des single-items nach Smither et al. (1993) eingeschätzt haben, bewerteten den BBM mit $M_{BBM} = 2.66$ ($SD = 1.35$) als berufsnäher im Vergleich zum BOMAT ($M_{BOMAT} = 3.33$, $SD = 0.94$; $t(63) = -4.45$, $p < .01$). Einen detaillierteren Einblick in die Akzeptanzbewertung ermöglichten die Ergebnisse des AKZEPT!-L, deren Mittelwerte, Standardabweichungen und Reliabilitäten sowie Korrelationen mit den Testergebnissen in Tabelle 2 aufgeführt sind. Damit der BBM als akzeptierter gelten kann, sollte er im Vergleich zum BOMAT signifikant höher hinsichtlich der Augenscheinvalidität und vergleichbar hinsichtlich der Kontrollierbarkeit, Messqualität und Belastungsfreiheit bewertet werden.

Tabelle 1: Mittelwerte, Standardabweichungen und Korrelationen für BBM, BOMAT und den zur Validierung eingesetzten Instrumenten & Variablen

	BBM			BOMAT	
	M (SD)	r(BBM)	N	r(BOMAT)	N
BBM	11.13 (3.40)	1	360	.75**	75
BOMAT	14.38 (5.38)	.75**	66	1	75
Abitur-gesamtnote	2.3 (0.57)	-.43**	232	-.45**	67
Bachelor-gesamtnote	1.86 (0.71)	-.39**	94	-.41**	28
IST-2000R					
Matrize	14.70 (2.86)	.51**	233	.36**	66
Analogie	11.53 (3.05)	.41**	233	.38**	66
Zahlenreihe	17.18 (3.32)	.44**	233	.41**	66
NEO-FFI					
Neuro-tizismus	1.87 (0.67)	-.05	229	-.01	66
Extraversion	2.33 (0.49)	-.04	229	.04	66
Offenheit	2.53 (0.40)	.20*	229	.31*	66
Verträglich-keit	2.68 (0.47)	-.05	229	.07	66
Gewissen-haftigkeit	3.02 (0.55)	.31**	229	.03	66

Anmerkung. * $p \leq .05$ (zweiseitig), ** $p \leq .01$ (zweiseitig)

Hinsichtlich der Dimension Augenscheinvalidität des AKZEPT!-L konnten signifikante Unterschiede zugunsten des BBM ($M_{BBM} = 3.68, SD = 0.75; M_{BOMAT} = 2.54, SD = 0.84; t(49) = -8.83, p < .01$) gefunden werden. Für beide Verfahren ließen sich darüber hinaus signifikante Korrelationen zwischen dem Testergebnis und der eingeschätzten Augenscheinvalidität des AKZEPT!-L finden ($r_{BBM} = .58, p < .01; r_{BOMAT} = .40, p < .01$). BBM und BOMAT unterschieden sich nicht statistisch bedeutsam in Hinsicht auf die Dimensionen Kontrollierbarkeit ($M_{BBM} = 5.00, SD = 1.03; M_{BOMAT} = 4.91, SD = 1.11; t(49) = -.87, p = .39$), Messqualität ($M_{BBM} = 3.45, SD = 0.85; M_{BOMAT} = 3.39, SD = 0.81; t(49) = -.39, p = .69$) und Belastungsfreiheit ($M_{BBM} = 3.44, SD = 1.12; M_{BOMAT} = 3.41, SD = 1.23; t(49) = -.29, p = .77$). Die Reliabilitäten der Skalen des AKZEPT!-L bewegten sich in einem mittleren Bereich ($.65 \leq \alpha \leq .87$), was auf die geringe und somit ökonomische Itemanzahl (4 Items pro Skala) zurückzuführen sein kann.

Zur Überprüfung der Testfairness wurden die Testwerte sowohl für den BBM als auch für den BOMAT hinsichtlich der Kontrollvariablen Geschlecht, Alter und Muttersprache überprüft. Es ergaben sich weder für das Geschlecht (BOMAT: $t(48) = 0.31, p < .65$; BBM: $t(48) = 1.34, p < .46$)

noch für die Muttersprache (BOMAT: $t(48) = 0.60, p < .55$; BBM: $t(48) = -.38, p < .79$) signifikante Mittelwertsunterschiede. Auch zwischen dem Alter und den beiden Testwerten wurden keine bedeutsamen Zusammenhänge ($r_{BOMAT} = -.05; p = n.s.; r_{BBM} = .01; p = n.s.$) gefunden.

