

Forschungsbericht

Anmerkungen zur Aussagekraft des Artikels „Validität des BIP und des NEO-PI-R“ (Hülshager, Specht & Spinath, 2006)

Projektteam Testentwicklung, 2018

Verfasser: Robin Merchel, Philip Frieg & Rüdiger Hossiep

Projektteam Testentwicklung
c/o Dr. R. Hossiep
Ruhr-Universität Bochum
Fakultät für Psychologie
IB 5/55
44780 Bochum

www.testentwicklung.de
bip@ruhr-uni-bochum.de
Fon 0234/32-24623



Link zum PDF

Vorbemerkung

Hülshager, Specht und Spinath (2006) setzten in ihrem Artikel „Validität des BIP und des NEO-PI-R“ unter anderem das *Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung* (BIP) ein, um Berufserfolgskriterien zu erklären. Die Studie fand durchaus Resonanz und wird in einschlägigen Publikationen zitiert (z.B. Borkenau, Friedel & Wolfradt, 2011; Kauffeld & Grohmann, 2011; Obermann, 2018), weswegen wir uns umfassend mit der Methodik der Studie auseinandersetzten. Hierbei mussten wir feststellen, dass die Studie unseres Erachtens so gravierende methodische Mängel aufweist, dass wir dies auch nach mehr als einer Dekade doch nicht unkommentiert lassen wollen und uns veranlasst gesehen haben, diese Stellungnahme zu verfassen. Die zentralen Kritikpunkte beziehen sich auf die Stichprobenszusammensetzung und die Operationalisierung der objektiven Berufserfolgskriterien (insbesondere Entgelt).

Während Hülshager et al. (2006) zum Ergebnis kommen, dass beide Verfahren vergleichbare inkrementelle Validität gegenüber dem jeweils anderen und eine vergleichbare Akzeptanz auf Seiten der Teilnehmer aufweisen, belegt eine aktuellere Metaanalyse (Shaffer & Postlethwaite, 2012) mit Blick auf Validitätsmaße die Überlegenheit kontextadäquater Verfahren im beruflichen Umfeld. Wir möchten mit dem vorliegenden Forschungsbericht eine erweiterte Perspektive geben und aufzeigen, dass die Problematiken in der Methodik bei Hülshager et al. (2006) unseres Erachtens die Ergebnisse, die auch von den Autoren als erwartungswidrig beurteilt wurden, beeinflusst haben können.

Zusammenfassung der Publikation von Hülshager et al. (2006)

Die Autoren wollten der Frage nachgehen, ob und inwieweit ein berufsbezogenes Testverfahren zur Erfassung der Persönlichkeit wie das Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung von Hossiep und Paschen (2003) einem nicht-berufsbezogenen, zumeist in der Forschung eingesetzten Verfahren wie dem NEO-PI-R von Ostendorf und Angleitner (2004) im Einsatz im Berufskontext in Hinblick auf objektive und subjektive Berufserfolgskriterien überlegen ist. Eine Stichprobe von $N = 90$ Berufstätigen wurde durch persönliche Ansprache oder Weitergabe der Fragebögen an Familie und Kollegen rekrutiert und füllte stets zunächst NEO-PI-R und danach das BIP aus.

Hierarchische Regressionen mit verschiedensten Kontrollvariablen (Alter, Geschlecht, Ausbildungsniveau, durchschnittliche Ausbildungsnote und Dauer der Tätigkeit) wurden von den Autoren mit dem Ergebnis durchgeführt, dass beide Verfahren bedeutsame Beiträge zur Varianzaufklärung objektiver und subjektiver Berufserfolgskriterien leisteten; die über das

jeweils andere Verfahren hinausgehende Varianzaufklärung war hierbei für beide Verfahren vergleichbar. Weiterhin zeigten sich in Bezug auf die beurteilte Akzeptanz keine substantiellen Unterschiede.

