

# **Evaluation und Qualitätsoptimierung in der Personalauswahl: Zwei Bewährungskontrollen**

*Michael Thielepape und Dr. Martin Kersting<sup>1</sup>*

Grundvoraussetzung für den Einsatz diagnostischer Personalauswahlinstrumente ist der Nachweis ihrer Qualität, eine Maxime, welche die DGP sich seit Beginn ihres Bestehens auf ihre Fahnen geschrieben hat, und eine Forderung, die seit der Installation der Qualitätsnorm DIN 33430 im Juni 2002 erneut an Aktualität gewonnen hat. Hossiep (1995, S. 75f.) beklagt die geringe Anzahl an empirischen Erfolgskontrollen von Personalauswahlverfahren und nennt in diesem Kontext die DGP als „rühmliche Ausnahme“. Es gab mehrere Gründe, wieder eine Evaluation durchzuführen:

- Die DGP unterzieht im Auftrag der öffentlichen Verwaltungen jedes Jahr ca. 20.000 Bewerberinnen und Bewerber einem Eignungstest, davon ca. 8.000 für die Laufbahn des gehobenen Dienstes. D.h. die prognostische Qualität der Auswahlinstrumente hat erhebliche Relevanz sowohl für die Personalbeschaffung und letztlich Qualifikation der Auszubildenden, Fachhochschulstudenten/innen und Mitarbeiter/innen unserer Kunden als auch für die berufliche Entscheidung und Entwicklung der Bewerberinnen und Bewerber. Qualitätssicherung und -optimierung darf daher kein singuläres Ereignis sein, sondern muss als kontinuierlicher Prozess institutionalisiert werden.
- Der schriftliche DGP-Test für den gehobenen Dienst wurde zum Saisonbeginn 01. 08. 1996, Einstellung August 1997, stark modifiziert (BIS-r-DGP), und eine Überprüfung der Prognosegüte wurde sinnvoll, nachdem die ersten Anwärtterinnen und Anwärtter die Laufbahnprüfung absolviert hatten.
- Kersting veröffentlichte bereits 1999 eine erste Erfolgskontrolle des BIS-r-DGP, die sich auf den Einsatz des Verfahrens bei einer großen Finanz- und Steuerverwaltung, Erfolgskriterium Zwischenprüfung, bezieht (DGP-Informationen Heft 56, 1999); die positiven Ergebnisse sind natürlich nicht direkt übertragbar auf die allgemeine Verwaltung, hatten aber bereits Auswirkungen auf die Modifikation des Testverfahrens für den gehobenen Dienst (seit August 2001 Test A1). Gleichwohl sollte die Treffsicherheit des Verfahrens für die allgemeine Verwaltung explizit überprüft werden.
- Der eigentliche aktuelle Anlass für die Berechnung der prognostischen Validität war aber die Tatsache, dass es seitens der Verwaltungen, die der Verwaltungsfachhochschule, kommunale Abteilung Braunschweig, ange-

---

<sup>1</sup> Dieser Bericht wurde erst ermöglicht durch die intensive Mitarbeit einer Reihe von DGP-Mitarbeiterinnen und -Mitarbeitern, denen wir zu großem Dank verpflichtet sind: Stephan Jerusel, Thomas Königer, Heike Konitz, Marina Leicher, Beate Scholz. Außerdem bedanken wir uns aufs Herzlichste bei allen auf Verwaltungsseite beteiligten Personalsachbearbeitern/innen, die durch das mühevollen Zusammentragen der Daten eine wichtige Voraussetzung für die Analyse der Prognosegüte der DGP-Personalauswahlverfahren lieferten.

