

# Die Vorhersage des Ausbildungserfolgs von Verwaltungsfachangestellten und kaufmännischen Berufen – Bewährungskontrolle der DGP-Testbatterie M2

Nina Ristel/Benjamin Haarhaus

## Zusammenfassung

Die DGP-Testbatterie M2 wurde anhand einer umfangreichen Stichprobe (N = 441) validiert. Es zeigen sich signifikante Zusammenhänge zwischen der Leistung im M2 und den Ergebnissen der Zwischenprüfung ( $r = .48$ ) sowie der Abschlussprüfung ( $r = .46$ ) der jeweiligen Ausbildungen. Dieses Ergebnis bestätigt, dass die Testbatterie M2 in der Lage ist, den Ausbildungserfolg valide vorherzusagen.

## 1 Bewährungskontrollen – Eine Tradition der DGP e.V.

Die erste Bewährungskontrolle der DGP wurde im Jahr 1960 veröffentlicht. Unter dem Titel „Zum prognostischen Wert psychologischer Eignungsuntersuchungen“ untersuchte Adolf Otto Jäger anhand von 45 Teilnehmern/innen eines REFA-Jahrganges (Grundlehrgang des Verbandes für Arbeitsstudien e.V.), wie gut sich die Prüfungsleistung aus dem DGP-Testergebnis vorhersagen lässt. Herausragend hoch ( $r = .80$ ) fiel der Zusammenhang zwischen dem DGP-Empfehlungsgrad und der nach einem halben Jahr abgelegten Prüfung aus. Dieses kaum zu überbietende Ergebnis war dem besonderen Umstand geschuldet, dass es sich nur um einen halbjährigen Beobachtungszeitraum handelte. Hinzu kommt, dass alle getesteten Teilnehmer/innen nach einem halben Jahr den Grundlehrgang abschlossen und somit die gesamte Stichprobe erhalten blieb. Jäger schließt seinen Artikel mit den Worten „Die Dringlichkeit weiterer Bewährungskontrollen und ihrer Veröffentlichung wurde hervorgehoben. [...] Es wurde angeregt auch bescheidenere Bewährungsergebnisse psychologischer Eignungsuntersuchungen mitzuteilen und den Voraussagewert [...] zu ermitteln.“

Seinen Empfehlungen sollte in den nächsten Jahren mehrfach gefolgt werden: Insgesamt wurden bis heute 32 Bewährungskontrollen durch die DGP e.V. durchgeführt. Damit ist die DGP im deutschsprachigen Raum eine der Institutionen mit den meisten veröffentlichten Bewährungskontrollen zur Personalauswahl. Sowohl die Stichprobengrößen als auch die Ergebnisse aller bisherigen Bewährungskontrollen unterlagen dabei großen Schwankungen. Um die Vielfalt in Erinnerung zu rufen, wird an einigen Beispielen der Versuch eines Schnelldurchlaufes durch die Geschichte dieser Untersuchungen gewagt.

Eine für den damaligen Zeitpunkt sehr große Stichprobe wurde von Klaus Althoff (1968), dem langjährigen Leiter der DGP, untersucht. Es handelte sich um 104 Regierungsinspektoren und 96 Revierförster. Die durchschnittliche Korrelation zwischen Empfehlungsgrad und Ausbildungserfolg betrug  $r = .40$ . Der gleiche Autor stellte bei einer speziellen Berufsgruppe (Polizeibeamtinnen und -beamten) im Jahre 1975 weiterhin eine herausragend gute Vorhersagekraft von  $r = .75$  des damaligen DGP-Verfahrens fest und bewies damit eine deutliche Überlegenheit gegenüber vorher genutzten internen Kriterien der Polizei. Eine kleinere Stichprobe wurde im Bereich der Krankenpflege von Michael Thielepape (1980) untersucht. Bei 42 Personen konnten Zusammenhänge des DGP-Verfahrens zum Ausbildungserfolg hergestellt

werden.