Anmerkungen

Stichprobe und Berufserfolgskriterien

Die Stichprobe besteht aus $N = 90$ Berufstätigen, die durch persönliche Ansprache der Versuchsleiterin rekrutiert wurden. Für die meisten relevanten Berechnungen (Regressionsanalysen) schrumpft dieser Datensatz weiter auf $N = 77$. Diese Stichprobe ist vergleichsweise klein. Bei Fragestellungen dieser Art sind größere Stichproben üblich, wie Tabelle 1 zeigt.

Tabelle 1: Beispiele für Stichprobengrößen bei Untersuchungen zu Konstrukt- und Kriteriumsvalidität.

Quelle	Fragestellung	N	
Hossiep & Paschen (2003)	Konstruktvalidität	363	(BIP und NEO-FFI)
		155	(BIP und 16-PF-R)
Hossiep & Paschen (2003)	Kriteriumsvalidität (BIP)	5.691	(berufliches Entgelt)
		5.208	(hierarchische Position)
Ostendorf & Angleitner (2004)	Konstruktvalidität (gemeinsame Faktorenanalyse)	242	(NEO-PI-R und Gießen-Test)
		236	(NEO-PI-R und Trierer Persönlichkeitsfragebogen)
		184	(NEO-PI-R und BIP)
Hülshager et al. (2006)	Kriteriumsvalidität (NEO-PI-R und BIP)	77	(Bruttoeinkommen)
		79	(Berufsstatus)
Hossiep & Krüger (2012)	Kriteriumsvalidität (BIP-6F)	6.234	(hierarchische Position)
		4.541	(Entgelt)
Frieg & Schulz (2013)	Konstruktvalidität (gemeinsame Faktorenanalyse)	414	(BIP und Deutsche Personality Research Form)
Frieg, Pohlmann & Hossiep (2014)	Kriteriumsvalidität (BIP-6F)	12.292	(hierarchische Position)
		12.017	(Entgelt)
		12.602	(subjektiver Berufserfolg)

Weiterhin ist hierbei als problematisch anzusehen, dass 20 der 90 Berufstätigen angaben, im Bereich der Finanzdienstleistungen zu arbeiten, wodurch diese Berufsgruppe überrepräsentiert ist. Die Datenbasis der Studie von Hülshager et al. (2006) ist also eher gering und für die heterogene Gruppe der Fach- und Führungskräfte nur eingeschränkt repräsentativ. Hier ist zu bedenken, dass für spezifische Stichproben unterschiedliche Zusammenhänge zwischen Skalen und Erfolgskriterien gelten können. So zeigten Hossiep, Mutwill und Schulz (2012), dass für Sachbearbeiter mit Blick auf Erfolgskriterien zum Teil andere BIP-Skalen relevant sind als für Vorstände. Dies kann durchaus auch für die Stichprobe von Hülshager (2006) gelten. Das heißt, die Ergebnisse sind möglicherweise nicht auf Berufstätige verschiedener Berufsgruppen und unterschiedlicher Hierarchieebenen zu übertragen - dies wird allerdings durch den Titel suggeriert.

Ein Blick auf die Kriterien lässt weitere Zweifel an der Belastbarkeit der Befunde aufkommen. Tabelle 2 vergleicht die Entgeltkategorien von Hossiep und Paschen (2003) mit denen von Hülshager et al. (2006).

Tabelle 2: Vergleich der Entgeltkategorien zweier Studien.

Hossiep & Paschen (2003)	Hülshager et al. (2006) (vermutet)¹
unter 60.000 DM	unter 12.000 EUR
60.000 bis 80.000 DM	12.000 bis 18.000 EUR
80.000 bis 100.000 DM	18.000 bis 24.000 EUR
100.000 bis 130.000 DM	24.000 bis 30.000 EUR
130.000 bis 170.000 DM	30.000 bis 36.000 EUR
170.000 bis 250.000 DM	36.000 bis 42.000 EUR
250.000 bis 400.000 DM	42.000 bis 48.000 EUR
größer als 400.000 DM	größer als 48.000 EUR

¹ Die konkreten Entgeltkategorien von Hülshager et al. (2006) vermuten wir in dieser Form auf Grundlage der Angaben im Artikel sowie schlussfolgernd aus Judge et al. (1999), da sichere Angaben nicht zu entnehmen waren.