<sup>2</sup> Die DIN 33430 formuliert „Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen“, siehe Heyse & Kersting, 2004; Kersting & Heyse, 2004 sowie das Heft 57 der DGP-Informationen mit drei Beiträgen zur DIN 33430, abrufbar unter [www.dgp.de](http://www.dgp.de). Kersting (2004a) geht in einer ausführlichen Darstellung der Frage nach, inwieweit die Verfahren der DGP den Anforderungen der DIN 33430 entsprechen.

schlossenen sind, Rückmeldungen über sehr hohe Durchfallquoten bei der Zwischenprüfung gab (Bewährungskontrolle 1).

- Da für die Landes- und Kommunalverwaltungen in Schleswig-Holstein bisher keine separate Erfolgskontrolle existierte, sollte parallel für diese seit 1974 von der DGP betreuten Verwaltungen ebenso eine Validitätsüberprüfung durchgeführt werden (Bewährungskontrolle 2).

## **1. DGP-Auswahlverfahren**

Eine ausführliche Beschreibung der DGP-Eignungsuntersuchung gehobener Dienst findet sich bei Kersting 1999, S. 43ff und in der – von der DGP abzurufenen – Broschüre „Eignungsdiagnostik“. Hier nur eine kurze Zusammenfassung des zweistufigen DGP-Eignungsuntersuchungsverfahrens für den gehobenen Dienst:

### **1.1 Schriftlicher Leistungstest**

Der schriftliche Leistungstest (BIS-r-DGP, Version 1, siehe Kersting und Beauducel, 1997, 2001) von ca. 4 Stunden Dauer erfasst und gewichtet (prozentuale Gewichte bei der Bestimmung des „Empfehlungsgrad Test“ in Klammern): Intellektuelle Leistungen wie Verarbeitungskapazität (30%), Einfallsmenge (5%), Merkfähigkeit (8%), Bearbeitungsgeschwindigkeit (10%); beim Umgang mit verbalem (15%), numerischem (10%) und figuralem (4%) Material; Büroarbeitsverhalten (6%), Rechtschreibung (2%) sowie Kenntnisse (10%) in den Wissensdomänen Politik, Wirtschaft, Gemeinschaftskunde, Literatur. Die Testleistungen werden anforderungsspezifisch gewichtet, zu einem Gesamtwert aggregiert und in einen sog. Testempfehlungsgrad auf einer 5-stufigen Skala (siehe unten) transformiert.

Die jeweiligen Verwaltungen nutzen das Testergebnis in Form des Testgesamtpunktwertes bzw. des Testempfehlungsgrades als Vorauswahlkriterium für die Zulassung zum zweiten Verfahrensteil, der Verhaltensbeurteilung. Die DGP empfiehlt, Bewerberinnen und Bewerber mit den Empfehlungsgraden „5“, „4,5“ und „4“ zum zweiten Verfahrensteil zuzulassen, bei einem Mangel an qualifizierten Personen werden de facto auch noch Personen mit einem Empfehlungsgrad von „3,5“, evtl. sogar „3“ bei der Einladung zum mündlichen Teil der Eignungsuntersuchungen berücksichtigt.

### **1.2 Verhaltensbeurteilung:**

Die Verhaltensbeurteilung basiert auf einem eintägigen Assessment-Center mit drei Stationen – Gruppendiskussion, Gruppenaufgabe, Einzelgespräch –, in denen die Bewerberinnen und Bewerber nach einem definierten System beobachtet und hinsichtlich folgender Merkmale bewertet werden (für eine ausführliche Beschreibung der DGP-Assessment-Center-Verfahren und deren Bewährung siehe Kersting (2003)):

- Sachbezogenheit
- Einfallsreichtum
- Soziale Kompetenz/Teamfähigkeit
- Kommunikation/Sprachlicher Ausdruck
- Zielorientierte Initiative
- Sicherheit/Belastbarkeit
- Berufsbezogene Motivation

Für jede der genannten Dimensionen sowie für das Insgesamt der Leistungen im Assessment-Center wird eine Verhaltensbeurteilung seitens der Psychologinnen und Psychologen abgegeben.

