



# Validierung des adaptiven Persönlichkeitsfragebogens shapes für die Personalauswahl

Richard T. Justenhoven, Katharina Lochner & Achim Preuß

Cut-e Group, Hamburg

## ZUSAMMENFASSUNG

Der adaptive, computergestützte Fragebogen shapes wurde für die unbeaufsichtigte Online-Vorauswahl im Rahmen der beruflichen Eignungsdiagnostik entwickelt. Zunächst wird die Konstruktion sowie die adaptive Messmethode von shapes ausführlich dargestellt. In Studie 1 wird die Retest-Reliabilität und damit die Zuverlässigkeit des Instrumentes überprüft ( $N = 517$ ). Die von shapes gemessenen Persönlichkeitsskalen weisen nach Bühner (2006) mehrheitlich gute bis sehr gute Reliabilitätswerte auf. In Studie 2 werden zum Zweck der konvergenten Validierung die Skalen des Fragebogens mit denen des NEO-FFI korreliert ( $N = 68$ ). Die gefundenen Zusammenhänge stützen die Validität des Instrumentes.

*Schlüsselbegriffe: shapes, adalloc, berufliche Eignungsdiagnostik, Retest-Reliabilität, Konvergente Validierung, NEO-FFI*

## 1 Einleitung

Die Messung persönlichkeitsrelevanter Konstrukte stellt seit langem einen wichtigen Faktor in der psychologischen Diagnostik und der beruflichen Personalauswahl dar (z. B. Barrick & Mount, 1991; Camera & Merenda, 2000; Tett, Barrick & Rothstein, 1991) und wird vermutlich noch weiter an Bedeutung gewinnen (Hough & Oswald, 2008; Rothstein & Goffin, 2006). Der Grund hierfür ist, dass sich Persönlichkeit in zahlreichen Studien als valider Prädiktor für Berufserfolg erwiesen hat (Hogan, 1998; Hurtz & Donovan, 2000) und zudem inkrementelle Validität zu kognitiven Fähigkeiten aufweist (Rothstein & Goffin, 2006).

Insgesamt wird der Einsatz von Persönlichkeitsfragebögen in der Personalauswahl häufig positiv diskutiert (z. B. Hell & Schuler, 2005; Hossiep & Mühlhaus, 2005; Jackson & Corr, 1998; Schuler, 2014), jedoch gibt es diesbezüglich auch Probleme. In dem vorliegenden Beitrag werden daher folgende Aspekte diskutiert: (1) wie valide eine Selbstbeschreibung zur Messung von Persönlichkeit ist, sowie, insbesondere im Kontext der Personalauswahl, wie verfälschbar Persönlichkeitsfragebögen sind und welche Auswirkung eine Verfälschung auf die Validität der Instrumente hat; (2) ob breite oder enge Eigenschaften eine bessere Vorhersage ermöglichen; (3) die soziale Akzeptanz, also wie derartige Instrumente von Bewerbern wahrgenommen werden; und (4) welche Besonderheiten sich durch die Nutzung moderner Technologien, insbesondere des Internets, ergeben.

Die beschriebenen Aspekte waren der Ausgangspunkt für die Entwicklung des adaptiven online-Fragebogens shapes, der speziell für den Einsatz im unüberwachten Online-Assessment konstruiert wurde. Dies geschah mit der Zielsetzung, die oben genannten Probleme beim

Einsatz von Persönlichkeitsfragebögen in der Personalauswahl zu lösen. Im Folgenden werden daher die oben genannten Kontroversen umrissen und es wird dargestellt, wie mit shapes und der Messmethode adalloc versucht wird, Lösungen für die genannten Probleme anzubieten. Die anschließenden Studien zielen darauf ab, die Reliabilität (Studie 1) und Validität (Studie 2) des Instrumentes zu untersuchen und zu stützen.

Bevor die angesprochenen Probleme, die ein Einsatz von Persönlichkeitsfragebögen in der Personalauswahl mit sich bringt, thematisiert werden, soll zunächst geklärt werden, welches Konzept die Differenz zwischen Kandidaten (zeit-) stabil erklären kann. Denn der Kern einer erfolgreichen Personalauswahl besteht darin, die Varianz zwischen Bewerbern zu betrachten und darauf basierend zwischen ihnen zu differenzieren, sodass letztlich der passende Kandidat für die zu besetzende Stelle ausgewählt werden kann (z. B. Asendorpf, 2012). Nach dem Eigenschaftsparadigma sorgen überdauernde Dispositionen dafür, dass eine Person in ähnlichen Situationen wiederkehrende Reaktionen zeigt (Asendorpf, 2012). Dispositionen bzw. Eigenschaften sind in ihrer Ausprägung jedoch nicht direkt mess- oder erfassbar. Allerdings kann von gezeigtem Verhalten einer Person auf dahinterliegende Dispositionen geschlossen werden (Eysenck & Keane, 2015). Das Konstrukt der Eigenschaften besitzt vielfältige Definitionen. In diesem Artikel wird der folgenden Definition von Angleitner und Riemann gefolgt: „Eine Eigenschaft ist ein Integrationsmittel, sie ist meist relativ generalisiert und überdauernd und durch die Konsistenz im Verhalten in unterschiedlichen Situationen erkennbar“ (Angleitner & Riemann, 2005, S. 94).

Die Persönlichkeit ist dabei als Summe der Eigenschaften einer Person definiert, wobei die einzelnen Eigenschaften

nicht dichotom, sondern kontinuierlich ausgeprägt sind. Bei der Persönlichkeitsdiagnostik gilt es folglich nicht zu ergründen, ob eine Eigenschaft vorhanden ist, sondern in welchem Maße sie bei der betreffenden Person ausgeprägt ist (Asendorpf, 2012; Eysenck & Keane, 2015).

Das Big Five Modell kann in diesem Zusammenhang als das etablierteste Persönlichkeitsmodell angesehen werden (Salgado, 2003; Warr, Bartram & Brown, 2005). Die Big Five Persönlichkeitsdimensionen werden häufig herangezogen, um berufliche Leistung oder Eignung vorherzusagen (Barrick & Mount, 1991). Eine genaue Beschreibung der Dimensionen (Neurotizismus, Extraversion, Offenheit für Erfahrung, Verträglichkeit sowie Gewissenhaftigkeit) kann Borkenau und Ostendorf (2008) entnommen werden.

## 2 Rahmenbedingungen

Wenn Persönlichkeitsfragebögen in der Personaldiagnostik zum Einsatz kommen, werden vor allem (1) die Verfälschbarkeit von Fragebögen, (2) die relative Breite der Eigenschaften, (3) die soziale Validität sowie (4) die Nutzung neuester Technologien thematisiert (z. B. Barak & English, 2002; Bartram 1999; Salgado, 1997; Schmidt & Hunter, 1998; Schuler, Hell, Trapmann, Schaar & Boramir, 2007; Vecchione, Alessandri & Barbaranelli, 2011).

Im Folgenden wird zunächst die Verfälschbarkeit diskutiert. Autoren berichten von mittelmäßigen bis hohen Übereinstimmungen zwischen Fremd- und Selbstbeurteilungen (Brandstätter, Filipp & Drescher, 1992; Moser et al., 1994; Moser, 1999; Zempel & Moser, 2005). Eine Abweichung zwischen Selbst- und Fremdbeurteilung kann unterschiedlich begründet sein. Hier ist vor allem zwischen Impression Management und Self-Deception zu unterscheiden (Paulhus, 1984). Impression Management ist die gewollte Verzerrung der Selbstbeurteilung, um erwünschter bzw. passender für eine Stelle wahrgenommen zu werden (McFarland, Ryan & Kriska, 2003). Self-Deception hingegen beschreibt die unbewusste Abweichung zwischen Selbst- und Fremdwahrnehmung, die auf einer unterschiedlichen Wahrnehmung der Realität basiert (Gur & Sackeim, 1979). Im Rahmen des Impression Management unterscheiden Mahar, Cologon und Duck (1995) zwischen *faking-good*, *faking-bad* und *faking a specific job-role*. Birkeland, Manson, Kisamore, Brannick und Smith (2006) gehen davon aus, dass Kandidaten Impression Management betreiben, wenn es zur Erreichung eines Zieles sinnvoll erscheint (ebenso Hogan, Barrett & Hogan, 2007). Dies konnte sowohl unter Laborbedingungen (Tett, Freund, Christiansen, Fox & Coaster, 2012) als auch in realen Settings (Griffith, Chemielowski & Yoshita, 2007; Griffith & Converse, 2012) nachgewiesen werden.

Für den Bereich der Personalauswahl liegen jedoch auch gegenteilige Befunde vor, wonach Bewerber kaum zu Verfälschung neigen (z. B. Hogan, Barret & Hogan, 2007). Allerdings ergaben die Metaanalysen von Birkeland et al. (2006) signifikante Unterschiede im Verhalten zwischen Bewerbern und Nicht-Bewerbern. Es zeigte sich, dass vor allem Dimensionen, die vermeintlich relevant für die zu besetzende Stelle waren, von den Bewerbern verzerrt wurden.