Rechnet man der Einfachheit halber um (1 DM = 0,5 EUR), so ergibt sich bei Hossiep und Paschen eine Spanne von 30.000 bis 200.000 EUR (plus offene Kategorie) und bei Hülshager et al. (2006) eine Spanne von 12.000 bis 48.000 EUR (plus offene Kategorie). Das Spektrum beginnt bei Hülshager et al. (2006) also bei einem deutlich niedrigeren Wert und ist insgesamt auch erheblich enger. Die Autoren rechtfertigen dies mit einer Anlehnung an Judge, Higgins, Thoresen und Barrick (1999), die sechs Kategorien von „1 = less than \$15,000“ bis „6 = \$50,000 and over“ verwendeten.

Hier stellt sich die Frage, ob mit einer solchen Skalierung überhaupt Berufserfolg (im Sinne eines hohen Entgeltes) sinnvoll erfasst werden kann. Vor allem die offene Kategorie „ab 48.000 EUR“ muss hier als problematisch angesehen werden. Das statistische Bundesamt (2017) gibt als durchschnittlichen Bruttomonatsverdienst Vollzeitbeschäftigter für 2006 2.950€ an, dies entspräche auf ein Jahr gerechnet 35.400€. Mit einem durchschnittlichen Gehalt für das Jahr 2006 läge man folglich bei Hülshager et al. (2006) bereits im oberen Bereich der Spannweite. Offensichtliche Unterschiede im Gehalt werden hier nivelliert. So muss es doch im Sinne eines Berufserfolgskriteriums unterschiedlich bewertet werden, ob eine Person 50.000 EUR oder 100.000 EUR p. a. erhält. Bei Hülshager et al. (2006) fallen diese beiden Summen aber in dieselbe Kategorie, weswegen diese Kategorisierung im überdurchschnittlichen Lohnbereich als nicht differenzierungsfähig angesehen werden muss.

Anscheinend besteht die Stichprobe tendenziell aus Personen, die über eher geringe Einkommen verfügen. Wenn dem so ist, ist es aber wenig zielführend, überhaupt Berechnungen zur Kriteriumsvalidität mit Blick auf das berufliche Entgelt anzustellen. Abbildung 1 zeigt die Entgeltkategorien mit den entsprechenden Häufigkeiten, die Grundlage für die BIP-Manualberechnungen (Hossiep & Paschen, 2003) waren.