### 1.3 Gesamturteil

Die Leistungs- und die Verhaltensbeurteilung werden im Verhältnis 1 : 1 zu einer DGP-Gesamtempfehlung zusammengefasst. Das Eignungsurteil der DGP und die Einstellungsentscheidung der Organisation sind der Sache nach und im Interesse der Organisation und der ausbildungssuchenden Personen klar voneinander zu trennen (siehe Jäger, 1970, S. 641; Schuler, 1996, S. 151). Bei der letztendlichen Einstellungsentscheidung werden neben den Empfehlungen der DGP auch inner- und überorganisationale Belange (Personalbedarf, Anzahl qualifizierter Bewerberinnen und Bewerber, Merkmale der aktuellen Mitarbeiterinnen und Mitarbeiter (z. B. Geschlecht, Alter, Kompetenzen) berücksichtigt.

### 1.4 Empfehlungsgradskala

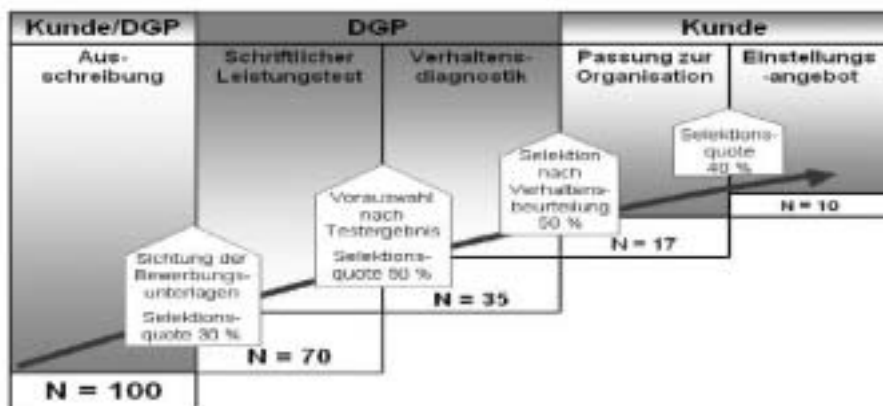
Die Beurteilungen für den Test, das Verhalten sowie die Gesamtbeurteilung werden auf der sog. Empfehlungsgradskala der DGP abgegeben. Diese Skala erstreckt sich in halben Schritten von dem negativen Wert „1“ bis zum positiven Wert „5“. Zur Bedeutung der einzelnen Skalenwerte: Eine gute Prognose liegt für Bewerberinnen und Bewerber mit einem Empfehlungsgrad von „5“, „den Anforderungen voll entsprechend“, vor. Die Empfehlungsgradstufe „4“ („den Anforderungen weitgehend entsprechend“) signalisiert leichte Einschränkungen in Bezug auf den zu erwartenden Ausbildungs- und Berufserfolg. Ein höheres Risiko ergibt sich bei der Empfehlungsgradstufe „3“, „den Anforderungen nur teilweise entsprechend“. Bei mit „3“ beurteilten Bewerbern bestehen seitens der DGP Bedenken gegenüber einer Zulassung. Die „Dreier-Kandidatinnen und -Kandidaten“ sind zwar nicht durchgängig leistungsschwach (das sind die Bewerber mit den Empfehlungsgraden „2“ und „1“), weisen aber neben Stärken auch Schwächen auf. Eine Prognose ist hier schwierig, es sind „Risikokandidaten“.