5 Diskussion

In der vorliegenden Studie wurde der BOMAT, ein nachweislich valider (Hossiep et al., 2004), aber eher abstrakter Matrizentest mit dem berufsbezogen gestalteten BBM unter besonderer Berücksichtigung des Aspekts der Akzeptanz kontrastiert. Ziel war es zu überprüfen, ob die Akzeptanz eines Matrizentests allein durch eine kontext- bzw. berufsbezogene Symbolgestaltung erhöht werden kann. Darüber hinaus sollte geprüft werden, ob eine akzeptanzfördernde Gestaltung nicht zulasten der psychometrischen Güte geht. Nur wenn ein berufsbezogen gestalteter Intelligenztest auch reliabel und valide misst, erscheint es lohnenswert, über seinen Einsatz im Vergleich zu abstrakt gestalteten Tests nachzudenken. „Es kommt 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).

Sowohl der BOMAT als auch der BBM erwiesen sich in der vorliegenden Stichprobe als hinreichend reliabel sowie trennscharf und deckten unterschiedliche Schwierigkeitsbereiche ab. Schwierigkeits- und Trennschärfeindizes des BBM entsprachen annähernd denen des BOMAT (Hossiep et al., 2001). Trotz geringerer Itemanzahl des BBM bewegte sich die Reliabilität beider Matrizentests in einem vergleichbaren Bereich (vgl. Bortz, 2005; Osburn, 2000), wobei die niedrigere Anzahl an Items des BBM zusätzlich dem Nebengütekriterium der Ökonomie entgegen kommt. Die Befunde zur Kriteriumsvalidität für den BBM weisen auf eine akzeptable Güte sowohl für das Kriterium *Abitur-gesamtnote* als auch für die *Bachelor- bzw. Vordiplom-gesamtnote* hin, die den retrospektiven Validitäten des BOMAT entsprachen. Die Koeffizienten beider Verfahren bewegten sich in einem ähnlichen Bereich wie die Ergebnisse von Hülshager et al. (2006) und unterstreichen die Güte von Intelligenztests zur Vorhersage von Ausbildungserfolg in Deutschland.

Tabelle 2: Akzeptanzeinschätzung für BBM und BOMAT mittels AKZEPT!-L (nach Kersting, 2008), n = 50

Akzeptanzdimensionen	BBM			BOMAT			BBM vs. BOMAT t-Wert
	M	SD	α	M	SD	α	
Kontrollierbarkeit	5.00	1.03	.74	4.91	1.11	.77	$t(49) = -0.87$
Messqualität	3.45	0.85	.67	3.39	0.81	.68	$t(49) = -0.39$
Augenscheinvalidität	3.68	0.75	.65	2.54	0.84	.77	$t(49) = -8.83**$
Belastungsfreiheit	3.44	1.12	.84	3.41	1.23	.87	$t(49) = -0.29$

Anmerkung. * $p \leq .05$ (zweiseitig), ** $p \leq .01$ (zweiseitig), t-Test für unabhängige Stichproben.

Die Korrelationen des BOMAT und des BBM mit den Skalen des NEO-FFI (Borkenau & Ostendorf, 2008) weisen auf eine divergente Validität hin, was einen weiteren Hinweis für Konstruktvalidität darstellt. Die signifikanten Korrelationen mit den Persönlichkeitsskalen *Offenheit für neue Erfahrung* und *Gewissenhaftigkeit* lassen sich dahin gehend interpretieren, dass Personen mit höheren Ausprägungen in diesen Dimensionen die Testaufgaben mit mehr Sorgfalt, Aufrichtigkeit sowie Interesse bearbeitet haben und dementsprechend höhere Leistungen erbrachten. Die Ergebnisse stehen mit bisherigen Befunden im Einklang, die positive Zusammenhänge einerseits zwischen der Testmotivation und den NEO-FFI Skalen *Offenheit für neue Erfahrung* und *Gewissenhaftigkeit* (Hunthausen et al., 2003) sowie andererseits zwischen Testmotivation und Intelligenz (Chan et al., 1997) unterstreichen. Bis hierhin kann für den BBM attestiert werden, dass er hinsichtlich der psychometrischen Güte als reliabel und valide einzustufen ist, also *gut misst*.