Die Verfälschbarkeit in der Personalauswahl ist darüber hinaus ein Thema, welches auch Beachtung von Seiten der Testkonstruktion erfordert (Jackson, Wroblewski & Ashton, 2000). Gerade bei Persönlichkeitsfragebögen mit einfachen Ratingskalen ist es möglich, das Ergebnis gezielt zu verfälschen (Griffith, Chemielowski & Yoshita, 2007). Ein in der Praxis häufig genutztes Vorgehen, um Verfälschungstendenzen zu begegnen, ist das Forced-Choice Format (Heggestad, Morrison, Reeve & McCloy, 2006). Bühner (2006) wie auch Bortz und Döring (2006) bezeichnen Forced-Choice-Formate als den effektivsten Weg Antworttendenzen und Verfälschungen entgegenzuwirken. Dabei muss der Teilnehmer den Grad seiner Zustimmung nicht zu jedem Item einzeln angeben, sondern sich entscheiden, welchem der zwei (oder mehr) präsentierten Items er eher zustimmt. Der Teilnehmer wird somit gezwungen, eine intra-individuelle Unterscheidung der Items vorzunehmen (Bartram, 2007; Hicks, 1970). Wichtig ist hierbei, dass die gleichzeitig präsentierten und abzuwägenden Items auf unterschiedliche Eigenschaften laden müssen (Baron, 1996). Die Ergebniswerte der einzelnen Eigenschaften sind daher in Bezug auf die anderen Werte der gleichen Person zu interpretieren (Hicks, 1970), es erfolgt also ein *intra-* und kein *interindividueller* Vergleich, d.h. ein Vergleich *innerhalb* einer Person und nicht *zwischen* verschiedenen Personen. Die so erzeugten Daten werden ipsative Daten genannt (Hicks, 1970).

Darüber hinaus stellt sich als zweiter Aspekt die Frage, was genau ein Persönlichkeitsfragebogen erfassen sollte. Eigenschaften können unterschiedlich breit definiert und erfasst werden (Amelang & Bartussek, 2001). Je mehr Persönlichkeitsmerkmale und Verhaltensweisen in einer Eigenschaft aggregiert werden, desto breiter wird diese (Visser & Du Toit, 2004). Das Big Five Modell bildet sehr breite Eigenschaften ab, die moderate bis gute Vorhersagen auf den Berufserfolg erlauben (Dudley, Orvis, Lebiecki & Cortina, 2006; Hurtz & Donovan, 2000). Dem stehen Befunde gegenüber die zeigen, dass enggefasste Eigenschaften Berufserfolg besser vorhersagen können bzw. inkrementelle Validität haben (z. B. Ashton, 1996; Paunonen & Ashton, 2001; de Vries, de Vries & Born, 2010).

Enggefasste Eigenschaften ermöglichen eine stärkere Differenzierung von Verhaltensweisen (Fehr, 2006). So konnten Sitser, van der Linden und Born (2013) zeigen, dass einzelne Facetten von Gewissenhaftigkeit bessere prädiktive Werte hinsichtlich konkreten Verhaltens erzielen als der Faktor selbst. Wenn ein Fragebogen lediglich engere, berufsbezogene Eigenschaften erfasst, ist darüber hinaus zu erwarten, dass sich auch die Augenscheinvalidität erhöht, da die entsprechenden Items ebenfalls einen stärkeren Bezug zu beruflichen Tätigkeiten aufweisen (Kersting, 2008; Lochner, Preuß & Lohff, 2015).

Der dritte Aspekt, der im Zusammenhang mit dem Einsatz von Persönlichkeitsfragebögen in der Personalauswahl diskutiert wird, ist die soziale Validität (Kersting, 2008). Wird ein Instrument von Bewerbern nicht positiv wahrgenommen, sinkt deren Motivation zur ernsthaften Bearbeitung und somit auch die prognostische Validität des Instruments. Zudem kann mangelnde Motivation zu einem Abbruch der Bearbeitung und folglich zu einem

Ausscheiden aus dem Bewerbungsprozess führen. In diesem Zusammenhang konnte Preuß (2006) in einer Studie im Bereich der Traineeauswahlaufzeigen, dass (1) die Drop-out Quote bei der Bearbeitung eines Fragebogens am Computer nach 15 Minuten deutlich ansteigt und (2) vor allem qualifizierte Bewerber das Verfahren abbrechen. Dies ist gerade in Zeiten des „War for Talent“ kein wünschenswerter Effekt.

Schließlich stellt auch die zunehmende Globalisierung neue Anforderungen an die psychologische Diagnostik, denen mit technischen Innovationen begegnet werden kann (Wiechmann & Ryan, 2003). Auf der einen Seite werden Bewerbungsprozesse internationaler und müssen dennoch ökonomisch und schnell durchführbar sein. Dieser Trend ist bereits beobachtbar und wird sich in Zukunft noch verstärken (z. B. Barak & English, 2002; Hertel, Konradt & Orlikowski, 2003). Insbesondere wenn Bewerber aus unterschiedlichen Teilen der Welt kommen und eine Vorauswahl innerhalb kurzer Zeit getroffen werden muss, bieten sich unbeaufsichtigte Online-Verfahren an (Kühlmann & Stahl, 2001; Lochner, Preuß & Lohff, 2015). Auf der anderen Seite ermöglichen die neuen technischen Gegebenheiten die tatsächliche Umsetzung von adaptivem Testen. Dies wiederum hilft die Attraktivität von Diagnostikinstrumenten bei Kandidaten zu steigern, da die Bearbeitungszeit durch adaptives Testen reduziert werden kann (Preuß, 2006). Die Nutzung neuester Technologien kann demnach auch als eine Voraussetzung hoher sozialer Validität angesehen werden (Roth, Wegge & Schmidt, 2007).

### 3 Instrument

Der Persönlichkeitsfragebogen *shapes* wurde daraufhin optimiert, dass er den genannten vier Kriterien Rechnung trägt. So waren die grundlegenden Konzeptionsgedanken, dass (1) er ein Forced-Choice-Format verwendet, um Verfälschungstendenzen zu minimieren, (2) er enge, berufsbezogene Eigenschaften erfasst, (3) er kurze Bearbeitungszeiten aufweist, um zum einen die Akzeptanz zu erhöhen und zum anderen den Drop-out zu minimieren und zuletzt (4) eine unüberwachte, computergestützte Administration möglich ist.

Zum Zeitpunkt der Fragebogenkonstruktion (2000-2003) existierte kein Persönlichkeitsfragebogen, der die bis dahin aus der Forschung entwickelten Vorteile des computerbasierten, adaptiven Testens (CAT) nutzte (siehe hierzu beispielhaft Bartram, 1996; Baron, 1996; Fan, 1998; Hicks, 1970; Rost & Hoberg, 1997; Sands, Waters & McBride, 1997; Saville & Willson, 1991; Stocking & Lewis, 2000; Tonidandel & Quiñones, 2000; Wainer, 2000; Yan, Lewis & Stocking, 1998).

Adaptive Verfahren sind nicht zwingend an die Unterstützung von Computern gebunden (Frey, 2007; Sands, Waters & McBride, 1997). Jedoch sind sie ohne technische Hilfestellung schwierig durchzuführen und auszuwerten (Zhou, Gierl & Cui, 2007), da nach Beantwortung eines ersten Items die Auswahl der Folge-Items vom Antwortverhalten des Teilnehmers abhängt (Frey, 2007; Wainer, 2000). Während adaptive Algorithmen im Bereich der Leistungsdiagnostik recht verbreitet sind, kommen sie

im Bereich der Persönlichkeitsdiagnostik kaum zum Einsatz (Lohff & Wehrmaker, 2008). Dies ist darauf zurückzuführen, dass bei Zunahme von adaptiven Regeln die klassische Testtheorie an ihre Grenzen stößt und folglich für viele etablierte Fragebögen nicht einsetzbar ist (Yan, Lewis & Stocking, 1998).

Dem zweiten Kriterium folgend ist der Fragebogen *shapes* so konstruiert, dass mit dem zugrundeliegenden Modell nur Eigenschaften abgebildet werden, die signifikanten Einfluss auf den beruflichen Erfolg haben. Es sollten ausschließlich Eigenschaftsdimensionen bzw. Facetten von Persönlichkeitsdimensionen berücksichtigt werden, die Varianz in Bezug auf beruflichen Erfolg aufklären (z. B. Leslie & Van Velsor, 1996; Lombardo & Eichinger, 2000; McCall & Lombardo, 1983), aber auch in der Praxis bereits etabliert und akzeptiert sind. Im Kontext des Fragebogens *shapes* werden diese Eigenschaftsdimensionen als Skalen bezeichnet.

Die Identifizierung der Skalen erfolgte mithilfe der qualitativen Inhaltsanalyse nach Mayring (2000) im Zeitraum von 2001-2003. Zunächst wurden verfügbare Kompetenzmodelle recherchiert. Auf diese Weise wurden 42 verschiedene, in der Praxis zu dieser Zeit genutzte Modelle zusammengetragen (z. B. Lominger CPM; PDI Profiler®; SHL Competency Framework). Zusätzlich führten fünf Interviewer mehr als 300 Interviews mit der Repetory Grid (Fay, Bell & Bannis, 2004) oder der Critical Incident Technik (Gatewood, Feild & Barrick, 2010) sowie über 50 Visionary Interviews (Smith, 2005) durch. Aus den gesammelten Daten, den Definitionen und Beschreibungen der Kompetenzmodelle, den erfolgsfördernden Verhaltensweisen sowie weiteren beruflichen Operationalisierungen wurde eine Liste erstellt, welche über 12.000 Einträge umfasste. Darüber hinaus wurde in den Interviews gezielt nach Eigenschaften gefragt, die für Berufserfolg relevant sind. Gemäß Mayrings (2000) zusammenfassender Inhaltsanalyse wurden die Antworten der Interviewpartner paraphrasiert, generalisiert, durch Selektion und Bündelung reduziert und schließlich in Kategorien zusammengefasst.