Die erste Bewährungskontrolle, in der explizit der mittlere Dienst untersucht wurde, stammt von Gisela Seggebruch (1982). Sie konnte besonders deutliche Zusammenhänge des Ausbildungserfolgs mit den Testkriterien verbale Kompetenz, Gedächtnisleistung, politisch-wirtschaftliche Bildung und dem damals so genannten „komplexeren bürotechnischen Arbeitsverhalten“ feststellen. Diese Untersuchung war darüber hinaus Grundlage dafür, auch im Verfahren für den mittleren Dienst Gruppenstationen in das mündliche Verfahren zu implementieren, da erstmals statistisch nachgewiesen wurde, wie hoch der Zusammenhang der persönlichen Einschätzung der DGP-Psychologinnen und -Psychologen und den Prüfungskriterien auch für den mittleren Dienst ausfiel ( $r = .59$ ).

Eine weitere Untersuchung, die sich ebenfalls auf das DGP-Verfahren für den mittleren Dienst bezieht, ist die zum Ausbildungserfolg des Angestelltenlehrgangs I (Bretz & Oldendörp, 1992). Es handelte sich zwar um das gleiche DGP-Verfahren wie für den mittleren Dienst, jedoch um eine gänzlich andere Zielgruppe, da die Aufstiegsbewerber wesentlich älter (Durchschnittsalter hier 28 Jahre) und schon länger im Verwaltungskontext tätig waren. Es wurden Daten von Aufstiegsbewerbern/innen über einen Zeitraum von sieben Jahren erfasst, die am Angestelltenlehrgang I teilgenommen hatten ( $N = 80$ ). Auch bei dieser Zielgruppe zeigte sich ein Zusammenhang von  $r = .48$  zwischen DGP-Empfehlungsgrad und der Abschlussprüfungsnote.

In den folgenden Jahren (z. B. Weber & Werner, 1983, Biernat & Scholz, 1995, Thielepape, 2003 sowie Rieke, Brendler & Knorz, 2007) wurden wiederholt der mittlere und der gehobene Dienst unter die Lupe genommen, ebenso wie der feuerwehrtechnische Dienst, die Krankenpflege, der Polizeivollzugsdienst sowie gewerbliche Berufe wie z. B. Gärtner/innen, Tierpfleger/innen, Fachkräfte für Abwassertechnik, Industriemechaniker/innen und einige mehr.

Es zeigt sich somit über die dokumentierten 53 Jahre der Erfahrungen mit Bewährungskontrollen in der DGP, dass diese regelmäßig zur Überprüfung der Qualität der verwendeten Personalauswahlverfahren für alle Berufsgruppen beigetragen haben. Die von der DGP eingesetzten Tests und Untersuchungsinstrumente wurden und werden auch in Zukunft immer wieder zum Ausbildungserfolg in Beziehung gesetzt, um durch statistische Analysen festzustellen, ob die Eignungsuntersuchungen der DGP eine relevante Prognose der Ausbildungsleistung erlauben und somit ihrem grundsätzlichen Zweck weiterhin entsprechen.

Die letzte Bewährungskontrolle (Haarhaus, Buchhester & Ristel, 2012) bezog sich auf die Daten von Personen, die für den gehobenen Dienst ausgewählt wurden und somit auf das A2-Verfahren der DGP. Um den eigenen Prinzipien treu zu bleiben, den Empfehlungen der DIN 33430 zu folgen und alle Verfahren regelmäßig zu überprüfen, folgt hiermit nun die Bewährungskontrolle für das aktuelle M2-Verfahren.

## **2 Empirische Bewährungskontrolle**

### **2.1 Angaben zum Verfahren**

Der schriftliche Leistungstest M2 überprüft die intellektuellen Fähigkeiten der Bewerber/innen mit den Schwerpunkten verbale und numerische Verarbeitungskapazität sowie die Merkfähigkeit. Zusätzlich zu diesen zentralen Elementen aus dem Berliner Intelligenzstrukturmodell werden die Arbeitseffizienz, die Rechtschreibung sowie verschiedene, anforderungsrelevante Kenntnisbereiche überprüft. Der M2 sagt

insbesondere die Eignung für solche Berufsbilder und Laufbahnen vorher, für die üblicherweise mittlere Bildungsabschlüsse vorausgesetzt werden, wie z. B. Verwaltungsfachangestellte und Bürokaufleute. Die Messgenauigkeit des Testbatterie beträgt  $\alpha = .94$ . Weitergehende Informationen zum Verfahren finden sich in den Verfahrenshinweisen, die über [www.dgp.de](http://www.dgp.de) frei verfügbar sind.