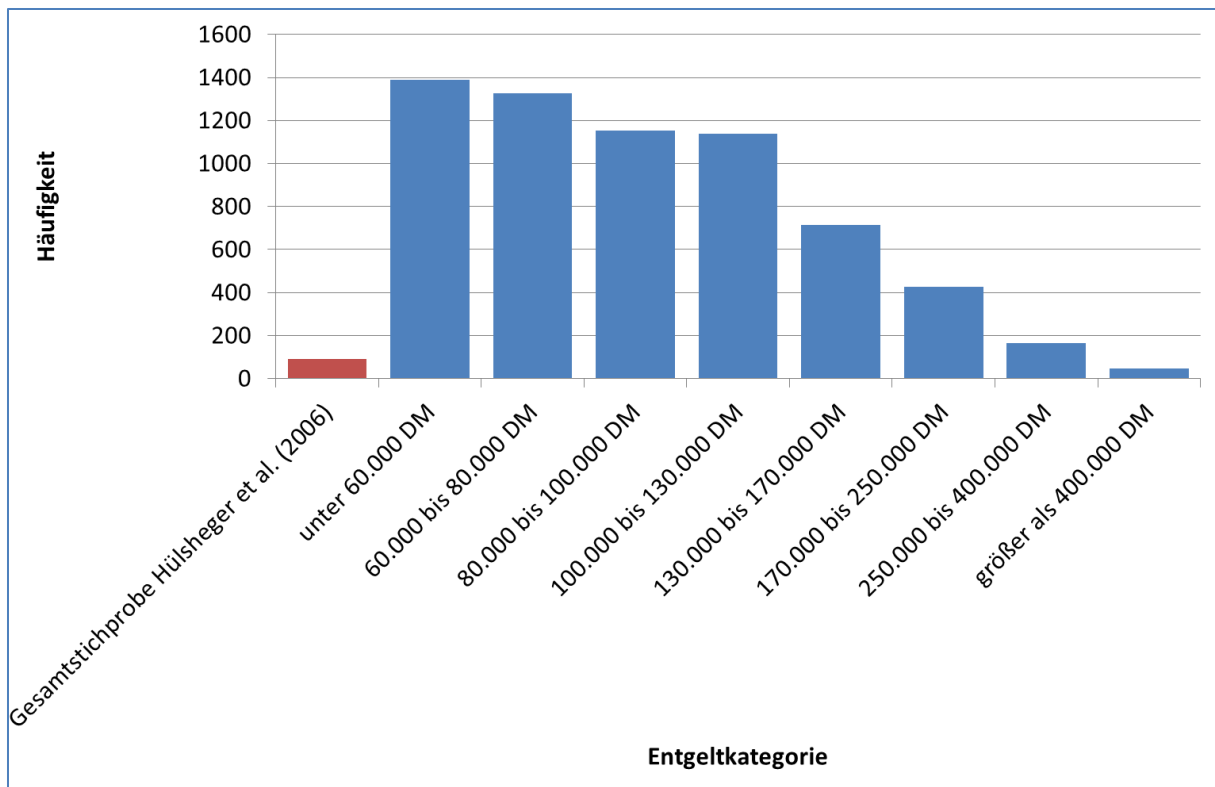


Abbildung 1: Stichprobengrößen in den Entgeltkategorien bei Hossiep und Paschen (2003) im Vergleich zur Gesamtstichprobe bei Hülshager et al. (2006).

Zusammenfassend muss infrage gestellt werden, ob mithilfe einer Stichprobe von $N = 90$ Personen, deren Einkommen größtenteils deutlich unter 48.000 EUR liegen muss (wie durch die Kategorisierung nahegelegt), belastbare Aussagen zum Zusammenhang zwischen Persönlichkeitsskalen und Berufserfolg getroffen werden können.

Berufsstatus als Kriteriumsvariable wurde von Hülshager et al. (2006) operationalisiert, indem dieser in nur vier Kategorien von „1 = sonstiger Arbeiter“ bis „4 = Inhaber und/oder Geschäftsführer eines Unternehmens“ aufgeteilt wurde. Fachkräfte als eine wichtige Zielgruppe des BIP dürften sich hierbei ausschließlich in den mittleren Kategorien 2 und 3 (bei denen nicht spezifiziert ist, welche Positionen sie überhaupt umfassen) wiederfinden – wie bereits beim Entgelt eine übermäßige Vereinfachung realer Unterschiede, die Differenzierung erschwert.

Akzeptanz bzw. Augenscheinvalidität

Eine weitere Schwäche der Studie bezieht sich auf das Thema Akzeptanz bzw. Augenscheinvalidität. Die Autoren stellen nachvollziehbar die Hypothese auf, dass das BIP aufgrund des expliziten Berufsbezugs eine höhere Akzeptanz erzielen müsste als das NEO-PI-R. Fraglich ist, ob dies anhand der Stichprobe überhaupt prüfbar ist, da es sich um kein Setting

handelt, in dem berufsbezogene Fragebogenverfahren üblicherweise zum Einsatz kommen (Personalauswahl, Personalentwicklung, Coaching etc.).

Zu der auch aus juristischer Perspektive (vgl. Maties & Wottawa, 2011) interessanten Frage „In welchem Ausmaß zeigt der Inhalt dieses Fragebogens Beziehungen zu Ihrer Berufstätigkeit?“ berichten die Autoren allerdings keine separaten Ergebnisse. Stattdessen werden lediglich Ergebnisse zu „Akzeptanzfaktoren“ berichtet, in die diese Frage mit einfließt und die für beide Verfahren jeweils ca. 50% der Varianzen erklären. Die Autoren berichten beim Vergleich der Akzeptanzwerte eine nicht signifikante Effektstärke von $d = .13$. Die Autoren begründen nicht, warum Itemergebnisse der durchaus aussagekräftigen Items nicht berichtet werden. Durch den Verzicht auf ca. 50% der Varianz könnten Ergebnisse unentdeckt bleiben bzw. verdeckt werden.