In einem nächsten Schritt wurden die im Interview erfassten Eigenschaften sowie Eigenschaftsdefinitionen aus der Literatur, vor allem aber aus etablierten Persönlichkeitsfragebögen zusammengetragen. Genutzt wurden unter anderem folgende Fragebögen: NEO-FFI, 16PF sowie der OPQ (u. a. Ostendorf & Angleitner, 2004; Saville & Holdsworth Ltd., 1993; Schuerger, 2001).

Anhand inhaltlicher und semantischer Übereinstimmung wurden die Eigenschaften der Fragebögen sowie die in den Interviews erwähnten Eigenschaften den Kategorien zugeordnet. Dabei wurden nur jene Eigenschaften, deren Definitionen und Beschreibungen inhaltliche Übereinstimmungen mit den Kategorien und damit mit den zuvor zusammengetragenen Verhaltensweisen aufwiesen, berücksichtigt.

Es konnten 18 Skalen identifiziert werden, welche inhaltlich die ermittelten Kategorien abbilden und deren bipolare Ausprägungen nicht durch die Definitionen der übrigen Eigenschaften erklärbar sind. Diese 18 Skalen, abge-

bildet in Tab. 1, sind in keinem der damals vorliegenden Persönlichkeitsmodelle vollständig enthalten. Sie sind jedoch erforderlich, um die ermittelten Kategorien und folglich die erfolgsfördernden Verhaltensweisen darzustellen.

**Tabelle 1: Definitionen der *shapes* Primärskalen (cut-e GmbH, 2013)**

Niedrige Ausprägung	Skala	Hohe Ausprägung		
Überlässt es gern anderen, Führung und Verantwortung zu übernehmen; konzentriert sich lieber auf die eigenen Aufgaben	leitend	Übernimmt gern und oft Führung und Verantwortung; sorgt laufend dafür, dass das Team sich auf die Ziele konzentriert	Arbeitet lieber mit konkreten, praktischen Aufgaben; empfindet die Diskussion über abstrakte Konzepte oder Theorien als überflüssig	konzeptionell Arbeitet gern mit abstrakten Konzepten, Modellen und Strategien; hat viel Spaß an intellektuellen Herausforderungen; denkt hochgradig vernetzt
Akzeptiert die Überzeugungen anderer immer als gegeben und vertraut allein auf die Macht der Fakten	überzeugend	Verhandelt und verkauft sehr gern; überzeugt andere scheinbar mühelos und gewinnt leicht deren Unterstützung	Entwickelt eher konventionelle Lösungen; bleibt im Rahmen des Üblichen und Erprobten; ist nüchtern und wenig kreativ	einfallsreich Denkt weit über den üblichen Rahmen hinaus; entwickelt viele unkonventionelle Lösungen und kreative Ideen
Bleibt bei offiziellen Anlässen sehr oft im Hintergrund; steht auch im vertrauten Umfeld äußerst ungern im Mittelpunkt der Aufmerksamkeit	souverän	Tritt selbstsicher und souverän auf und fühlt sich bei formellen Anlässen oder Präsentationen stets unbefangen und wohl	Steht Neuerungen abwartend gegenüber; schätzt bewährte Vorgehensweisen; konzentriert sich vor allem auf die Risiken von Innovation und Veränderung	offen für Veränderungen Unterstützt neue Ideen; hat Spaß an Veränderungen; sucht ständig nach Verbesserungsmöglichkeiten und stellt Traditionen gern in Frage
Ist still und zurückhaltend; hält lieber etwas Distanz und hat Mühe, neue Kontakte zu knüpfen	kontaktfreudig	Ist stets lebhaft, gesprächig und gesellig; kann andere unterhalten und pflegt ein sehr großes Netzwerk an Kontakten	Sichert sich gern ab; braucht die Unterstützung anderer für eigene Vorhaben; äußert eigene Meinungen nur sehr vorsichtig und zurückhaltend	unabhängig Handelt eigeninitiativ und selbständig; folgt betont den eigenen Überzeugungen; äußert die eigene Meinung immer sehr offen und freimütig
Arbeitet in der Regel allein und konzentriert an einer Sache; folgt nur dem eigenen Tempo und versäumt es häufiger, Informationen weiterzugeben	verträglich	Ist besonders teamorientiert und rücksichtsvoll; arbeitet sehr gern mit anderen zusammen und ist stets hilfsbereit und angenehm im Umgang	Steckt sich sicher erreichbare Ziele und vermeidet allzu große Herausforderungen; schätzt eine ausgewogene Work-Life Balance und stellt dafür die Karriere hintenan	ehrgeizig Steckt sich höchst ambitionierte Ziele; sucht beständig nach Herausforderungen; will unbedingt vorankommen und erfolgreich sein
Denkt sehr sachlich und rational; hinterfragt selten die Gründe für Verhalten anderer und agiert daher oft unüberlegt oder unbedacht	beobachtend	Achtet sehr auf die Reaktionen anderer; beobachtet und analysiert die Motive und Verhaltensweisen anderer genau	Empfindet Konkurrenz als störend für die Harmonie im Team; kann anderen den Erfolg überlassen, solange die eigenen Ziele erreicht sind	wettbewerrend Will im Vergleich mit anderen unbedingt erfolgreich sein und wird durch Wettbewerb angespornt; ist entschlossen zu gewinnen; spricht gern über eigene Erfolge
Denkt selten langfristig und strategisch, sondern konzentriert sich auf kurzfristig dringende Aufgaben; übersieht die Konsequenzen eigener Handlungen	umsichtig	Berücksichtigt eine breite Spanne von Einflussfaktoren von Plänen; denkt weit voraus und nimmt eine langfristige, umfassende Perspektive ein	Nimmt sich Anforderungen sehr zu Herzen und ist emotional sehr beteiligt; wird unter Druck schnell nervös und hektisch	ausdauernd Zeigt ausgesprochen viel Energie und Entschlossenheit; geht die Arbeit mit Optimismus an; bleibt unter Druck immer konzentriert und leistungsfähig
Sucht nach der optimalen Lösung und nimmt Umwege in Kauf, um alle Risiken auszuschließen; könnte seine Aufgaben sehr viel effektiver abschließen	ergebnisorientiert	Geht äußerst zielgerichtet und pragmatisch an Aufgaben heran; konzentriert sich stets auf das Erreichen von Ergebnissen; löst Probleme sehr schnell und effektiv		
Fühlt sich ohne klare Pläne und Prioritäten wohl; ist flexibel und kann mit Chaos und schnellen Veränderungen gut umgehen	planvoll	Ist ausgesprochen systematisch und gut organisiert; setzt immer klare Prioritäten und plant voraus; schätzt den Zeitbedarf von Aktivitäten realistisch ein		
Möchte Dinge vorantreiben und setzt sich dafür über Regeln und Vorschriften hinweg; stößt vieles an, überlässt die Details aber gern anderen	pflichtbewusst	Bringt Arbeiten stets zu Ende und achtet detailliert auf die Qualität; hält sich streng an Verfahrensweisen, Regeln und Vorschriften		
Beurteilt Probleme und Sachverhalte intuitiv und hat nicht das Bedürfnis, den Dingen auf den Grund zu gehen; sein/ihr Urteil wirkt daher oft undifferenziert	analysierend	Evaluiert Informationen und Zahlenmaterial sehr gründlich; versteht mühelos komplexe Sachverhalte und erkennt schnell den Kern eines Problems		

Damit standen die vom Fragebogen zu erfassenden Konstrukte fest. Der Fragebogen sollte jedoch zusätzlich ein Itemformat verwenden, welches einerseits Verfälschungstendenzen entgegenwirken und andererseits sozial akzeptiert sein sollte. Wie bereits beschrieben, bieten sich Forced-Choice-Formate an, um Verfälschungstendenzen zu minimieren. Mit steigender Anzahl zu erfassender Dimensionen werden Forced-Choice-Fragebögen jedoch sehr lang, was sich wie beschrieben negativ auf die Motivation auswirkt. Daher sollte die neue Messmethode auch eine Verkürzung des Instruments bei gleichzeitig valider Erfassung der Persönlichkeitsmerkmale ermöglichen. Auf Basis dieser Überlegungen resultierte die Messmethode *adalloc* (adaptive allocation of consent, dt. adaptive Zuweisung der Zustimmung).

Die Messmethode *adalloc* sieht vor, dass jeweils drei zu verschiedenen Skalen gehörige Items gleichzeitig präsentiert werden (Triplette). Der Kandidat kann insgesamt sechs Punkte gemäß seine Zustimmung auf diese Items verteilen. Im Gegensatz zu klassisch ipsativen Verfahren ist es zusätzlich möglich, gleich viele Punkte auf mehrere Aussagen zu verteilen; außerdem müssen nicht alle Punkte verteilt werden. In Abb.1 ist beispielhaft eine Triplette dargestellt.

Abb. 1: Beispielhafte Darstellung einer Triplette des Fragebogens *shapes*

Um jede der 18 Skalen exakt einmal abzubilden sind sechs Tripletten notwendig. Diese sechs Tripletten bilden einen Sektor. Jede Skala wird durch acht Items erfragt, sodass der Fragebogen folglich aus acht Sektoren besteht. Abbildung 2 stellt die Zusammensetzung eines Sektors bildlich dar.