## 2.2 Datenerhebung und Methodik

Um eine möglichst große und repräsentative Stichprobe zu gewinnen, wurden ab April 2013 alle Kunden, die in den Jahren 2007 bis 2010 mehr als 15 Bewerberinnen und Bewerber mit dem M2-Verfahren getestet hatten, per E-Mail kontaktiert. In diesem Anschreiben befand sich eine vorgefertigte Tabelle, in die die Kunden die damaligen Leistungen im DGP-Test sowie die Abschlussnoten der Laufbahnprüfungen eintragen sollten. Die ausgefüllten Tabellen wurden (selbstverständlich unter Beachtung von Datenschutzvorgaben) anschließend per E-Mail oder per Post an die DGP zurückgesandt. Ebenso nutzten einige Kunden das Angebot, die Daten telefonisch zu übermitteln. Die Datensammlung wurde im Juni 2013 beendet, da so sichergestellt werden konnte, dass auch die Daten aus den letzten Prüfungsphasen berücksichtigt werden konnten.

Um die unterschiedlich gehandhabte Notengebung zu standardisieren, wurden zunächst alle Prüfungsergebnisse in Punktwerte umgerechnet. Zur Erhebung des Zusammenhangs zwischen DGP-Testergebnis und Prüfungsleistung in der Ausbildung zu untersuchen, wurden anschließend Korrelationskoeffizienten berechnet. Korrelationskoeffizienten liegen zwischen  $r = -1$  (absolut negativer Zusammenhang) und  $r = +1$  (absolut positiver Zusammenhang). Ein Korrelationskoeffizient von  $r = 0$  bedeutet, dass kein Zusammenhang zwischen Prädiktor und Kriterium besteht.

## 2.3 Stichprobe

Insgesamt konnten 441 Datensätze von 40 verschiedenen Verwaltungen und Institutionen gesammelt werden, was nach Sichtung aller bisherigen Bewährungskontrollen eine der größten Stichproben für eine Bewährungskontrolle in der Geschichte der DGP ist. Da 37 der ausgewählten Bewerber/innen die Ausbildung nicht abgeschlossen hatten, gingen die Daten von insgesamt 404 Personen in die Untersuchung ein. Weil bei einigen Kunden zudem keine vollständige Rekapitulation aller Daten möglich war, schwankt der Stichprobenumfang je nach betrachteten Variablen zwischen 384 und 400 Personen (vgl. Tabelle 1).

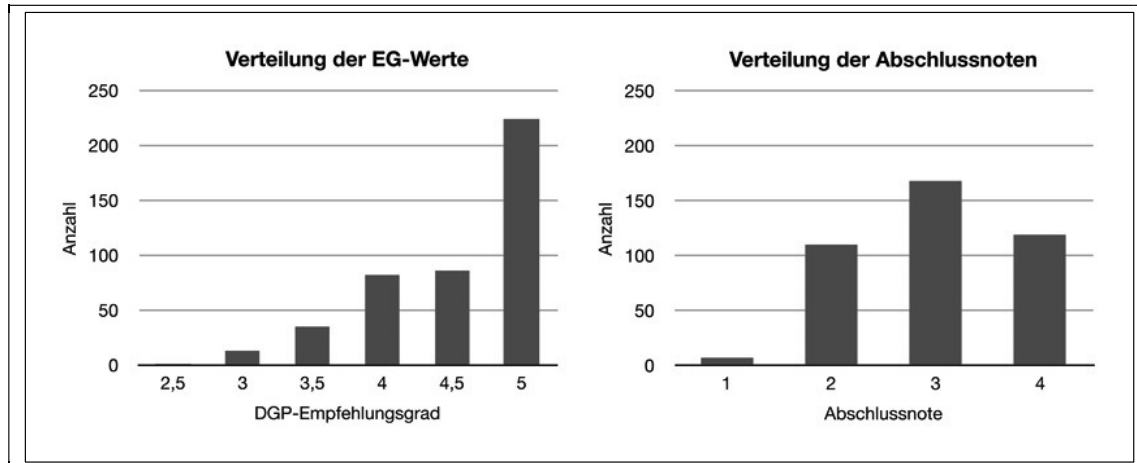
*Tabelle 1:* Deskriptive Statistiken der Prädiktoren und Kriterien