In der Diskussion bei Hülshager et al. (2006) wird darauf hingewiesen, dass es sich möglicherweise negativ auf die Akzeptanzwerte des BIP ausgewirkt haben könnte, dass das BIP im Anschluss an das NEO-PI-R beantwortet wurde. Sprich: Die Ermüdung kann zu geringeren Akzeptanzwerten geführt haben. Dies ist plausibel. NEO-PI-R und BIP haben schließlich zusammengenommen immerhin 450 Items. Die Bearbeitung einer solch langen Testbatterie stößt auch aufgrund von Ermüdungseffekten nach aller Erfahrung an die Grenzen der Zumutbarkeit. Gegebenenfalls erklärt dies auch die geringe Stichprobengröße – aber eben auch tendenziell geringere Akzeptanzwerte für das zweite der beiden Testverfahren (hier: BIP-Fragebogen). Hier bleibt die naheliegende Frage offen, warum nicht eine Überkreuzapplikation der Verfahren stattgefunden hat.

Weiterhin schreiben die Autoren „Aufgrund des stärkeren Arbeitsbezugs des BIP wird angenommen, dass das BIP eine höhere Akzeptanz erfährt als das NEO-PI-R“ (Hülshager et al., 2006, S. 137). Die kurz zuvor zitierte Quelle von Hossiep, Paschen und Mühlhaus (2000) spricht nicht davon, dass ein berufsbezogenes Verfahren in jedem Falle eine höhere Akzeptanz erzielen muss als ein nicht-berufsbezogenes Verfahren, sondern dass dieser Zusammenhang in berufsbezogenen Situationen wie Auswahl-situationen zu erwarten ist. Gleichwohl wird in der Studie von Hülshager et al. (2006) nicht deutlich, dass die beiden Verfahren in einem berufsbezogenen Kontext eingesetzt wurden – im Gegenteil, es heißt „Darüber hinaus gaben zahlreiche Teilnehmer Fragebögen an weitere Personen aus ihrem privaten oder beruflichen Umfeld weiter“ (S.138). Das bedeutet, dass vermutlich kein Teilnehmer die Verfahren in einer realen Berufssituation wie einer Auswahl-situation beantwortete und wohl auch Fragebögen explizit außerhalb jeglichen Berufskontextes ausgefüllt worden sind. Es überrascht nicht, dass dort keine überlegene Akzeptanz des BIP gefunden wurde; schließlich wurde die unzulässige

Verletzung der Privatsphäre, die nach Hossiep et al. (2000) ursächlich für die verringerte Akzeptanz nicht berufsbezogener Verfahren ist, hier außerhalb von berufsbezogenen Situationen weniger empfunden.

Es ergäbe sich folglich als sinnvollere Hypothese, dass das BIP in echten Bewerbungssituationen im Vergleich zum NEO-PI-R in höherem Maße akzeptiert wird. Diese wäre auch in tatsächlichen Bewerbungen und nicht durch die Weitergabe an private Kontakte zu prüfen.

Ergebnisse der Regressionen

Unseres Erachtens ist die Auswahl der Kontrollvariablen (Alter, Geschlecht, Ausbildungsniveau, Durchschnittsnote und Dauer der Tätigkeit) zur Überprüfung der Validität eines psychometrischen Persönlichkeitstests unglücklich. Hülshager et al. (2006) begründen die Auswahl lediglich mit der Beziehung zwischen demographischen Variablen sowie Aspekten des Humankapitals mit Berufserfolgskriterien. Insbesondere bei den Variablen Ausbildungsniveau und der durchschnittlichen Ausbildungsnote ist jedoch davon auszugehen, dass viele der zu erwartenden Zusammenhänge durch die Verwendung dieser Kontrollvariablen verdeckt werden. So ist plausibel anzunehmen, dass Persönlichkeitsmerkmale (z. B. Leistungsmotivation) den Ausbildungs-/Studienerfolg beeinflussen. Dies wiederum wird sich schließlich auf den Berufserfolg auswirken, daher erscheint eine Verwendung der durchschnittlichen Ausbildungsnote als Kriteriumsvariable wie bei Hossiep und Paschen (2003) sinnvoller. Ausbildungsniveau und Note wären plausible vermittelnde Variablen und sollten als solche nicht kontrolliert werden.