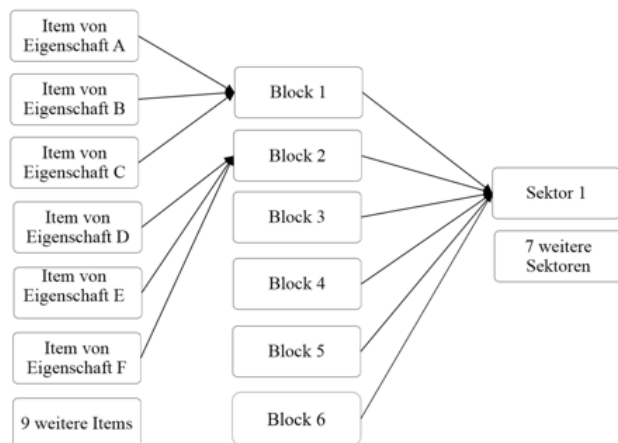


Abb. 2: Zusammensetzung des Fragebogens *shapes* nach der *adalloc*-Methode

Im ersten Sektor werden randomisiert Tripletten erstellt, sodass die Ausgangskonstellation bei jeder Durchführung des Fragebogens eine andere ist. Durch die Randomisierung wird es unmöglich, Musterlösungen zu verbreiten, da jeder Kandidat eine andere Ausgangskonstellation von Tripletten hat. Ab dem zweiten Sektor erfolgt die Zuordnung der Items zu den Tripletten adaptiv. Dazu werden die Skalen nach jedem Sektor gemäß ihrem Wert sortiert. Die Tripletten im jeweils folgenden Sektor werden daraufhin so geformt, dass Items von Skalen gemeinsam präsentiert werden, deren Wert ähnlich hoch ist. Nach dem ersten Sektor berechnet sich dieser Wert aus der Anzahl der Punkte, die der Kandidat auf das für die jeweilige Skala präsentierte Item verteilt hat.

Ab dem zweiten Sektor wird dieser Punktwert zusätzlich mit einem sogenannten Blockgewicht gewichtet: Eine Triplette mit Items, die zuvor hoch bewertet wurden, erhält ein höheres Blockgewicht als eine Triplette mit Items, die zuvor niedrig bewertet wurden. Das hat zur

Folge, dass ein Item mit hohem Blockgewicht, welches mit sechs Punkten bewertet wird, einen größeren Einfluss auf den Skalenwert hat, als ein gleichbewertetes Item mit niedrigem Blockgewicht (Lohff & Wehrmaker, 2008). Dieses Blockgewicht erlaubt eine Verkürzung des Fragebogens, da nicht alle Skalen miteinander verglichen werden müssen, wie es bei einer ipsativen Darbietung ohne Blockgewichte der Fall wäre.

Die Messmethode ist demnach ein Hybrid-Format, welches sowohl normative als auch ipsative Elemente vereint und nach Bartram (1996) und Hicks (1970) dadurch nicht eindeutig einer der beiden Formate zugeordnet werden kann.

#### 4 Studie 1

In dieser Studie wird die Reliabilität des Instrumentes untersucht. Bei den *shapes* Daten handelt es sich, wie bereits erläutert, um hybride Daten. Dies bedeutet, dass der Wert einer Person auf einer Skala abhängig von den Werten auf den anderen Skalen des Instruments ist (Bartram, 1996; Hicks, 1970). Die Skalen sind demnach korreliert und diese Korrelation ist umso höher, je weniger Skalen das Instrument enthält (Baron, 1996). Daher sind bei der Berechnung der Reliabilität einige methodische Besonderheiten zu beachten.

Als Reliabilitätsmaße werden üblicherweise die interne Konsistenz, die Split-half-Reliabilität, die Paralleltest-Reliabilität oder die Retest-Reliabilität verwendet (Bortz & Döring, 2006). In Bezug auf die interne Konsistenz zeigte Tenopyr (1988), dass diese bei ipsativen Daten unter gewissen Umständen überschätzt wird: Besteht ein Instrument aus mehreren Skalen und ist eine davon hochreliabel, so ergeben sich hohe interne Konsistenzen für die anderen Skalen, selbst wenn die zugehörigen Items völlig randomisiert bearbeitet wurden.

Bartram (1996) kommt bei seinem Vergleich von normativen und nachträglich ipsativierten Daten zu dem Schluss, dass die internen Konsistenzen letzterer geringer sind als jene der normativen Daten, aus denen nachträglich die ipsativen Daten erzeugt wurden. Zudem fallen die Split-half-Reliabilitäten ipsativer Daten geringer aus als jene normativer Daten (Baron, 1996). Die Verwendung der internen Konsistenz und der Split-half-Reliabilität sind daher problematisch für ipsative Daten. Dagegen führt die Retest-Reliabilität offenbar nicht zu einer Verzerrung der Reliabilitätsschätzung (Saville & Holdsworth Ltd., 1993; Baron, 1996). Da zudem Messmethodik und Aufbau von *shapes* zum Zeitpunkt der Studie einzigartig waren und es folglich kein paralleles Instrument gab, mit dem die Paralleltest-Reliabilität hätte bestimmt werden können, wurde die Retest-Reliabilität als Reliabilitätsmaß verwendet.

Die Retest-Reliabilität sollte darüber hinaus auch einen ersten Hinweis auf mögliche Verzerrungen durch die (randomisierte) Zusammenstellung der Ausgangstripletten geben. Im Rahmen zweier ausführlicher Studien wurden diese Effekte geprüft (Justenhoven, 2014). In beiden Studien konnte gezeigt werden, dass die Zuweisung der anfänglichen Tripletten keinen signifikanten Einfluss auf

das Ergebnis und somit auf die Reliabilität des Instrumentes hat.

Relevantes Kriterium bei der Berechnung der Retest-Reliabilität ist, dass nur Bearbeitungen berücksichtigt werden können, bei denen sich die gemessene Eigenschaft der Personen nicht geändert hat und bei denen die Bearbeitungsmotivation gleichbleibend ist. Ändert sich dies grundlegend, so muss davon ausgegangen werden, dass die Ergebnisse nicht mehr übereinstimmend sein können (Bühner, 2006). Folglich kann nicht auf doppelte Probebearbeitungen oder Doppelbearbeitungen innerhalb von (Personal-)Entwicklungsprogrammen zurückgegriffen werden. Aus diesem Grund wurden ausschließlich Bearbeitungen im Rahmen von Bewerbungsprozessen in die Analyse aufgenommen.

#### 4.1 Methode

Zugrunde liegt ein Datensatz mit  $N = 517$  Kandidaten, die im Rahmen von Bewerbungsverfahren in Norwegen den Fragebogen zweimal bearbeitet haben. Zwischen den Administrationen liegen mindestens sieben Tage. Das Alter der Teilnehmer liegt zwischen 23 und 66 Jahren ( $M = 40.91$ ,  $SD = 9.25$ ). 289 Personen, (56%) sind männlich und 228 Personen (44%) weiblich.

#### 4.2 Ergebnis

Die Retest-Reliabilitäten der 18 Skalen sind in Tabelle 2 aufgeführt. Die Werte liegen zwischen .55 und .84, wobei 15 der 18 Skalen eine Reliabilität größer gleich .69 aufweisen. Letztere sind als gute bis sehr gute Werte zu interpretieren ist (Bühner, 2006). Lediglich die Skalen *ergebnisorientiert*, *unabhängig* und *ausdauernd* weisen geringere Reliabilitäten auf.

**Tabelle 2: Reliabilitäten der *shapes* Primärskalen**

Skalen	Retest-Reliabilität
leitend	.82
überzeugend	.77
souverän	.79
kontaktfreudig	.81
verträglich	.77
beobachtend	.78
umsichtig	.69
ergebnisorientiert	.55
planvoll	.71
pflichtbewusst	.75
analysierend	.82
konzeptionell	.76
einfallsreich	.82
offen für Veränderungen	.69
unabhängig	.64
ehrgeizig	.69
wettbewerbend	.84
ausdauernd	.66

#### 4.3 Diskussion

Insgesamt sind die Retest-Reliabilitätswerte als sehr zufriedenstellend anzusehen. Somit kann festgehalten werden, dass das Instrument den Großteil der Merkmale zuverlässig misst. Allerdings unterscheiden sich die einzelnen Skalen im Hinblick auf ihre Retest-Reliabilitätswerte. So weist die Skala ergebnisorientiert eine Retest-Reliabilität von lediglich .55 auf. Lohff und Wehrmaker (2008) berichten jedoch von Restest-Werten dieser Skala von .82. Es gilt demnach zu evaluieren, worin die unterschiedlichen Werte begründet sind.

So könnten beispielsweise Retest Studien mit mehreren Messwiederholungen zu unterschiedlichen Zeitpunkten hilfreiche Informationen beispielsweise für die Frage liefern, ob und welchen Einfluss unterschiedliche Retest-Intervalle auf die Reliabilität haben. Gleichwohl kann auf Basis dieser Ergebnisse gefolgert werden, dass die zu messenden Konstrukte bzw. Skalen zuverlässig gemessen werden.