### 1.5 Sequenzielles, mehrstufiges Vorgehen

Die DGP-Eignungsuntersuchung stellt nur einen Teil des gesamten Selektionsprozesses dar. Dieser Gesamtselektionsprozess lässt sich unter Berücksich-

tigung der Erfahrungen der DGP ablauf- und zahlenmäßig etwa wie in Abbildung 1 darstellen (Die Zahlen entspringen den aus vielen Rückmeldegesprächen mit Personalverantwortlichen gewonnenen Informationen, sie sind als Durchschnittswerte zu verstehen. Im Einzelfall wird es aber aufgrund variierender Angebots- und Auswahlbedingungen und spezifischer Bedarfe stark davon abweichende Quoten geben.):

**Abbildung 1: Sequenzielles, mehrstufiges Vorgehen der Personalauswahl gehobener Dienst**



## 2. Bewährungskontrolle für die Kommunen der Niedersächsischen Fachhochschule für Verwaltung und Rechtspflege, kommunale Abteilung Braunschweig (Bewährungskontrolle 1)

Nachdem Anfang 2002 – wie oben beschrieben – von einigen Kommunalverwaltungen aus der Braunschweiger Region – aber auch aus Hannover – negative Rückmeldungen wegen der hohen Misserfolgsrate der Studierenden in der Zwischenprüfung an die DGP Hannover gegeben wurden, gab es einige Fachgespräche zwecks Ursachenfindung (s.u.). Danach machte die DGP das Angebot, eine genauere Analyse des Zusammenhangs zwischen Testleistung, Verhaltensbeurteilung und den Empfehlungen der DGP einerseits und dem Erfolg bzw. Misserfolg im Studium andererseits durchzuführen. Es wurden unter aktiver Mitarbeit der Personalstellen der beteiligten Verwaltungen die Daten der Studenten mit Zwischen- und Laufbahnprüfungsergebnissen zusammengetragen und den Ergebnissen aus den Eignungsuntersuchungen zugeordnet - dies geschah unter Beachtung der Datenschutzvorschriften für die DGP anonymisiert. Betroffen waren die Einstellungsjahrgänge 1997 - 1999, beteiligt waren 5 Stadt- bzw. Kreisverwaltungen. Für die Berechnung der Beziehungen zur Zwischenprüfung standen die Daten von 74, für die Abschlussprüfung die Daten von 58 Personen (Empfehlungsgrad Test bzw. von 52 Personen für den Empfehlungsgrad Verhalten sowie den Empfehlungsgrad Gesamt) zur Verfügung.

## 2.1 Ergebnisse der Eignungsuntersuchung: Prädiktoren

Tabelle 1 zeigt für die Bewerberinnen und Bewerber der Braunschweiger Untersuchungsstichprobe, die an der Laufbahnprüfung teilgenommen haben, die Häufigkeitsverteilung der testbasierten Empfehlungsgrade.

**Tabelle 1: Verteilung der DGP-Testempfehlungsgrade, Bewährungskontrolle 1**

DGP-Empfehlungsgrad	Testergebnis (N = 58)		Gesamtergebnis (N = 52)	
	Häufigkeit	%	Häufigkeit	%
Entspricht den Anforderungen 5	28	48,3 %	10	19,2 %
4,5	12	20,7 %	17	32,7 %
Entspricht weitgehend den Anforderungen 4	10	17,2 %	24	46,2 %
3,5	8	13,8 %	1	1,9 %
Entspricht nur teilweise den Anforderungen 3				
2,5				
Entspricht weitgehend nicht den Anforderungen 2				
1,5				
Entspricht nicht den Anforderungen 1				

Es handelt sich, wie es für einen Selektionsprozess typisch ist, um eine extrem asymmetrische, sog. abgeschnittene Verteilung. In Folge der Beachtung der DGP-Empfehlungen wird von den Verwaltungen in der Regel nur Personen mit positiven Empfehlungsgraden ein Ausbildungsangebot unterbreitet, so dass testmäßig schwächere Bewerberinnen und Bewerber nicht oder nur sehr selten das Fachhochschulstudium antreten können. Über die Bewährung der nicht eingestellten Bewerberinnen und Bewerber kann keine Aussage gemacht werden. Wie in der Darstellung anderer Bewährungskontrollen der DGP (siehe z. B. die „Krankenschwestern-Bewährungskontrolle“, Thielepape, 2003) beschrieben wurde, ist die abgeschnittene Verteilung ein Grund dafür, dass die Zusammenhänge zwischen den Ergebnissen in der DGP-Eignungsuntersuchung und dem Ausbildungserfolg unterschätzt werden.

