

Der Nutzen onlinebasierter Vorauswahl – Validierung des DGP E2

Zusammenfassung

Um den Aufwand für die Auswahl geeigneter Bewerber/innen zu reduzieren, stehen Institutionen verschiedene Möglichkeiten zur Vorselektion zur Verfügung. Darunter fallen neben dem Anschreiben, Lebenslauf und Schulnoten der Bewerber/innen auch die Leistung in onlinebasierten Leistungstests. In der vorliegenden Studie wurde anhand eines realen Auswahlprozesses untersucht, inwiefern der onlinebasierte Leistungstest E2 und Schulnoten in der Lage sind, die Leistung in einem validen, schriftlichen Auswahlverfahren vorherzusagen. Es zeigt sich, dass der E2 die Leistung im schriftlichen Auswahlverfahren valide vorhersagt und damit Fehler bei der Auswahlentscheidung reduziert. Schulnoten sind hingegen als Prädiktor nicht geeignet. In Kombination mit einem hochvaliden Paper-Pencil-Test wie dem DGP M2 reduziert der E2 die Kosten und erhöht die Effizienz der Personalauswahl.

1 Personalauswahl als Hürdenlauf

Zur Auswahl geeigneter Bewerberinnen und Bewerber wird meist nicht nur ein einzelnes Auswahlinstrument herangezogen. Der Prozess gleicht vielmehr einem Hürdenlauf: Zunächst werden die Bewerbungsunterlagen auf Vollständigkeit und Einhaltung bestimmter (formaler) Mindestanforderungen (wie z. B. ein bestimmter Schulabschluss) geprüft. Nur die potentiell geeigneten Bewerberinnen und Bewerber werden anschließend, je nach Auswahlstrategie, zu einem Leistungstest, Assessmentcenter und/oder Vorstellungsgespräch eingeladen. Die Auswahl und Zusammensetzung der verschiedenen Stationen ist von einer Vielzahl von Faktoren abhängig. Besonders wichtig sind dabei Validität und Ökonomie. Während die Validität, also die Vorhersagekraft der eingesetzten Verfahren, möglichst hoch sein soll, sollte der Aufwand möglichst gering sein. Da besonders valide Verfahren, wie z. B. Assessment Center und strukturierte Interviews (Schmidt & Hunter, 1998), in der Regel zeit- und kostenintensiver sind als weniger valide Verfahren, findet in den frühen Auswahlphasen häufig eine Negativselektion statt, so dass die valideren Verfahren mit einem ausgedünnten Bewerberpool durchgeführt werden können.

Ein Vorauswahlinstrument, das sich in den letzten Jahren immer größerer Beliebtheit erfreut, ist der onlinebasierte Vorauswahltest. Es handelt sich dabei um Leistungs- und/oder Persönlichkeitstests, die von den Bewerberinnen und Bewerbern per Internet von Zuhause aus durchgeführt werden. Das Ziel onlinebasierter Vorauswahltests besteht darin, die für die ausgeschriebene(n) Stelle(n) ungeeigneten Bewerber/innen möglichst früh im Auswahlprozess zu identifizieren. Für die testende Institu-

tion bietet dies den Vorteil, dass nur die erfolgversprechenden Kandidaten/innen zum Leistungstest oder Assessmentcenter vor Ort eingeladen werden. Dies reduziert nicht nur die Kosten für die Durchführung der Auswahlverfahren, sondern auch den personellen Aufwand (die Sichtung der Bewerbungsunterlagen kann in der ersten Stufe entfallen) und den administrativen Aufwand (Reservierung unnötig großer Räumlichkeiten). Aber nicht nur die einstellenden Organisationen profitieren von der Vorauswahl im Internet: Bewerberinnen und Bewerber, die bereits an der ersten Hürde scheitern, ersparen sich sowohl die Anreise zu als auch die Teilnahme an weiteren Auswahlverfahren, wie z. B. mehrstündige schriftliche Leistungstests oder Assessment Center, die mit großer Wahrscheinlichkeit nicht zum gewünschten Ergebnis geführt hätten.