Die höchste Einzelkorrelation mit dem Entgelt weist die BIP-Skala Führungsmotivation auf ($r = .33^{**}$; Hossiep & Paschen, 2003). Das ist plausibel, wenn man annimmt, dass Personen, die führen wollen (also eine hohe Führungsmotivation aufweisen) auch später eine Führungsposition innehaben und damit verbunden mehr Geld verdienen. In der Studie von Hülshager et al. (2006) beträgt die Einzelkorrelation zwischen Führungsmotivation und Bruttoeinkommen $r = .35^{**}$ (S. 142).

Im Rahmen der Regressionsanalyse wird zur Führungsmotivation später ein (nicht signifikantes) Betagewicht von $\beta = .04$ berichtet (Hülshager et al., 2006). Wenngleich bei der Interpretation von Betagewichten stets Vorsicht geboten sein muss (Field, 2013), drängt sich die Schlussfolgerung auf, dass der eigentlich zu erwartende Beitrag der Führungsmotivation durch das Design der

Regression, der problematischen Operationalisierung (vgl. Gehaltskategorien) und der Stichprobe nicht zutage tritt. Datenbasis und Operationalisierung geben dies schlicht nicht her. Führungsmotivation als Skala ist im NEO-PI-R in dieser Form nicht enthalten. Dies ist also einer der Aspekte, in dem das BIP dem NEO-PI-R in der Kriteriumsvalidität überlegen sein sollte.

Vorsicht ist nach unserer Einschätzung auch bei der Interpretation weiterer Betagewichte geboten. So schreiben Hülshager et al. (2006): „Darüber hinaus zeigte sich, dass Personen mit geringer Handlungsorientierung tendenziell einen höheren Berufsstatus erreichten als Personen, die hoch handlungsorientiert waren“ (S. 145). Diese Interpretation ist nicht nur kontraintuitiv, sondern sachlich unrichtig. Vielmehr existiert – die Gültigkeit der z. T. fragwürdigen Ergebnisse von Hülshager et al. (2006) vorausgesetzt – lediglich ein linear negativer Zusammenhang unter der Voraussetzung, dass die weiteren Variablen des Modells konstant gehalten werden (Field, 2013). Schäfer (2016) betont, dass „die isolierte Vorhersagekraft eines Prädiktors abhängig von allen anderen Prädiktoren beurteilt wird“ (S.193), sodass das Betagewicht die von den anderen Prädiktoren bereinigte Vorhersagekraft ist.

Hossiep und Paschen (2003) fanden positive Betagewichte der Handlungsorientierung auf verschiedenste Erfolgsmaße im Studium, bei denen zu erwarten ist, dass sie wiederum mit Erfolgsmaßen des Berufslebens in positiver Beziehung stehen. Die mögliche Kausalkette zwischen Handlungsorientierung über Noten zu beruflichen Erfolgskriterien wird, wie oben begründet, durch die Verwendung der Note als Kontrollvariable in der Regression nicht mehr sichtbar, da die Handlungsorientierung um die Note bereinigt wurde.

Für den Vergleich der Verfahren werden die korrigierten Bestimmtheitsmaße herangezogen, die „das Verhältnis zwischen Stichprobengröße und Zahl der Regressoren berücksichtigen“ (Hülshager et al., 2006, S. 139). Dieses Verhältnis ist insbesondere beim BIP mit 14 Skalen bei einer für die Regressionsanalysen effektiven Stichprobengröße von $N = 77$ als schlecht anzusehen, worin auch die Ursache zu sehen ist, dass „sich [bei der Erklärung des Bruttoeinkommens] sogar das korrigierte Bestimmtheitsmaß durch das Hinzunehmen der BIP-Dimensionen [verringerte]“ (Hülshager et al., 2006, S. 139). Spätestens an dieser Stelle hätte den Autoren evident werden müssen, dass eine Regression mit dieser hohen Anzahl an Regressoren und dieser geringen Stichprobengröße wenig belastbare Aussagen treffen kann.