	N	Mittelwert	Streuung	Minimum	Maximum
M2 Empfehlungsgrad	441	4,53	0,57	2,50	5,00
M2 Gesamtpunkt看	384	107,64	4,93	98,00	125,00
Punktwert der Zwischenprüfung	400	9,80	2,54	2,00	14,50
Punktwert der Abschlussprüfung	404	10,29	2,07	5,60	14,60

Da in der Regel nur Personen mit guten Testleistungen zur Ausbildung zugelassen werden, ist die Verteilung der Testpunktwerte stark asymmetrisch (vgl. Abbildung 1). Die Testleistungen der Bewerberinnen und Bewerber liegen mit 107,64 Punkten

deutlich über dem Durchschnitt von 100 Punkten. Auch der durchschnittliche Empfehlungsgrad von 4,53 verdeutlicht, dass hauptsächlich Bewerberinnen und Bewerber ausgewählt wurden, die für die jeweiligen Ausbildungen geeignet sind.

Abbildung 1: Verteilung der EG-Werte und Abschlussnoten



## 2.4 Ergebnisse

Wenn das Auswahlverfahren das misst, was es messen soll, erwarten wir einen positiven Zusammenhang zwischen der Leistung im DGP-Test (Prädiktor) und in der Ausbildung (Kriterium). Tabelle 2 gibt die Ergebnisse der Korrelationsanalyse wieder. Es zeigt sich, dass alle Korrelationen zwischen den Prädiktoren und Kriterien statistisch signifikant sind ( $p < .01$ ). Aufgrund der abgeschnittenen Verteilung, werden die wahren Zusammenhänge jedoch unterschätzt (Stelzl, 2005). Aus diesem Grund wurden alle Korrelationen gegen Einschränkung der Prädiktorvarianz korrigiert (Wiberg & Sundström, 2009). Die prognostischen Validitäten des DGP-Empfehlungsgrads und Gesamtpunktwerts liegen zwischen  $r = .46$  und  $r = .48$ .

Tabelle 2: Zusammenhang zwischen Vorhersagemaßen (Empfehlungsgrad / Gesamtpunktwert M2) und Ausbildungserfolg (Abschlussnote / Note der Zwischenprüfung)

Prädiktor	Kriterium	N	Validität (unkorrigiert)	Validität (korrigiert) <sup>1</sup>
Empfehlungsgrad	Abschlussnote	404	$r = .20^{**}$	0,46
Empfehlungsgrad	Zwischenprüfung	400	$r = .21^{**}$	0,47
Gesamtpunktwert	Abschlussnote	348	$r = .29^{**}$	0,46
Gesamtpunktwert	Zwischenprüfung	343	$r = .31^{**}$	0,48

Anmerkungen: <sup>\*\*</sup> Die Korrelation ist auf dem Niveau von  $\alpha = .01$  statistisch signifikant. <sup>1</sup> Korrektur gegen Einschränkung der Prädiktorvarianz nach Thorndike case 2 (Wiberg & Sundström, 2009).

## 3 Diskussion und Ausblick

Insgesamt 441 Personen von 40 Institutionen konnten in diese Untersuchung mit einbezogen werden. Es zeigen sich hoch signifikante Zusammenhänge zwischen dem DGP-Empfehlungsgrad und den jeweiligen Prüfungsnoten (Zwischen- und Abschlussprüfung). Die Prognostischen Validitäten zwischen  $r = .46$  und  $r = .48$  belegen, dass

der M2 für Verwaltungsfachangestellte und kaufmännische Berufsgruppen ein valides Auswahlinstrument ist, das DIN-gerecht den Ausbildungserfolg vorhersagt. Dieses Ergebnis reiht sich gut in die Historie früherer Bewährungskontrollen ein, die überdurchschnittlich häufig Validitäten im Bereich von  $r = .30$  bis  $r = .60$  ergaben. Es wird erneut deutlich, dass auch das aktuelle M2-Verfahren allen Anforderungen entspricht und vorhersagt, was es vorhersagen soll – nämlich den Ausbildungserfolg der Bewerberinnen und Bewerber.