Eine in der Praxis weit verbreitete Alternative zur Vorselektion durch Persönlichkeits- oder Leistungstests besteht in der Nutzung von Schulnoten. Dies ist u. a. darauf zurückzuführen, dass sich Schulnoten ohne großen Aufwand erheben lassen und sie einen scheinbar objektiven Vergleich zwischen Bewerber/innen zulassen. Ein weiterer Grund könnte darin bestehen, dass insbesondere bei Ausbildungsplätzen und Einstiegspositionen wenige sonstige Kriterien vorliegen anhand derer sich die Eignung feststellen lässt. Trotz der starken Nutzung von Schulnoten im Selektionskontext weisen empirische Untersuchungen eher auf schwache Zusammenhänge zwischen Schulnoten und beruflicher Leistung hin (Roth, Be Vier, Switzer, & Schippman, 1996; Schuler, 2000).

Wie gut sich bestimmte Kriterien zur Vorselektion eignen, lässt sich empirisch überprüfen. Ist ein Kriterium geeignet, sollte es die Leistung der Bewerber/innen im jeweils nächsten Prozessschritt vorhersagen. Eine Person mit hoher Ausprägung auf diesem Kriterium (z. B. gute Leistung im Vortest, gute Noten etc.) sollte also im nächsten Auswahlschritt ebenfalls gut abschneiden. Umgekehrt sollten Personen mit niedriger Ausprägung auf dem gewählten Kriterium die nächste Auswahlhürde mit geringerer Wahrscheinlichkeit überstehen. Statistisch gesprochen erwarten wir also eine positive Korrelation zwischen den Kriterien in beiden Auswahlritten. Ziel dieser Studie ist es, diesen Zusammenhang für den Test „E2“ empirisch zu überprüfen und der Vorhersagekraft von Schulnoten vergleichend gegenüberzustellen.

2 Angaben zu den Verfahren

E2: Der E2 ist ein onlinebasierter Vorauswahltest, der die kognitive Leistungsfähigkeit der Bewerberinnen und Bewerber auf den drei Dimensionen numerische Verarbeitungskapazität, verbale Verarbeitungskapazität und Verarbeitungsgeschwindigkeit erfasst. Um den Test durchzuführen, erhalten die Bewerberinnen und Bewerber vom Auftraggeber Zugangsdaten per E-Mail und melden sich mit diesen innerhalb eines vorher festgelegten Zeitraums auf der Online-Testplattform der DGP e.V. an. Dort führen sie den Test selbständig unter Zeitbegrenzung durch. Nach Ablauf des Testzeitraums

erhält der Auftraggeber einen Ergebnisbericht, der die Testergebnisse der Teilnehmer/innen in den drei Teilbereichen enthält. Zusätzlich enthält der Bericht eine Empfehlung, ob der/die Bewerber/in zum schriftlichen (Präsenz-)Test eingeladen werden sollte oder nicht.

M2: Der *M2* ist ein schriftlicher Auswahltest zur Vorhersage beruflicher Eignung. Er sagt insbesondere die Eignung für solche Berufsbilder und Laufbahnen vorher, für die üblicherweise mittlere Bildungsabschlüsse vorausgesetzt werden, wie z. B. Verwaltungsfachangestellte und Bürokaufleute. Der *M2* prüft neben numerischer und verbaler Verarbeitungskapazität die Merkfähigkeit, Arbeitseffizienz, die Rechtschreibung sowie verschiedene anforderungsrelevante Kenntnisbereiche. Im Gegensatz zum *E2* wird der *M2* nicht computerbasiert, sondern als Paper-Pencil-Version unter Aufsicht durchgeführt.