Zuletzt erstaunt es, dass einerseits ausgewählte Ergebnisse der Regression der Facetten des NEO-PI-R diskutiert werden, die Betagewichte jedoch im Ergebnisteil nicht berichtet werden.

Alles in allem ziehen wir die Aussagekraft der durchgeführten regressionsanalytischen Berechnungen in Zweifel, vor allem aufgrund der inadäquaten Wahl der Kontrollvariablen, der fragwürdigen Interpretation der Betagewichte sowie des unausgewogenen Verhältnisses von Regressoren und Stichprobengröße.

Zusammenfassung

Die von Hülshager et al. (2006) untersuchte Fragestellung der Kriteriumsvalidität und der Akzeptanz eines berufsbezogenen und eines nicht-berufsbezogenen Verfahrens ist für die Organisationspsychologie eine höchst relevante, wurde jedoch bedauerlicherweise mangelhaft umgesetzt. Die geringe und offenbar nicht das gesamte Spektrum beruflicher Erfolgskriterien abdeckende Stichprobe, die Operationalisierung des Berufsstatus in vier unscharf definierte Gruppen, die Zusammenfassung der Akzeptanzitems in nur einen Faktor und die Interpretation der Regressionen sind zentrale methodische Kritikpunkte. Der Ablauf außerhalb echter Auswahl-situationen, der das Ausfüllen des BIP nach dem NEO-PI-R und damit eine kaum zumutbare Itemanzahl von 450 beinhaltete, kann für die nicht erhöhte Akzeptanz des BIP verantwortlich zeichnen. Weiterhin dürften auch die Wahl der Kontrollvariablen und das unzureichende Verhältnis des Stichprobenumfangs zu der Anzahl der Regressoren essentielle Zusammenhänge verdecken.

Alternativen zum Vorgehen

Wir sehen durchaus Alternativen zum Vorgehen bei Hülshager et al. (2006). Die Studie berührt thematisch einerseits das Thema Konstruktvalidität, da zwei Testverfahren eingesetzt wurden. Dann geht es ganz explizit auch um die Kriteriumsvalidität (z. B. Entgelt als objektives Kriterium). Die von den Autoren vorgenommenen Regressionsanalysen, in denen schlicht alle Skalen beider Verfahren jeweils eingeschlossen wurden, halten wir zur Klärung der Fragestellung für nicht zielführend, da beide Verfahren inhaltliche Schnittmengen haben (z. B. Extraversion im NEO-PI-R und Kontaktfähigkeit im BIP oder auch Neurotizismus im NEO-PI-R und emotionale Stabilität im BIP).

Um den Datensatz entsprechend zu analysieren, würden wir folgendes Vorgehen empfehlen: Ein früher Schritt sollte eine gemeinsame Faktorenanalyse über alle NEO-PI-R- und BIP-Skalen sein, um inhaltliche Schnittmengen der beiden Verfahren zu identifizieren. Dieses Vorgehen ist bei Studien zur Konstruktvalidität üblich (vgl. Frieg & Schulz, 2013; Ostendorf & Angleitner, 2004).

So können die inhaltlichen Aspekte identifiziert werden, die beide Verfahren messen oder die nur jeweils durch eines der Verfahren abgedeckt werden. Sind auf diese Weise Gemeinsamkeiten und Unterschiede identifiziert, so können im nächsten Schritt die entstandenen Faktoren mit Blick auf die Zusammenhänge zu einem differenzierungsfähigen Kriterium (z. B. berufliches Entgelt) analysiert werden.