### 5 Studie 2

In dieser Studie werden zum Zwecke der konvergenten Validierung die Skalen des *shapes* Fragebogens mit den Big Five Persönlichkeitsdimensionen korreliert. Die Hypothesen zu den Interkorrelationen zwischen den *shapes* Skalen und den NEO-FFI Faktoren wurden zum einen auf Grundlage der von Lohff und Wehrmaker (2008) berichteten Faktorladungen von *shapes* auf den acht Faktoren berufsbezogenen Verhaltens nach Kurz und Bartram (2002) abgeleitet. Zum anderen leiteten inhaltliche Überlegungen – Ähnlichkeit der Skalendefinitionen und Items – die Hypothesengenerierung. Tabelle 3 zeigt überblicksartig die angenommenen Zusammenhänge. In den Zeilen sind die *shapes*-Skalen abgetragen, in den Spalten die Big Five Faktoren. In den Fällen, in denen ein Zusammenhang vermutet wird, ist in den Zellen jeweils ein Plus für einen vermuteten positiven und ein Minus für einen vermuteten negativen Zusammenhang eingetragen. Zur Begründung sind jeweils stichwortartig die Aspekte der Big Five Faktoren aufgeführt, welche sich mit den jeweiligen *shapes*-Skalen überschneiden und somit den vermuteten Zusammenhang begründen.

Zusammenfassend werden die folgenden Zusammenhänge vermutet: Neurotische Personen sind unsicher. Daher meiden sie alle Situationen, in denen sie im Fokus sind (leitend, überzeugend, souverän) und versuchen sich Sicherheit durch Regelkonformität herzustellen (pflichtbewusst). Extravertierte Personen sind durchsetzungsfähig (leitend, überzeugend), kontaktfreudig und treffen Entscheidungen eher auf Basis von Gefühlen als auf Basis von Fakten (nicht analysierend). Offene Personen stellen bestehende Regeln und Pflichten in Frage (nicht pflichtbewusst), sind offen für Veränderungen und haben neue Ideen. Verträgliche Menschen haben ein Interesse an anderen (beobachtend) und sind eher ein Teil des Teams (verträglich) als dessen Leiter (nicht leitend). Gewissenhafte Menschen schließlich planen sorgfältig (planvoll) und bringen Arbeiten zu Ende (pflichtbewusst). Dabei gehen sie eher konventionell vor

(nicht einfallsreich). Sie sind vorsichtig und verhandeln daher nicht gerne.

**Tabelle 3: Erwartete Zusammenhänge zwischen den *shapes*-Skalen und den NEO-Faktoren**

	N	E	O	A	C
<b>leitend</b>	- Unwohlsein und Unsicherheit in Gruppen	+ Durchsetzungsfähigkeit und Dominanz		- Nachgiebigkeit und Sanftheit	
<b>überzeugend</b>	- Unsicherheit	+ Durchsetzungsfähigkeit und Dominanz			- Vorsicht
<b>souverän</b>	- Ängstlichkeit und Besorgtheit				
<b>kontaktfreudig</b>		+ Lebhaftigkeit und Kontaktfreude			
<b>verträglich</b>				+ Verträglichkeit	
<b>beobachtend</b>				+ Interesse an anderen	
<b>umsichtig</b>					
<b>ergebnisorientiert</b>					
<b>planvoll</b>			- Offenheit und Flexibilität		+ Planung und Priorisierung
<b>pflichtbewusst</b>	+ Sicherheit durch Sich-Halten an Regeln		- Infragestellen von Bestehendem		+ Arbeiten zu Ende bringen
<b>analysierend</b>		- Gefühlsbetontheit			
<b>konzeptionell</b>					
<b>einfallsreich</b>			+ Entwicklung unkonventioneller Ideen		- Befolgen von Regeln
<b>offen f. Veränd.</b>			+ Offenheit		
<b>unabhängig</b>	+ ehrzeigig				
<b>wettbewerwend</b>				- Teamorientierung	
<b>ausdauernd</b>					

Anmerkung: +: positiver Zusammenhang erwartet; -: negativer Zusammenhang erwartet

## 5.1 Methode

Die Stichprobe umfasst  $N = 68$  Teilnehmer, 38 (56%) männliche und 30 (44%) weibliche, die im Jahr 2014 an einem Wettbewerb zur Ermittlung des „Best Student“ teilnahmen. Die Teilnehmer bearbeiteten zunächst online und unbeaufsichtigt den Fragebogen *shapes* und am Finaltag der Veranstaltung, einem Tag, der Elemente eines typischen Assessment Centers beinhaltet, den NEO-FFI (Borkenau & Ostendorf, 2008).

Bei der vorliegenden Studie handelt es sich um eine konvergente und diskriminante Validierung des Instruments *shapes*. Hierfür wird üblicherweise auf eine Multi-Trait-Multit-Method (MTMM) Analyse nach Campbell und Fiske (1959) zurückgegriffen. Sie stellt einen systematischen Vergleich der Eigenschaften dar, die mit verschiedenen Methoden erhoben wurden. In diesem Fall also einen Vergleich der mit *shapes* und der mit NEO-FFI erhobenen Eigenschaften. Konvergente Validität wird angenommen, wenn gleiche Eigenschaften, die mit verschiedenen Instrumenten erhoben wurden, hoch korrelieren. Diskriminante Validität liegt vor, wenn verschiedene Konstrukte, die mit dem gleichen sowie mit unterschiedlichen Instrumenten erhoben wurden, niedrig korrelieren (Campbell & Fiske, 1959). Konkret werden bedeutsame und signifikante Korrelationen zwischen den oben aufgeführten *shapes*-

Skalen und Big Five Faktoren erwartet, sowie niedrige bis keine Korrelationen zwischen den nicht aufgeführten *shapes*-Skalen und Big Five Faktoren.

## 5.2 Ergebnis

Die Korrelationsmatrix ist in Tabelle 4 dargestellt. Für die Skala Neurotizismus zeigt sich erwartungskonform ein signifikanter positiver Zusammenhang mit *pflichtbewusst* ( $r = .30$ ). Signifikant negative Korrelationen der Skala Neurotizismus finden sich nur mit *souverän* ( $r = -.38$ ) und *überzeugend* ( $r = -.24$ ), jedoch nicht mit *leitend*.

Zwischen dem Faktor Extraversion und den Skalen *überzeugend* ( $r = .26$ ), *souverän* ( $r = .36$ ) und *kontaktfreudig* ( $r = .51$ ) sind die Korrelationen erwartungsgemäß signifikant positiv. Ebenfalls den Hypothesen entsprechend zeigt sich ein signifikant negativer Zusammenhang zwischen Extraversion und *analysierend* ( $r = -.43$ ). Darüber hinaus zeigt sich für diesen Faktor ein signifikant negativer Zusammenhang mit *offen für Veränderungen* ( $r = -.30$ ) und *planvoll* ( $r = -.29$ ), was nicht angenommen worden war.

Zwischen dem Faktor Offenheit für Neues und den Skalen *einfallsreich* ( $r = .54$ ) und *offen für Veränderungen* ( $r = .24$ ) zeigen sich erwartungsgemäß signifikant positive Korrelationen, negative dagegen erwartungsgemäß mit *planvoll* ( $r = -.29$ ) und *pflichtbewusst* ( $r = -.33$ ).

Verträglichkeit korreliert signifikant positiv mit *verträglich* ( $r = .36$ ), nicht jedoch mit *beobachtend*. Zwar gibt es wie erwartet einen signifikanten negativen Zusammenhang zwischen Verträglichkeit und *wettbewerwend* ( $r = -.42$ ), jedoch nicht mit der Skala *leitend*.

Schließlich korreliert Gewissenhaftigkeit erwartungsgemäß signifikant positiv mit *pflichtbewusst* ( $r = .39$ ) und *planvoll* ( $r = .26$ ). Allerdings sind die angenommenen negativen Korrelationen von Gewissenhaftigkeit mit *überzeugend* oder *einfallsreich* nicht signifikant.

**Tabelle 4: Korrelationsmatrix von *shapes* und NEO-FFI**

	Neurotizismus	Extraversion	Offenheit	Verträglichkeit	Gewissenhaftigkeit
leitend	-.19	.18	.01	-.19	-.03
überzeugend	-.24*	.26*	-.04	-.04	-.15
souverän	-.38*	.36*	.23	.07	-.01
kontaktfreudig	-.04	.51*	-.07	.24*	-.16
verträglich	.17	-.02	-.33*	.36*	-.14
beobachtend	.05	.06	.20	.17	-.12
umsichtig	.01	-.11	-.08	-.08	.05
ergebnisorientiert	.10	.01	.17	.31*	.13
planvoll	.00	-.29*	-.29*	-.06	.26*
pflichtbewusst	.30*	-.03	-.33*	.40*	.39*
analysierend	.03	-.43*	-.19	-.26*	-.06
konzeptionell	.04	-.15	.27*	-.08	-.07
einfallsreich	-.11	-.01	.54*	.01	.08
offen für Veränderung	-.16	-.30*	.24*	-.15	.04
unabhängig	.13	-.13	.06	-.22	-.18
ehrzeigig	.04	-.02	-.12	-.19	.33*
wettbewerwend	.07	-.02	-.18	-.42*	-.15
ausdauernd	-.12	-.01	.12	.11	.25*

\*Anmerkung: Die Korrelation ist auf dem Niveau von 5% (zweiseitig) signifikant

## 6 Diskussion

In der vorliegenden Studie wurden die Zusammenhänge der *shapes* Skalen mit den Faktoren des NEO-FFI korreliert. Größtenteils zeigten sich in den Daten die erwarteten Zusammenhänge, jedoch nicht in allen Fällen. Im Einzelnen lassen sich die Ergebnisse wie folgt beschreiben: Neurotische Personen fühlen sich in Situationen unwohl, in denen sie im Fokus stehen und beschreiben sich entsprechend nicht als überzeugend und souverän im Fragebogen *shapes*. Tendenziell übernehmen sie auch nicht gerne die Leitung von Gruppen, wobei der Zusammenhang zwischen neurotisch und der *shapes* Skala leitend nicht signifikant ist. Neurotische Personen beschreiben sich darüber hinaus als pflichtbewusst, was sich dahingehend interpretieren lässt, dass sie sich Sicherheit zu verschaffen versuchen, indem sie sich an Regeln halten. Ein solches Vorgehen mag vermeintlich auch ein Scheitern unwahrscheinlicher machen, da man sich an Bewährtes halten kann.