3 Angaben zur Stichprobe

Für die ausgeschriebenen Stellen als Verwaltungsfachangestellte gingen insgesamt 624 Bewerbungen (69% Frauen, 31% Männer) ein. Die Bewerberinnen und Bewerber waren durchschnittlich 22,8 Jahre (Streuung: 5) alt. Von den 624 zum *E2*-Verfahren eingeladenen Bewerber/innen nahmen insgesamt 495 teil. Auf Basis der *E2*-Ergebnisse und der Schulnoten der Bewerber/innen wurden anschließend 417 Personen zum Auswahlverfahren vor Ort eingeladen. Die Verwendung der Schulnote als zweites Selektionskriterium führt dazu, dass auch Personen mit schlechten *E2*-Ergebnissen (aber guten Schulnoten) berücksichtigt wurden. Dies ist zwar für den Auswahlprozess unvorteilhaft, hat für die Validierungsstudie aber den großen Vorteil, dass die Prädiktorvarianz, also die Streuung der *E2*-Punktwerte, weniger eingeschränkt ist und sich Zusammenhänge somit besser darstellen lassen. Knapp 80% der Bewerber/innen (336) kamen der Einladung nach und bearbeiteten das schriftliche *M2*-Verfahren vor Ort.

Die Testleistungen der Stichprobe sind als durchschnittlich zu bezeichnen (vgl. Tabelle 1). So erreichten die Bewerberinnen und Bewerber durchschnittlich 97,3 Punkte (Streuung: 13) im *E2* und 100,9 Punkte (Streuung: 8,3) im *M2* und liegen damit nahe am Populationsmittelwert von 100 Punkten. Die Schulnoten zeichnen ein ähnliches Bild: Sowohl die Abschlussnoten als auch die Noten in den einzelnen Fächern weisen auf durchschnittliche Leistungen in der Gesamtstichprobe hin.

Tabelle 1: Deskriptive Statistiken der Testpunktwerte und Noten.

	Mittelwert	Streuung	Minimum	Maximum
1. E2 Punktwert	97,3	13,0	70,0	130,0
2. M2 Punktwert	100,9	8,3	78,0	124,0
3. Abschlussnote	8,8	1,7	4,7	14,0
4. Note Mathematik	7,3	2,8	2,0	15,0
5. Note Fremdsprache	7,9	2,5	2,0	14,0
6. Note Deutsch	8,3	2,3	2,0	14,0

Anmerkungen: Alle Noten sind in Punktwerten angegeben: 14 = sehr gut, 11 = gut, 8 = befriedigend, 5 = ausreichend, 2 = mangelhaft

4 Ergebnisse der Validierungsstudie

Zur Ermittlung der prognostischen Validität des E2-Verfahrens wurde der Korrelationskoeffizient für den Zusammenhang von E2-Punktwert und M2-Punktwert berechnet (vgl. Tabelle 2). Es ergibt sich eine Validität von $r = .51$ ($p < .01$). Personen, die im Vorauswahltest gut abschneiden, werden also mit großer Wahrscheinlichkeit auch im schriftlichen Test vor Ort gute Leistungen erbringen. Der umgekehrte Fall gilt gleichermaßen: Ein schlechtes Abschneiden im E2 geht in der Regel mit schlechten Leistungen im schriftlichen Haupttest einher. Die empirischen Ergebnisse attestieren dem E2 damit eine sehr gute prognostische Validität.

Die Korrelation zwischen der letzten Schulnote der Bewerber/innen und der Testleistung im M2 ergibt sich zu $r = .06$ (n.s.). Es besteht folglich kein statistisch bedeutsamer Zusammenhang zwischen Schulnote und Testleistung. Dieses Ergebnis zeigt sich konsistent über verschiedene Fächer hinweg. Die Noten in Mathematik, gewählter Fremdsprache und Deutsch korrelieren zwar stark untereinander, jedoch nicht mit der Testleistung im M2. Die Schulnoten der Bewerber/innen sind folglich als Vorauswahlinstrument für die schriftliche Eignungsuntersuchung ungeeignet.

Tabelle 2: Korrelationen zwischen den Testpunktwerten und Schulnoten.