Literatur

- Borkenau, P., Friedel, A., & Wolfradt, U. (2011). Standardisierte Persönlichkeitsfragebögen. In L. F. Hornke, M. Amelang, & M. Kersting (Hrsg.), *Persönlichkeitsdiagnostik* (S. 1-72). Göttingen: Hogrefe.
- Field, A. (2013). *Discovering Statistics using IBM SPSS Statistics*. London: Sage Publications.
- Frieg, P., Pohlmann, L. & Hossiep, R. (2014). *Gütekriterien der Forschungsversion des BIP - 6F* (Forschungsbericht). Bochum: Ruhr-Universität, Projektteam Testentwicklung.
- Frieg, P. & Schulz, R. (2013). *Persönlichkeitsmessung – allgemein und/oder berufsbezogen? Eine Untersuchung zur Konstruktvalidität von PRF und BIP* (Forschungsbericht). Bochum: Ruhr-Universität, Projektteam Testentwicklung.
- Hossiep, R. & Krüger, C. (2012). *Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung - 6 Faktoren (BIP-6F)*. Göttingen: Hogrefe.
- Hossiep, R., Mutwill, A. & Schulz, R. (2012). „Das Geheimnis meines Erfolgs...“ - Berufserfolgsattributionen im Top-Management. In Arbeitskreis Assessment Center e.V. (Hrsg.), *Talent(e) entdecken und fördern: Personaldiagnostik als Wettbewerbsvorteil. Dokumentation zum 8. Deutschen Assessment-Center-Kongress 2012* (S. 426-444). Lengerich: Pabst Science Publishers.
- Hossiep, R. & Paschen, M. (2003). *Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung – BIP* (2. Aufl.). Göttingen: Hogrefe.
- Hossiep, R., Paschen, M., & Mühlhaus, O. (2000). *Persönlichkeitstests im Personalmanagement: Grundlagen, Instrumente und Anwendungen*. Göttingen: Hogrefe.

Hülshager, U. R., Specht, E. & Spinath, F. M. (2006). Validität des BIP und des NEO-PI-R. Wie geeignet sind ein berufsbezogener und ein nicht explizit berufsbezogener Persönlichkeitstest zur Erklärung von Berufserfolg? *Zeitschrift für Arbeits- und Organisationspsychologie*, 50(3), 135-147.

Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrick, M. R. (1999). The Big Five personality traits, general mental ability, and career success across the life span. *Personnel Psychology*, 52(3), 621-652.

Kauffeld, S., & Grohmann, A. (2011). Personalauswahl. In S. Kauffeld (Hrsg.), *Arbeits-, Organisations- und Personalpsychologie* (S. 93-112). Berlin: Springer.

Maties, M. & Wottawa, H. (2011). Nicht immer erlaubt: Wann Eignungstests künftig zulässig sind. *Wirtschaftspsychologie aktuell*, 18(4), 28-32.

Obermann, C. (2018). *Assessment Center* (6. Aufl.). Wiesbaden: Springer Gabler.

Ostendorf, F., & Angleitner, A. (2004). *NEO-Persönlichkeitsinventar nach Costa und McCrae, Revidierte Fassung*. Göttingen: Hogrefe.

Schäfer, T. (2016). *Methodenlehre und Statistik*. Wiesbaden: Springer.

Shaffer, J. A. & Postlethwaite, B. E. (2012). A matter of context: A meta-analytic investigation of the relative validity of contextualized and noncontextualized personality measures. *Personnel Psychology*, 65(3), 445-494.

Statistisches Bundesamt (2017). *Durchschnittliche Bruttomonatsverdienste*. Verfügbar unter: <https://www.destatis.de/DE/ZahlenFakten/GesamtwirtschaftUmwelt/VerdiensteArbeitskosten/VerdiensteVerdienstunterschiede/Tabellen/Bruttomonatsverdienste.htm>

Bitte zitieren Sie diesen Artikel wie folgt:

Merchel, R., Frieg, P. & Hossiep, R. (2018). *Anmerkungen zur Aussagekraft des Artikels „Validität des BIP und des NEO-PI-R“ (Hülshager, Specht & Spinath, 2006)* (Forschungsbericht). Bochum: Ruhr-Universität, Projektteam Testentwicklung.