Für den BBM konnte eine aus Probandensicht hohe berufliche Augenscheinvalidität bestätigt werden. Die Testgestaltung beeinflusste in dieser Stichprobe die Akzeptanzeinschätzung zugunsten des BBM mit berufsbezogenen Symbolen im Vergleich zum BOMAT. Aus Probandensicht scheint die berufsbezogene Testgestaltung (BBM) die Augenscheinvalidität zu erhöhen und gleichzeitig wird der Test wie der BOMAT als kontrollierbar, belastungsfrei und messgenau empfunden. Der BBM *misst also gut* und *ist gut* bzw. sogar *besser* akzeptiert hinsichtlich der beruflichen Augenscheinvalidität.

Um den Gültigkeitsanspruch der Befunde zu erweitern, sollte eine Validierung von BOMAT und BBM in einem randomisierten Design/Setting an einer berufstätigen Bewerberstichprobe realisiert werden. Dieses Vorhaben erscheint unter zweierlei Aspekten lohnenswert. Einerseits würde die Akzeptanzeinschätzung aus der Perspektive von Bewerbern unter Realbedingungen die Aussagekraft der Ergebnisse erhöhen, da bislang lediglich Studierende untersucht worden sind und die Generalisierbarkeit von Akzeptanzuntersuchungen eignungsdiagnostischer Instrumente, in der ausschließlich oder überwiegend Studierende befragt wurden, angezweifelt werden kann (vgl. Kersting, 1998). Andererseits fehlt in dieser Studie die Validierung an einem prospektiven beruflichen Erfolgskriterium. Denn gerade die prognostische Validität von Intelligenztests bei relativ kurzer Durchführungszeit macht dieses eignungsdiagnostische Verfahren für die Praxis so attraktiv. Es ist aber bereits eine Folgestudie in Kooperation mit einem deutschen Industrieunternehmen geplant, in welcher der BBM explizit mit den firmenspezifischen Symbolen des kooperierenden Unternehmens getestet werden soll. Ein weiterer Aspekt, den es in Folgeuntersuchungen zu optimieren gilt, liegt in der Versuchsplanung. Die online-basierte Testung mit dem hinsichtlich Modul 2-4 frei wählbaren Umfang ermöglichte zwar eine ökonomische und probandenfreundliche Erhebung, jedoch variierten die Stichprobenumfänge der einzelnen Module stark. Dieses kann ein Hinweis dafür sein, dass einer Vielzahl der Probanden eine individuelle Ergebnisrückmeldung über die Module 1 bis 3 (BBM, IST-2000R, NEO-FFI) ausreichte und eine weitere Testung ihrer kognitiven Leistungsfähigkeit (BOMAT) in Modul 4 redundant erschien. Die Probanden,

die alle Module vollständig bearbeitet haben, stellen möglicherweise eine Positivselektion hinsichtlich bestimmter Eigenschaften wie überdurchschnittliche Testmotivation oder Leistungsfähigkeit dar. Da studentische Stichproben jedoch grundsätzlich im oberen Fähigkeitsbereich anzusiedeln sind, sollten mögliche auf die Leistungsfähigkeit zurückzuführende Verzerrungen gering ausgefallen sein.

Hinsichtlich der Akzeptanzmessung bleibt festzuhalten, dass durch das mehrdimensionale Instrument AKZEPT!-L (Kersting, 2008) eine differenzierte Bewertung der Akzeptanz im Vergleich zum verwendeten single-item zur berufsbezogenen Augenscheinvalidität nach Smither et al. (1993) möglich wurde. Es scheint eine lohnenswerte Aufgabe der Wissenschaft zu sein, sich über bestehende (wissenschaftlich fundierte) Inventare zur Akzeptanzmessung zu verständigen, so dass zukünftig eine einheitlichere und vergleichbare Akzeptanzforschung stattfinden kann.