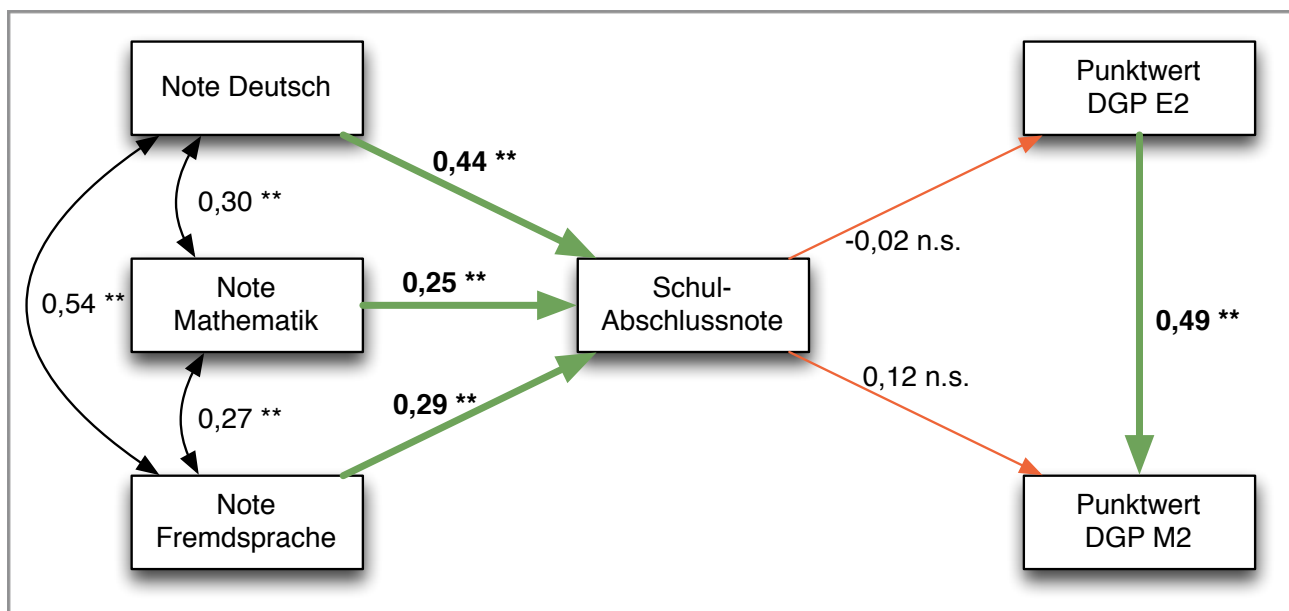
	1.	2.	3.	4.	5.	6.
1. E2 Punktwert	-	.513**	.064	.066	.089	-.019
2. M2 Punktwert		-	.075	.017	.049	-.020
3. Abschlussnote			-	.511**	.515**	.678**
4. Note Mathematik				-	.268**	.337**
5. Note Fremdsprache					-	.457**
6. Note Deutsch						-

Anmerkungen: ** $p < .01$

Um die Zusammenhänge zwischen den schulischen Leistungen und den Leistungen in den Auswahlverfahren zu verdeutlichen, wurde zusätzlich eine Pfadanalyse (Kline, 2011) durchgeführt (vgl. Abbildung 1). In Pfadanalysen können, wie in normalen Regressionsanalysen auch, mehrere Variablen simultan betrachtet werden. Beispielsweise könnte man annehmen, dass sich das Ergebnis des schriftlichen Auswahlverfahrens nicht nur aus dem Punktwert des Vortests, sondern zusätzlich aus der letzten Schulnote der Bewerber/innen vorhersagen lässt. Die Pfadanalyse bietet gegenüber der Regressionsanalyse den Vorteil, dass eine Variable (wie z. B. der E2-Punktwert) gleichzeitig als vorhergesagte und als vorhersagende Variable fungieren kann. Zum anderen lassen sich Zusammenhänge zwischen Variablen in der Pfadanalyse sehr anschaulich darstellen, was die Interpretation der Ergebnisse erleichtert.

Das Datenmuster der Korrelationsanalyse wird auch in der Pfadanalyse sichtbar. Es zeigt sich, dass die letzte Schulnote der Bewerber/innen weder die Leistung im Vorauswahltest, noch die Ergebnisse des schriftlichen M2-Verfahrens vorhersagt (rote Pfeile). Im Gegensatz dazu sagt der onlinebasierte Vorauswahltest E2 die Ergebnisse im schriftlichen Test zuverlässig vorher (grüner Pfeil).

Abbildung 1: Pfadanalyse zum Zusammenhang zwischen Schulnoten und Testleistungen.