Extravertierte Personen fühlen sich in Gruppen wohl und sind durchsetzungsfähig. Entsprechend beschreiben sie sich im *shapes* Fragebogen als kontaktfreudig, souverän und überzeugend. Dies bedeutet jedoch nicht notwendigerweise, dass sie auch die Leitung in einer Gruppe übernehmen wollen. Sie treffen Entscheidungen eher auf der Basis von Gefühlen als auf der von Fakten, beschreiben sich also nicht als analysierend.

Menschen, die offen für Erfahrungen sind, beschreiben sich in *shapes* erwartungsgemäß als offen für Veränderungen und einfallsreich. Dagegen beschreiben sie sich nicht als planvoll und auch nicht als pflichtbewusst. Dies könnte man so interpretieren, dass sie Veränderungen zwar durch ihre Offenheit und ihre Ideen anstoßen, dass aber jemand anderes die Veränderungen auch tatsächlich umsetzen und die Dinge zu Ende bringen muss.

Verträgliche Menschen müssen nicht notwendigerweise ein Interesse am Verhalten und an den Motiven anderer haben, sie könnten auch aus anderen Gründen verträglich sein. Sie gehen zwar ungerne den Wettbewerb mit anderen ein. Tendenziell möchten sie auch nicht die Leitung einer Gruppe übernehmen, aber ähnlich wie bei neurotischen Personen scheint auch dies nicht immer der Fall zu sein – es gibt einen negativen, aber nicht signifikanten Zusammenhang zwischen NEO-FFI verträglich und *shapes* leitend.

Gewissenhafte Menschen gehen erwartungsgemäß planvoll und pflichtbewusst an Aufgaben heran. Dies bedeutet jedoch nicht, dass sie keine neuen Ideen hervorbringen oder niemand anderen von ihren Ideen überzeugen können.

Insgesamt lässt sich sagen, dass *shapes* ein etwas differenzierteres Bild der Persönlichkeit liefert als der NEO-FFI, dabei aber nicht so umfassend ist wie letzterer. Während der NEO-FFI sehr breite Eigenschaften erfasst, sind diese im *shapes* Modell deutlich schmaler. Dies ist auch ein möglicher Erklärungsansatz für die erwarteten, jedoch eher geringen Korrelationen, wie z.B. zwischen Offenheit für Neues und *offen für Veränderungen*. Während die

Items des Big Five Modells sehr allgemein gehalten sind und eine breite Spanne an Verhaltensmustern abdecken, spiegeln die Items des *shapes* Modells konkret das Verhalten im beruflichen Kontext wieder. So werden, anders als im NEO-FFI, im *shapes* Aspekte wie Kunst oder Essensgewohnheiten nicht abgebildet. Für die Diagnostik bedeutet dies, dass mit Hilfe von *shapes* die Passung von Kandidaten zu recht differenzierten Anforderungsprofilen geprüft werden kann. So lässt sich beispielsweise unterscheiden, ob sich jemand „nur“ gut durchsetzen kann oder ob er tatsächlich die Leitung einer Gruppe übernehmen möchte. Es lässt sich ableiten, ob jemand möglicherweise einen Wandel im Unternehmen anstoßen wird (offen für Veränderungen, einfallsreich) und ob er auch in der Lage sein wird, die Veränderungen umzusetzen und zu Ende zu bringen (planvoll und pflichtbewusst).

Dennoch sind einige Aspekte der Studie kritisch zu hinterfragen. So ist die zugrundeliegende Stichprobengröße von  $N = 68$  ist zwar ausreichend, dennoch relativ niedrig. Vor allem gilt zu bedenken, dass die gewählte Stichprobe in Bezug auf Bildungsgrad und Motivation hochgradig vorselektiert ist. Beworben haben sich  $N = 696$  Studenten um an dem Finaltag teilnehmen zu können. Die NEO-FFI Ergebnisse liegen allerdings nur für die 10% Finalisten vor, die an dem Finaltag tatsächlich teilnehmen durften. In Bezug auf wissenschaftliche Güte wäre zu wünschen gewesen, wenn die Studie nicht nur die 68 Finalisten, sondern die Gesamtheit der Bewerber ( $N = 696$ ) umfasst hätte.

## 7 Fazit

In diesem Artikel wurde der Persönlichkeitsfragebogen *shapes* sowie die dazugehörige Messmethode *adalloc* detailliert betrachtet. Anhand zweier Studien wurde gezeigt, dass der Persönlichkeitsfragebogen *shapes* reliabel ist und damit eine wesentliche Voraussetzung für den Einsatz in der Personalauswahl erfüllt. Zudem zeigte eine Validierung mit dem BigFive-Instrument NEO-FFI, dass *shapes* leicht andere im Sinne von engeren Traits erfasst als der NEO-FFI.

Einleitend wurde dargelegt, wie die erfassten Skalen mittels inhaltsanalytischer Reduktion identifiziert wurden. Es kann diskutiert werden, inwieweit das Modell alle Eigenschaften umfasst, die in der beruflichen Praxis relevant sind. Darüber hinaus konnte mit diesem Verfahren weder sichergestellt werden, dass die Skalen gleiche Breite aufweisen noch, dass sie ausreichende Trennschärfen besitzen. Andererseits wurde durch diese Methode der Anwendbarkeit und Ökonomie für die berufliche Praxis Rechnung getragen. So wird beispielsweise im NEO-PI-R Ästhetik als Facette von Offenheit für Erfahrung erfasst (Sarges & Wottawa, 2001). Es ist jedoch anzunehmen, dass diese Eigenschaftsdimension nur wenig Varianzaufklärung bei der Vorhersage von Berufserfolg erzeugt. Da verschiedene berufsnahe Quellen die Basis für die Eigenschaften von *shapes* bilden, ist anzunehmen, dass die Skalen und folglich auch der Fragebogen in Summe eine höhere Varianzaufklärung als der NEO-PI-R bei der Vorhersage beruflich erfolgreichem Verhaltens erzeugt. Hierzu ist jedoch weitere Forschung mit empirischen Belegen gefordert.



Zu Beginn des Artikels wurden verschiedene Probleme aufgezeigt, denen aktuell eingesetzte Verfahren begegnen müssen. Eines der aufgezeigten Merkmale ist die Absicherung gegen Verfälschungstendenzen. Durch die Darstellung und Messmethode des Fragebogens *shapes* können die Vorteile ipsativer Datenformate übernommen werden. Weiterhin kann die Bearbeitungsdauer gegenüber Fragebögen mit rein ipsativen Formaten deutlich gesenkt werden, was von den Teilnehmern höchstwahrscheinlich positiv wahrgenommen wird und die soziale Akzeptanz erhöht.

In Studie 1 wurde die Retest-Reliabilität des Instruments geprüft. Aus den Ergebnissen kann gefolgert werden, dass das Instrument eine zuverlässige Messung ermöglicht. Das ist nach Asendorpf (2007) besonders für Persönlichkeitseigenschaften relevant, da oftmals – so auch im Kontext eines Bewerbungsverfahrens – nur auf die Selbstbeschreibung zurückgegriffen werden kann. Hierbei spielen Verfälschungstendenzen sowie Diskrepanzen zwischen Selbst- und Fremdwahrnehmung eine überaus große Rolle, weswegen die teils sehr guten Reliabilitätswerte umso aussagekräftiger für das Verfahren sprechen. Jedoch wurden in Studie 1 lediglich Retest-Reliabilitäten als Maß für die Zuverlässigkeit herangezogen. Denn, wie Lohff und Wehrmaker (2008) darlegen, ist für das Instrument eine normale Cronbachs-Alpha Berechnung problematisch. Die interne Konsistenz wird in Ermangelung anderer Methoden daher mit Hilfe eines modifizierten Alpha ermittelt. Diese Modifikation müsste zunächst in einer Studie im Hinblick auf ihre Genauigkeit und Robustheit geprüft werden.

In Studie 2 wurde eine konvergente Validierung im Sinne einer MTMM anhand des Fünf-Faktoren-Modells durchgeführt. Es wurde untersucht, ob die Skalen des *shapes* und des NEO-FFI erwartungskonform korrelieren. Diese Frage kann überwiegend mit ja beantwortet werden. Zumeist konnten die Hypothesen angenommen werden, jedoch wurden auch klare Unterschiede aufgezeigt. Ein Erklärungsansatz hierfür liefert die relative Breite der Eigenschaften, die bei den beiden Instrumenten stark variiert. Dennoch ist besonders in diesem Bereich weitere Forschung wünschenswert, sei es, um die hier gefundenen Ergebnisse zu replizieren. So zeigt sich beispielsweise bei der Eigenschaft Extraversion eine Vielzahl signifikanter Zusammenhänge. Mehrere Autoren gehen davon aus, dass das Ausmaß dieser Eigenschaft zum einen die Job Performance moderiert, vor allem aber die Entscheidung, ob ein Arbeitsvertrag angeboten wird (Bartram, 2005; Hossiep, Paschen & Mühlhaus, 2005; Hülsheger & Maier, 2008; Kurz, Bartram & Baron, 2004). Insofern ist es nicht verwunderlich, dass eine Anzahl von Skalen mit dieser Dimension korreliert, jedoch sollte diesbezüglich untersucht werden, ob die betreffenden Skalen des Fragebogens *shapes* ausreichende Trennschärfen aufweisen.