Bis hierhin kann konstatiert werden, dass der BBM ein zeitökonomisches, psychometrisch valides und sozial akzeptiertes Testverfahren darstellt. Anliegen dieser Studie war es zu untersuchen, ob psychologische Testverfahren durch einfache Modifikationen an Augenscheinvalidität gewinnen und so möglicherweise vermehrt Zugang in die Unternehmenspraxis erfahren können. Wenn wissenschaftlich fundierte Instrumente wie der BOMAT aus Akzeptanzgründen nicht den Weg in die Praxis finden, stellt die Konstruktion von berufsbezogenen Tests mit kontext- oder branchenspezifischen Symbolen einen lohnenswerten Lösungsweg dar. Eine akzeptanzförderliche Testgestaltung macht jedoch keinesfalls eine theoriegeleitete sowie methodisch bzw. testtheoretisch fundierte Testkonstruktion und abschließende Evaluation entbehrlich. Insbesondere für Intelligenztests anderer Formate (z.B.: IST-2000R) sollte ein eindeutig erkennbarer Bezug zu den Arbeitsanforderungen eines Arbeitsplatzes zur Akzeptanzerhöhung - wie von der DIN 33430 gefordert (Deutsches Institut für Normung e. V., 2002) - beitragen. Ein Beispiel für eine Kombination aus Arbeitsprobe und Testverfahren zur Erfassung der berufsbezogenen Intelligenz für büro- und kaufmännische Tätigkeiten stellt der AZUBI-BK (Görllich & Schuler, 2010) dar, der empirisch nachweislich *gut misst* und aus Sicht der Bewerber auch *gut ist*. Damit schlussendlich Intelligenztests auch in deutschen Unternehmen vermehrt Einzug in die Eignungsdiagnostik finden, sollte neben einem stärkeren Austausch zwischen Praxis und Wissenschaft auch die Konstruktion von validen, berufs- und anforderungsbezogenen Intelligenztest Aufgabe zukünftiger Forschung sein.

6 Danksagung

Wir bedanken uns herzlichst bei den Testautoren des BOMAT, Rüdiger Hossiep und Michele Leidorf (ehemals Hasella), sowie dem Cheflektor des Hogrefe Verlags, Jörg Hampe, für die freundliche und unkomplizierte Unterstützung unseres Forschungsvorhabens.

7 Literaturverzeichnis

- Ackerman, P. L. (1994). Intelligence, attention, and learning: Maximal and typical performance. In D. K. Detterman (Ed.), *Theories of Intelligence* (Current Topics in Human Intelligence, Vol. 4, pp. 1-27). Norwood: Ablex Publishing Corporation.
- 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. (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. & 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.
- 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.
- Birnbaum, M. H. (2004). Human Research and Data Collection via the Internet. *Annual Review of Psychology*, 55, 803-832.
- Borkeu, P. & Ostendorf, F. (Hrsg.) (2008). *NEO-Fünf-Faktoren Inventar nach Costa und McCrae (NEO-FFI)* (2. Aufl.). 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. (2005). *Statistik für Human- und Sozialwissenschaftler* (6. Aufl.). Wien: Springer.
- 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.
- Deutsches Institut für Normung e. V. (Hrsg.) (2002). *DIN 33430: Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen*. Berlin: Beuth.
- Eleftheriou, A. & Robertson, I. (1999). A Survey of Management Selection Practices in Greece. *International Journal of Selection and Assessment*, 7, 203-208.
- Gilliland, S. W. (1993). The Perceived Fairness of Selection Systems: An Organizational Justice Perspective. *The Academy of Management Review*, 18, 694-734.
- Göhs, N. & Dick, J. (2001). Testverfahren bei der Personalauswahl - Qualitätssuche im intransparenten Markt. *Personal - Zeitschrift für Human Resource Management*, 53, 46-48.
- Görlich, Y. & Schuler H. (2010). *Arbeitsprobe zur berufsbezogenen Intelligenz - Technische und handwerkliche Tätigkeiten (AZUBI-TH)* (2. Auflage). 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. (2003). *Kognitive Leistungsfähigkeit in der Berufsdiagnostik: Forschungsüberblick und Entwicklung eines neuen Instruments zur Erfassung Schlussfolgernden und Kreativen Denkens*. Berlin: dissertation.de.
- Hell, B., Boramir, I., Schaar, H. & Schuler, H. (2006). Interne Personalauswahl und Personalentwicklung in deutschen Unternehmen. *Wirtschaftspsychologie*, 8, 2-22.
- Hossiep, R., Turck, D. & Hasella, M. (Hrsg.) (2001). *Bochumer Matrizentest - advanced - short version (BOMAT - A)*. Göttingen: Hogrefe.
- Hossiep, R., Turck, D. & Hasella, M. (2004). Bochumer Matrizentest - advanced - short version (BOMAT-A). In W. Sarges & H. Wottawa (Hrsg.). *Handbuch wirtschaftspsychologischer Testverfahren. Band I: Personalpsychologische Instrumente* (2. Aufl., S. 167-172). Lengerich: Pabst Science Publishers.
- Hülsheger, U. R. & Anderson, N. (2009). Applicant perspectives in selection: Going beyond preference reactions. *International Journal of Selection and Assessment*, 17, 335-345.
- Hülsheger, U. R. & Maier, G. W. (2008). Persönlichkeitseigenschaften, Intelligenz und Erfolg im Beruf: Eine Bestandsaufnahme internationaler und nationaler Forschung. *Psychologische Rundschau*, 59, 108-122.
- Hülsheger, 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.
- 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.
- Hunter, J. E. & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 340-362.
- Kersting, M. (1998). Differentielle Aspekte der sozialen Akzeptanz von Intelligenztests und Problemlöseszenarien als Personalauswahlverfahren. *Zeitschrift für Arbeits- und Organisationspsychologie*, 42, 61-75.
- Kersting, M. (2006). Zur Beurteilung der Qualität von Tests: Resümee und Neubeginn. *Psychologische Rundschau*, 57, 243-253.
- Kersting, M. (2008). Zur Akzeptanz von Intelligenz- und Leistungstests. *Report Psychologie*, 33, 420-433.
- König, C. J., Klehe, U.-C., Berchtold, M. & Kleinmann, M. (2010). Reasons for being selective when choosing per-