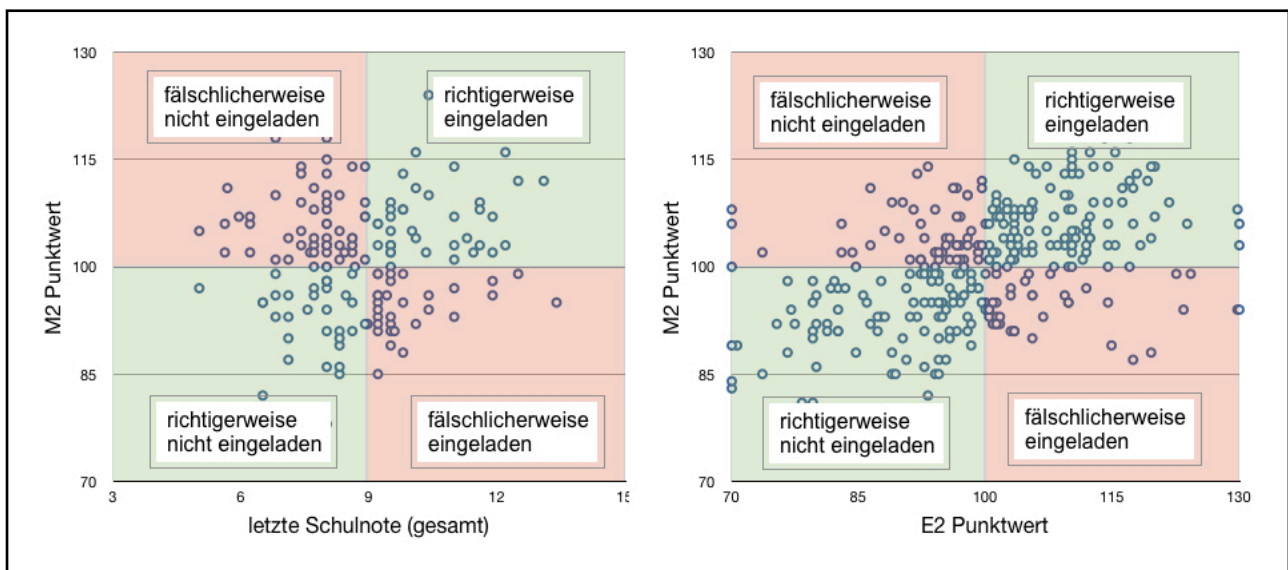


Anmerkungen: ** $p < .01$, n.s. = nicht signifikant, einköpfige Pfeile entsprechen beta-Gewichten linearer Regressionen, doppelköpfige Pfeile entsprechen Korrelationen, Modellstatistiken: $df = 6$, $p = 0.24$, CFI = 0.99, TLI = 0.98, RMSEA = 0.05

5 Diskussion

Die praktische Bedeutsamkeit dieser Ergebnisse wird bei Betrachtung der Streudiagramme deutlich. Abbildung 2 zeigt die Zusammenhänge von M2-Punktwert und Schulnote (linke Abbildung) und E2-Ergebnis (rechte Abbildung). Die vier Quadranten ergeben sich durch die Festlegung „kritischer“ Noten bzw. Punktwerte, die überschritten werden müssen, damit man zum schriftlichen Test eingeladen wird (im Beispiel Schulnote 3+ bzw. 100 Punkte im E2) in Kombination mit einem kritischen Wert, ab dem der M2 als „bestanden“ gilt (im Beispiel ebenfalls 100 Punkte). In der praktischen Anwendung sind die kritischen Werte von einer Vielzahl von Faktoren (wie z. B. der Zahl der Bewerber/innen und den Kosten einer falschen Auswahlentscheidung) abhängig und dienen hier nur der Veranschaulichung.

Abbildung 2: Zusammenhang von M2-Punktwert und Schulnote (links) und E2-Ergebnis (rechts).



Anmerkungen: Die kritischen Werte (Schulnote 3+, 100 Punkte in E2 und M2) dienen nur der Veranschaulichung.