Nach Doppler und Lauterburg (2002) wird computergestützte (Personal)Diagnostik zunehmend notwendig. *Shapes* trägt dieser Notwendigkeit Rechnung und bietet Kandidaten wie auch Unternehmen durch Einhaltung der einleitend genannten Kriterien somit konkreten Mehrwert. Darüber hinaus hat sich *shapes* in den Studien als reli-

ables und valides Instrument erwiesen.

Durch die neu entwickelte Messmethode ist adaptives Testen erstmals auch für eine Selbstbeurteilung im Rahmen der Persönlichkeitsdiagnostik möglich. Durch den Einsatz des Fragebogens in der Personalauswahl werden z. B. international agierende Unternehmen unterstützt, unternehmensweit einheitliche Standards zu definieren indem überall das gleiche Instrument eingesetzt wird. Die Kandidaten müssen für die Testung keine Anreise in Kauf nehmen, was zuvor ein Selektionskriterium gewesen sein kann. Die Selbstbeurteilung kann in einem sehr frühen Stadium des Bewerbungsprozesses berücksichtigt werden, was ein umfassenderes Bild des Kandidaten und dadurch einen qualitativ hochwertigeren Prozess ermöglicht.

Allerdings besteht weiterhin noch viel Forschungsbedarf. So sollten weitere Überprüfungen hinsichtlich des Messmodells stattfinden. In diesem Zuge kann auch die Forschung zur Überprüfung der Gütekriterien bei ipsativen und hybriden Daten, wie bei Brown und Maydeu-Olivares (2012) oder Meade (2004) angesprochen, weiter vorangetrieben werden. Da der Fragebogen *shapes* zumeist in der beruflichen Eignungsdiagnostik Anwendung findet, ist in diesem Bereich eine Kriteriumsvalidierung von besonderem Interesse. Darüber hinaus wäre eine Konstruktvalidierung mit einem einschlägig validierten berufsbezogenen Persönlichkeitsinventar wünschenswert. Besonders die Ergebnisse der Studie 2 bedürfen weiterer Belege, da trotz der weiterführenden Ergebnisse keine Konstruktvalidität aufgezeigt werden konnte. Zu diesem Zweck könnte auch eine konfirmatorische Faktorenanalyse durchgeführt werden.

## 8 Literatur

- Amelang, M. & Bartussek, D. (2001). *Differentielle Psychologie und Persönlichkeitsforschung* (5. Aufl.). Stuttgart: Kohlhammer.
- Angleitner, A. & Riemann, R. (2005). Eigenschaftstheoretische Ansätze. In H. Weber & T. Rammsayer (Hrsg.), *Handbuch der Persönlichkeitspsychologie und Differentiellen Psychologie* (S. 93-103). Göttingen: Hogrefe.
- Asendorpf, J. B. (2012). *Psychologie der Persönlichkeit* (5. Aufl.). Berlin: Springer.
- Ashton, M. C. (1996). Personality and job performance: the importance of narrow traits. *Journal of Organizational Behavior*, 19, 289-303.
- Barak, A. & English, N. (2002). Prospects and Limitations of Psychological Testing on the Internet. *Journal of Technology in Human Science*, 19, 65-89.
- Bartram, D. (2007). Increasing validity with forced-choice criterion measurement formats. *International Journal of Selection and Assessment*, 15, 263-272.
- Bartram, D. (2005). The Great Eight Competencies: A Criterion-Centric Approach to Validation. *Journal of Applied Psychology*, 90, 1185-1203.
- Bartram, D. (1999). Testing and the Internet: Current realities, issues and future possibilities. *Selection and Development Review*, 15, 3-12.

- Bartram, D. (1996). The relationship between ipsatized and normative measures of personality. *Journal of Occupational and Organizational Psychology*, *69*, 25-39.
- Baron, H. (1996). Strength and Limitations of Ipsative Measurements. *Journal of Occupational and Organizational Psychology*, *69*, 49-56.
- Barrick, M. R. & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*, 1-25.
- Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, *14*, 317-334.
- Borkenau, P. & Ostendorf, F. (2008). *NEO-Fünf-Faktoren-Inventar (NEO-FFI) nach Costa und McCrae*. Göttingen: Hogrefe.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation: Für Human- und Sozialwissenschaftler* (4. Aufl.). Berlin: Springer.
- Brandstätter, H., Filipp, G. & Drescher, P. (1992). Zur Validität der 16 Persönlichkeits-Adjektivskalen (16PA). *Zeitschrift für Differentielle und Diagnostische Psychologie*, *13*, 53-67.
- Brown, A. & Maydeu-Olivares, A. (2012). Fitting a Thurstonian IRT model to forced-choice data using Mplus. *Behavior Research Methods*, *44*, 1135-1147.
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson.
- Bühner, M. & Ziegler, M. (2009). *Statistik für Psychologen und Sozialwissenschaftler*. München: Pearson.
- Camera, W. J. & Merenda, P. F. (2000). Using personality tests in preemployment screening: Issues raised in *Soroka v. Dayton Corporation*. *Psychology, Public Policy and Law*, *6*, 1164-1186.
- Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81-105. cut-e GmbH (2013). *shapes Inside*. Hamburg.
- De Vries, A., de Vries, R. & Born, M. P. (2010). Broad versus narrow traits: Conscientiousness and honesty-humility as predictors of academic criteria. *European Journal of Personality*, *25*, 336-348.
- Doppler, K. & Lauterburg, C. (2002). *Change Management. Den Unternehmenswandel gestalten* (10. Aufl.). Frankfurt: Campus Verlag.
- Dudley, N. M., Orvis, K. A., Lebiecki, J. E. & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology*, *91*, 40-57.
- Eysenck, M. W. & Keane, M (2015). *Cognitive Psychology. A Student's Handbook* (7. Aufl.). New York: Psychology Press.
- Fan, X. (1998). Item response theory and classical test theory: An empirical comparison of their item/person statistics. *Educational and Psychological Measurement*, *58*, 357-381.
- Fay, F., Bell, R. & Bannis, D. (2004). *A Manual for Repertory Grid Technique*. United Kingdom: Chichester.
- Fehr, T. (2006). Big Five: Die fünf grundlegenden Dimensionen der Persönlichkeit und ihre 30 Facetten. In W. Simon (Hrsg.), *Persönlichkeitsmodelle und Persönlichkeitstests. 15 Persönlichkeitsmodelle für Personalauswahl, Persönlichkeitsentwicklung, Training und Coaching* (S. 113-135). Offenbach: Gabal.
- Frey, A. (2007). Adaptives Testen. In: H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 261-278). Berlin: Springer.
- Gatewood, R. D., Feild, H. S. & Barrick, M. (2010). *Human Resource Selection* (7. Aufl.). South-Western: Mason.
- Griffith, R. L., Chmielowski, T. & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review*, *36*, 341-357.
- Griffith, R. L., & Converse, P. D. (2012). The rules of evidence and the prevalence of applicant faking. In M. Ziegler, C. MacCann, & R. D. Roberts (Hrsg.), *New perspectives on faking in personality assessment* (S. 34-51). New York: Oxford University Press.
- Gur, R. C., & Sackeim, H. A. (1979). Self-deception: A concept in search of a phenomenon. *Journal of Personality and Social Psychology*, *37*, 147-69.
- Heggstad, E. D., Morrison, M., Reeve C. L. & McCloy, R. A. (2006). Forced-Choice Assessments of Personality for Selection: Evaluating Issues of Normative Assessment and Faking Resistance. *Journal of Applied Psychology*, *91*, 9-24.
- Hell, B. & Schuler, H. (2005). Verfahren der Studierendenauswahl aus Sicht der Bewerber. *Empirische Pädagogik*, *19*, 361-376.
- Hertel, G., Konradt, U. & Orlikowski, B. (2003). Ziele und Strategien von E-Assessment aus Sicht der psychologischen Personalauswahl. In U. Konradt & W. Sarges (Hrsg.), *E-recruitment und E-Assessment* (S. 37-53). Göttingen: Verlag für Angewandte Psychologie.
- Hicks, L. E. (1970). Some properties of ipsative, normative and forced choice normative measures. *Psychological Bulletin*, *74*, 167-184.
- Hogan, J. (1998). Personality and Job Performance. *Human Performance*, *11*, 125-127.
- Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology*, *92*, 1270-1285.
- Hossiep, R., Paschen, M. & Mühlhaus, O. (2005). *Persönlichkeitstest im Personalmanagement*. Göttingen: Hogrefe.
- Hough, L. M. & Oswald, F. L. (2008). Personality Testing and Industrial-Organizational Psychology: Reflections, Progress, and Prospects. *Industrial and Organizational Psychology*, *1*, 272-290.
- Hülshager, U. R. & Maier, G. W. (2008). Persönlichkeitsei-