Trotz der DGP-seitigen Zufriedenheit mit diesem Ergebnis, besteht weiterhin der Ehrgeiz, alle Verfahren ständig zu aktualisieren und zu optimieren. Nachdem im vergangenen Jahr der A3 und in 2013 der K3 als überarbeitete Versionen der Vorgänger A2 und K2 eingeführt wurden und neben neuen Inhalten auch eine neue, noch übersichtlichere Optik besitzen, ist es nach dieser Bewährungskontrolle das Ziel der DGP, auch den M2 zu aktualisieren und durch ein zeitgemäßes Design auch optisch in neuem Glanz erscheinen zu lassen.

## Literatur

Althoff, K. (1968). Bericht über Bewährungskontrollen der Eignungsuntersuchungen im Land Niedersachsen, DGP-Information, 07/1968 (Heft 4)

Althoff, K. (1975) Die Vorhersage des Ausbildungserfolges von Aufstiegsbeamten der Kriminalpolizei. DGP-Information, 04/1975 (Heft14).

Biernat, J. & Scholz, B. (1995). Eine Bewährungskontrolle des DGP-Auswahlverfahrens für den mittleren feuerwehrtechnischen Dienst. DGP-Information, 12/1995 (Heft 54).

Bretz, E. & Oldendörp, H. (1992). Bewährungskontrolle: Vorhersage des Ausbildungserfolges im Angestellten-Lehrgang I. DGP-Information, 03/1992 (Heft 51).

DIN 33430 (2002): Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen. Berlin: Beuth.

Haarhaus, B., Buchhester, S., & Ristel, N. (2012). Bewährungskontrolle des A2-Verfahrens: Hintergründe, Ergebnisse und Konsequenzen. DGP-Informationen, 53(62), 5-10.

Jäger, A. (1960). Zum prognostischen Wert psychologischer Eignungsuntersuchungen – Eine Bewährungsstudie. DGP-Informationen, 08/1960.

Rieke, J., Brendler, K. & Knorz, C. (2007). Auswahl von Bewerbern/innen in den gewerblichen Berufen: Eine Bewährungskontrolle für den Test „Gewerbliche Ausbildung“ (GA) . DGP-Informationen 08/2007. (Heft 59)

Seggebruch, G. (1982). Bewährungskontrolle I – Vorhersage des Berufserfolges von Anwärtern des mittleren Dienstes einer Kommunalverwaltung. DGP-Informationen, 12/1982 (Heft 42).

Stelzl, I. (2005). Fehler und Fallen der Statistik: für Psychologen, Pädagogen und Sozialwissenschaftler. Münster: Waxmann.

Thielepape, M. (1980). Untersuchung über die Vorhersage des Krankenpflegeerfolges von Krankenschwestern und -pflegern – Eine Bewährungskontrolle. DGP-Information, 07/1980 (Heft 18).

Thielepape, M. (2003). Ergebnisse einer Bewährungskontrolle der DGP-Eignungstests für eine große Krankenpflegeschule in Niedersachsen. DGP-Information, 04/2003 (Heft 57).

Weber, W. & Werner, R.R. (1983). Auswahlverfahren und Prüfungsleistungen bei Anwärtern des mittleren und gehobenen nichttechnischen Dienstes – Bericht über zwei Bewährungskontrollen. DGP-Information, 12/1984 (Heft 44).

Wiberg, M., & Sundström, A. (2009). A comparison of two approaches to correction of restriction of range in correlation analysis. Practical Assessment, Research & Evaluation, 14(5), 2009.

#### **Korrespondenzanschriften der Autoren:**

Dipl.-Psych. Nina Ristel  
Deutsche Gesellschaft für Personalwesen e.V.  
Stammestr. 40 D  
30459 Hannover  
ristel@dgp.de

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