Die Scatterplots zeigen, dass die Vorhersage einer zukünftigen Leistung immer fehlerbehaftet ist. Dabei können zwei verschiedene Arten von Fehlern auftreten: Man lädt Personen *nicht* zum schriftlichen Test ein, obwohl sie ihn bestanden hätten (Fehler 1. Art, Quadrant oben links), oder man lädt Personen zum schriftlichen Test ein, die diesen nicht bestehen würden (Fehler 2. Art, Quadrant unten rechts). Beide Fehler sind für die testende Institution mit Kosten verbunden: Werden Personen fälschlicherweise nicht eingeladen, verliert die Institution potentiell geeignete Bewerberinnen und Bewerber, was sich vor allem bei schwacher Bewerberlage negativ auf die Leistungs- und Wettbewerbsfähigkeit auswirken kann. Die Teilnahme ungeeigneter Bewerberinnen und Bewerber am schriftlichen Haupttest verursacht wiederum unnötige Kosten für Korrespondenz, Anreise und Testdurchführung. Es ist daher im Interesse der testenden Institution und der Bewerber/innen, beide Fehlerarten so gering wie möglich zu halten. Die Ergebnisse zeigen, dass dies nur mit validen Vorauswahlverfahren möglich ist: Im Vergleich zu den weniger validen Schulnoten reduziert der E2 – unter Annahme der exemplarisch gewählten kritischen Werte – den Fehler 1. Art in unserer Stichprobe um 25% und der Fehler 2. Art um 15%.

Eine Frage, die sich bei Testungen im Internet immer stellt, ist die der „Schummler“. Die Möglichkeiten der Täuschung sind vielfältig und reichen von der Nutzung unerlaubter Hilfsmittel (z. B. Wikipedia bei Wissensfragen oder Taschenrechner bei Mathematikaufgaben) bis hin zur gemeinsamen Testbearbeitung mit mehreren Personen vor dem Bildschirm. Obwohl die testende Institution grundsätzlich keine Möglichkeit hat, Schummeln zu verhindern, schränkt dies die Nützlichkeit in-

ternetbasierter Tests nicht ein: Erstens ist der E2 als reiner Vorauswahltest konzipiert, der **nur** in Kombination mit einem anschließenden schriftlichen Test unter Aufsicht durchgeführt wird. Dieses Vorgehen, das auch in führenden Fachzeitschriften empfohlen wird (Tippins et al., 2006), stellt sicher, dass Täuschungsmanöver spätestens im nächsten Auswahlschritt enttarnt werden. Zweitens wird den Teilnehmern/innen zu Beginn des Tests mitgeteilt, dass sie bei ausreichender Leistung in jedem Fall einen Test unter Aufsicht durchführen müssen, was die Motivation zum Schummeln einschränkt. Drittens sind die Aufgaben so gestellt, dass unerlaubte Hilfsmittel keinen großen Zusatznutzen bringen. Dies wird beispielsweise durch Textrechenaufgaben erreicht, bei denen die Leistung hauptsächlich darin besteht, den richtigen Rechenweg herauszufinden, und nicht im Ausrechnen des Ergebnisses. Zudem sind alle Aufgaben unter einem Zeitlimit zu bearbeiten, so dass für ausführliche Recherchen im Internet keine Zeit bleibt.

Offen bleibt die Frage, warum der Zusammenhang zwischen Schulnoten und Ergebnis im schriftlichen Haupttest so schwach ausfällt. Eine mögliche Ursache liegt darin, dass Schulnoten nachweislich wenig objektiv, reliabel und valide sind. So können die gleichen Leistungen in Abhängigkeit vom bewertenden Lehrer (Birkel & Birkel, 2002), vom Klassendurchschnitt (Ingenkamp, 1995), sowie von Schultyp und Bundesland (Klieme, 2003) unterschiedlich bewertet werden. Es ist auch bekannt, dass Mädchen im Durchschnitt bessere Schulnoten erhalten als Jungen (Helbig, 2012), und zwar unabhängig von ihrer kognitiven Leistungsfähigkeit (Blossfeld et al., 2009). Eine weitere mögliche Ursache besteht darin, dass Schulnoten andere Kompetenzen abbilden als die, die für das Bestehen eines Eignungstests wie dem M2 notwendig sind. Während im M2 hauptsächlich die kognitive Leistungsfähigkeit der Bewerber/innen überprüft wird, fließen in Schulnoten auch Verhaltensaspekte wie aktive Teilnahme am Unterricht, adäquates Sozialverhalten oder Fehlzeiten mit ein, die wenig mit der kognitiven Leistungsfähigkeit zu tun haben.