- enschaften, Intelligenz und Erfolg im Beruf. *Psychologische Rundschau*, 59, 108-122.
- Hurtz, G. M. & Donovan, J. J. (2000). Personality and Job Performance: The Big Five revisited. *Journal of Applied Psychology*, 85, 869-879.
- Jackson, C. J. & Corr, P. J. (1998). Personality- performance correlations at work: individual and aggregate levels of analyses. *Personality and Individual Differences*, 24, 393-403.
- Jackson, D. N., Wroblewski, V.R. & Ashton, M.C. (2000). The impact of faking on employment tests: Does forced-choice offer a solution? *Human Performance*, 13, 371-388.
- Justenhoven, R. T. (2014). *Adaptive allocation of – Innovative Itemformate zur Messung von Persönlichkeit*. Unveröffentlichte Masterarbeit. Hamburg: Hochschule Fresenius.
- Kanning, U. P. (2014). Prozess und Methoden der Personalentwicklung. In H. Schuler & U. P. Kanning (Hrsg.), *Lehrbuch der Personalpsychologie* (3. Aufl.), (S. 501-562). Göttingen: Hogrefe.
- Kersting, M (2008). Zur Akzeptanz von Intelligenz- und Leistungstests. *Report Psychologie*, 33, 420-433.
- Kühlmann, T. M. & Stahl, G. K. (2001). Problemfelder des internationalen Personaleinsatzes. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 533-557). Göttingen: Hogrefe.
- Kurz, R. & Bartram, D. (2002). Competency and individual performance: Modeling the world of work. In I. T. Robertson, M. Callinan & D. Bartram (Hrsg.), *Organizational effectiveness: The role of psychology* (S. 227-255). Chichester: Wiley.
- Kurz, R., Bartram, D. & Baron, H. (2004). Assessing potential and performance at work: The Great Eight competencies. *Proceedings of the British Psychological Society*, 4, 91-95.
- Leslie, J. B. & Van Velsor, E. (1996). *A Look at Derailment Today: North America and Europe*. Greensboro, NC: Center for Creative Leadership.
- Lochner, K., Preuß, A. & Lohff, A. (2015). Innovative Wege im Executive Assessment. In S. Weinert & K. Stulle (Hrsg.), *Executive Assessment. Instrumente, Trends, Herausforderungen* (S. 171-186). Berlin: Springer.
- Lohff, A. & Wehrmaker, M. (2008). Adaloc™ – adaptive scales for online questionnaires. In W. Sarges & D. Scheffer (Hrsg.), *Innovative Ansätze für die Eignungsdiagnostik* (S. 239-251). Göttingen: Hogrefe.
- Lombardo, M. M. & Eichinger, R. W. (2000). *The Leadership Machine: Architecture to Develop Leaders for Any Future*. Minneapolis, MN: Lominger Limited Inc.
- Mahar, D., Cologon, J., & Duck, J. (1995). Response strategies when faking personality questionnaires in a vocational selection setting. *Personality and Individual Differences*, 18, 605-609.
- Mayring, P. (2000). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (7. Aufl.). Weinheim: Deutscher Studien Verlag.
- McCall, M. W. & Lombardo, M. M. (1983). *Off the Track: Why and How Successful Executives Get Derailed* (Technical Report / Center for Creative Leadership, No. 21). Greensboro, NC: Center for Creative Leadership.
- McFarland, L. A., Ryan, A. M., & Kriska, S. D. (2003). Impression Management Use and Effectiveness Across Assessment Methods. *Journal of Management*, 29, 641-661.
- Meade, A. M (2004). Psychometric problems and issues involved with creating and using ipsative measures for selection. *Journal of Occupational and Organizational Psychology*, 77, 531-551.
- Moser, K. (1999). Selbstbeurteilung beruflicher Leistung. Überblick und offene Fragen. *Psychologische Rundschau*, 50, 14- 25.
- Moser, K., Donat, M., Schuler, H., Funke, U. & Roloff, K. (1994). Validität der Selbstbeurteilung beruflicher Leistung. Eine Untersuchung im Bereich industrieller Forschung und Entwicklung. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 41, 474-500.
- Ostendorf, F. & Angleitner, A. (2004). *NEO- Persönlichkeitsinventar nach Costa und McCrae: NEO- PI-R*. Göttingen: Hogrefe.
- Paulhus, D. L. (1984). Two-component models of social desirable responding. *Personality Processes and Individual Differences*, 46, 598-609.
- Paunonen, S. V. & Ashton, M. C. (2001). Big Five Predictors of Academic Achievement. *Journal of Research in Personality*, 35, 78-90.
- Preuß, A. (2006). *Making online ability tests forgery-proof*. Presentation at the 26<sup>th</sup> International Congress of Applied Psychology, Athens.
- Roth, C., Wegge, J. & Schmidt, K.-H. (2007). Konsequenzen des demographischen Wandels für das Management von Humanressourcen. *Zeitschrift für Personalpsychologie*, 6, 99-116.
- Rothstein, M. G., & Goffin, R. D. (2006). The use of personality measures in personnel selection: What does current research support? *Human Resource Management Review*, 16, 155-180.
- Rost, D. H. & Hoberg, K. (1997). Itempositionsveränderungen in Persönlichkeitsfragebögen: Methodischer Kunstfehler oder tolerierbare Praxis? *Diagnostica*, 43, 97-112.
- Salgado, J. F. (2003). Predicting job performance using FFM and non-FFM personality measures. *Journal of Occupational and Organizational Psychology*, 76, 323-346.
- Salgado, J. F. (1997). The five factor model of personality and job performance in the European Community. *Journal of Applied Psychology*, 82, 30-43.
- Sands, W. A., Waters, B. K. & McBride, B. K. (1997). *Computerized Adaptive Testing: From Inquiry to Operation*. Washington, DC: APA.

- Sarges, W. & Wottawa, H. (2001). *Handbuch wirtschaftspsychologischer Testverfahren*. Berlin: Science Publishers.
- Saville & Holdsworth Ltd. (1993) *OPQ Factor Model Manual and Users' Guide*. Thames Ditton, UK: Saville & Holdsworth Ltd.
- Saville, P. & Willson, E. (1991). The reliability and validity of normative and ipsative approaches in the measurement of personality. *Journal of Occupational Psychology*, *64*, 219-238.
- 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.
- Schuler, H. (2014). Arbeits- und Anforderungsanalyse. In Schuler & U. P. Kanning (Hrsg.), *Lehrbuch der Personalpsychologie* (3. Aufl.), (S. 61-98). Göttingen: Hogrefe.
- Schuler, H., Hell, B., Trapmann, S., Schaar, H. & Boramir, I. (2007) Die Nutzung psychologischer Verfahren der externen Personalauswahl in deutschen Unternehmen. *Zeitschrift für Personalpsychologie*, *6*, 60-70.
- Schuerger, J.M. (2001). *16PF Adolescent Personality Questionnaire*. Champaign, IL: IPAT.
- Sitser, T., van der Linden, D. & Born, M. P. (2013). Predicting Sales Performance with Personality Measures: the Use of the General Factor of Personality, the Big Five and Narrow Traits. *Human Performance*, *26*, 126-149.
- Smith, A. (2005). *Management competencies for support service managers in the Irish health services: identifying management skills and attributes*. Dublin: Technical Report.
- Stocking, M. L. & Lewis, C. (2000). Methods of controlling the exposure of items in CAT. In: W. J. van der Linden & C. A. W. Glas (Hrsg.), *Computerized adaptive Testing: Theory and practice* (S. 163-182). Norwell MA: Kluwer.
- Tenopyr, M. L. (1988) Artefactual reliability of forced-choice scales. *Journal of Applied Psychology*, *73*, 749-751.
- Tett, R. P., Freund, K. A., Christiansen, N. D., Fox, K. E., & Coaster, J. (2012). Faking on self-report emotional intelligence and personality tests: Effects of faking opportunity, cognitive ability and job type. *Personality and Individual Differences*, *52*, 195-201.
- Tett, R. P., Jackson, D. N. & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology*, *44*, 703-745.
- Tonidandel, S. & Quiñones, M. A. (2000). Psychological Reactions to Adaptive Testing. *International Journal of Selection and Assessment*, *8*, 7-15.
- Vecchione, M., Alessandri, G., & Barbaranelli, C. (2011). The five factor model in personnel selection: Measurement equivalence between applicant and nonapplicant groups. *Personality and Individual Differences*, *52*, 503-508.
- Visser, D. & Du Toit, J. M. (2004). Using the occupational personality questionnaires (OPQ) for measuring broad traits. *Journal of Industrial Psychology*, *30*, 65-77.
- Wainer, H. (2000). *Computerized adaptive testing: A primer*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Warr, P., Bartram, D. & Brown, A. (2005). Big Five validity: Aggregation method matters. *Journal of Occupational and Organizational Psychology*, *78*, 377-386.
- Wiechmann, D. & Ryan, A. M. (2003). Reactions to Computerized Testing in Selection Contexts. *International Journal of Selection and Assessment*, *11*, 215-229.
- Yan, D., Lewis, C. & Stocking, M. (1998). *Adaptive Testing without IRT*. San Diego: Annual Meeting of the National Council on Measurement in Education (NCME).
- Zempel, J. & Moser, K. (2005). Feedback als Moderator der Validität von Selbstbeurteilungen. *Zeitschrift für Personalpsychologie*, *4*, 19-27.
- Zhou, J., Gierl, M. J. & Cui, Y. (2007). *Computerized attribute-adaptive testing: A new computerized adaptive testing approach incorporating cognitive psychology*. Minneapolis: GMAC Conference on Computerized Adaptive Testing.

#### Korrespondenzadresse

Richard Justenhoven  
 Großer Burstah 18-32  
 20457 Hamburg  
 DEUTSCHLAND  
 richard.justenhoven@cut-e.com