- sonnel 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. *Psychologische Rundschau*, 60, 82-98.
- Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J. & Couper, M. (2004). Psychological Research Online: Report of Board of Scientific Affairs' Advisory Group on the Conduct of Research on the Internet. *American Psychologist*, 59, 105-117.
- 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.
- Liepmann, D., Beauducel, A., Brocke, B. & Amthauer, R. (Hrsg.) (2007). *Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R)* (2. Aufl.). Göttingen: Hogrefe.
- 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.
- Osburn, H. G. (2000). Coefficient Alpha and Related Internal Consistency Reliability Coefficients. *Psychological Methods*, 5, 343-355.
- Ostapczuk, M., Musch, J. & Lieberei, W. (2011). Der „Analytische Test“: Validierung eines neuen eignungsdiagnostischen Instruments zur Erfassung von schlussfolgerndem Denken (AT). *Zeitschrift für Arbeits- und Organisationspsychologie*, 55, 1-16.
- Ryan, A. M. & 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 validity in practice. *Personnel Psychology*, 52, 359-392.
- Ryan, A. M. & Ployhart, R. E. (2000). Applicants' Perceptions of Selection Procedures and Decisions: A Critical Review and Agenda for the Future. *Journal of Management*, 26, 565-606.
- Sackmann, S. & Elbe, M. (2000). Tendenzen und Ergebnisse empirischer Personalforschung der 90er Jahre in West-Deutschland. *Zeitschrift für Personalforschung*, 14, 131-157.
- 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. (2003). International validity generalization of GMA and cognitive abilities: A European community meta-analysis. *Personnel Psychology*, 56, 573-605.
- Sarges, W. & Wottawa, H. (Hrsg.). (2004). *Handbuch wirtschaftspsychologischer Testverfahren: Band I: Personalpsychologische Instrumente* (2. Aufl.). Lengerich: Pabst Science Publishers.
- 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.
- 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.
- Schuler, H. (1990). Personalauswahl aus der Sicht der Bewerber: Zum Erleben eignungsdiagnostischer Situationen. *Zeitschrift für Arbeits- und Organisationspsychologie*, 34, 184-191.
- Schuler, H. (2000). *Psychologische Personalauswahl: Einführung in die Eignungsdiagnostik* (3. Aufl.). Göttingen: Verlag für Angewandte Psychologie.
- Schuler, H., Frier, D. & Kauffmann, M. (Hrsg.) (1993). *Personalauswahl im europäischen Vergleich* (Beiträge zur Organisationspsychologie, Bd. 13). Göttingen: Verlag 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.
- Schuler, H. & Stehle, W. (1983). Neue Entwicklungen des Assessment-Center-Ansatzes beurteilt unter dem Aspekt der sozialen Validität. *Zeitschrift für Arbeits- und Organisationspsychologie*, 27, 33-44.
- Shrout, P. E. & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420-428.
- 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. *The American Journal of Psychology*, 15, 201-292.
- 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.
- Wirtz, M. A. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe.

Korrespondenzadresse:

MSc Psych., Nils Benit
 University of Hildesheim
 Department of Psychology
 Marienburger Platz 22
 D - 31141 Hildesheim
 GERMANY
 nils.benit@uni-hildesheim.de