6 Fazit

Um den Personalauswahlprozess auch bei großen Bewerberzahlen ökonomisch zu gestalten ohne Validität einzubüßen, ist eine Vorselektion der Bewerber/innen unumgänglich. In einer empirischen Studie mit 624 Personen wurde deutlich, dass der E2 die Leistung im schriftlichen Haupttest zuverlässig vorhersagt und sich damit als Vorauswahlkriterium in mehrstufigen Auswahlverfahren eignet. Schulnoten stellten sich demgegenüber als äußerst schlechte Prädiktoren der Testleistungen heraus, was sich zum einen auf ihre schwache Messgüte und zum anderen auf ihre zur Vorhersage ungeeigneten Messinhalte zurückführen lässt. Gemeinsam mit dem nachweislich validen M2-Verfahren (Ristel & Haarhaus, 2013; in dieser Ausgabe) ist der E2 in der Lage, die Kosten von Personalauswahlprozessen zu reduzieren und gleichzeitig ihre Effizienz zu erhöhen.

Literatur

- Birkel, P., & Birkel, C. (2002). Wie einig sind sich Lehrer bei der Aufsatzbeurteilung? Eine Replikationsstudie zur Untersuchung von Rudolf Weiss. *Psychologie in Erziehung und Unterricht*, 49(3), 219-224.
- Blossfeld, H.-P., Bos, W., Hannover, B., Lenzen, D., Müller-Böling, D., & Prenzel, M. (2009). *Geschlechterdifferenzen im Bildungssystem. Jahresgutachten 2009 Aktionsrat Bildung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Helbig, M. (2012). Warum bekommen Jungen schlechtere Schulnoten als Mädchen? Ein sozialpsychologischer Erklärungsansatz. *Zeitschrift für Bildungsforschung*, 2(1), 41-54.
- Ingenkamp, K. (Ed.). (1995). *Die Fragwürdigkeit der Zensurengebung*. Weinheim: Beltz.
- Klieme, E. (2003). Benotungsmaßstäbe an Schulen: Pädagogische Praxis und institutionelle Bedingungen. Eine empirische Analyse auf der Basis der PISA-Studie. In H. Döbert, B. von Kopp, R. Martini & M. Weiß (Eds.), *Bildung vor neuen Herausforderungen* (pp. 195-211). Neuwied: Luchterhand.
- Kline, R. (2011). *Principles and Practice of Structural Equation Modeling*. New York, London: The Guilford Press.
- Ristel, N., & Haarhaus, B. (2013). Bewährungskontrolle des M2-Verfahrens – Die größte Stichprobe der DGP-Geschichte. *DGP-Informationen*, 54(63), x-y.
- Roth, P. L., Be Vier, C. A., Switzer, F. S., & Schippman, J. S. (1996). Meta-analyzing the relationship between grades and job performance. *Journal of Applied Psychology*, 81(5), 548-556.
- Schmidt, F., & Hunter, J. (1998). The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings. *Psychological Bulletin*, 124(2), 262-274.
- Schuler, H. (2000). Noten und Studien- und Berufserfolg. In D. Rost (Ed.), *Handwörterbuch Pädagogische Psychologie*, 4. Aufl. (pp. 599-606). Weinheim: Psychologie Verlags Union.
- Tippins, N. T., Beaty, J., Drasgow, F., Gibson, W. M., Pearlman, K., Segall, D. O., & Shepherd, W. (2006). Unproctored Internet Testing in Employment Setting. *Personnel Psychology*, 59(1), 189-225.

Korrespondenzanschriften des Autors:

Dipl.-Psych. Benjamin Haarhaus
Deutsche Gesellschaft für Personalwesen e.V.
Hohenzollernstr. 11-13
40211 Düsseldorf
haarhaus@dgp.de