Wie stark und wie einseitig die Gesamtbewerbergruppe durch den Prozess der Personalauswahl – und evtl. durch Selbstselektionseffekte der Bewerber – ausgelesen wird, zeigt sich z.B. darin, dass der durchschnittliche Wert für den wichtigsten DGP-Einzelprediktor „Verarbeitungskapazität“ für die Gruppe der 58

Personen bei 109,7 und somit eine Standardabweichung über dem für die repräsentative Bewerbergruppe bestimmten Wert 100 liegt!

Interessant ist auch eine Betrachtung der durchschnittlichen Leistungen der Bewerberinnen und Bewerber über die Jahre hinweg. Eine solche Betrachtung ist möglich, da die DGP pro Jahr ca. 8000 Bewerberinnen und Bewerber für den gehobenen Dienst testet. Beispielhaft soll hier der Anteil der Bewerberinnen und Bewerber mit dem testbasierten Empfehlungsgrad (EG) „5“ besprochen werden, also Personen, die von der DGP aufgrund ihrer Testleistungen uneingeschränkt empfohlen werden: Ein Jahr nach der Einführung des neuen Tests „BIS-r-DGP“, also 1997, erzielten rund 16% der Bewerberinnen und Bewerber den Empfehlungsgrad „5“. In den folgenden Jahren sank die prozentuale Häufigkeit der Bewerberinnen und Bewerber mit dem EG 5 kontinuierlich bis auf den Tiefpunkt des Einstellungsjahrganges 2001 von 9,54%, d.h. die Qualität der Bewerberinnen und Bewerber für den öffentlichen Dienst reduzierte sich – in erster Linie in inverser Abhängigkeit zur wirtschaftlichen Situation in Deutschland. Bei kritischer Wirtschaftslage steigt die Attraktivität des öffentlichen Dienstes als potenzieller Arbeitgeber wieder an, und es bewerben sich auch qualitativ bessere Bewerberinnen und Bewerber. Der Anteil der Bewerberinnen und Bewerber mit dem EG 5 stieg im Einstellungsjahr 2002 auf 14,76% und verblieb seitdem etwa auf diesem Niveau (Einstellungsjahr 2003 = 15,39%, 2004 = 14,54%). Diese Effekte zeigen sich natürlich in Abhängigkeit von regionalen und strukturellen Gegebenheiten in unterschiedlicher Ausprägung. Die Zahlen verdeutlichen, dass man von der Qualität einer begrenzten Gruppe von Bewerberinnen und Bewerbern nicht ohne weiteres auf das allgemeine Niveau der Bewerbungsgruppe rückschließen kann. Althoff hatte bereits 1997 gezeigt, dass Frauen, die sich beim Auswärtigen Amt bewerben, deutlich (test-)leistungsstärker sind als Frauen, die sich bei der allgemeinen Verwaltung bewerben. Die Qualität von Bewerberinnen und Bewerbern ist immer auch eine Frage der Attraktivität der um Bewerberinnen und Bewerber nachsuchenden Organisation (zum Personalmarketing siehe Kersting, 1997; Thielepape, 1990). Dieser Attraktivität der Organisation für Bewerberinnen und Bewerber kommt in der Zukunft aufgrund des demographisch bedingten Rückgangs an Bewerberinnen und Bewerbern eine maßgebliche Bedeutung zu.

## **2.2 Ergebnisse der Laufbahnprüfung**

Wegen der primären Ausrichtung dieser Analyse auf den Gesamtausbildungserfolg werden zuerst die Ergebnisse der Laufbahnprüfung und die Prognosegüte der Eignungsuntersuchung hinsichtlich der Laufbahnprüfung dargestellt. Die Verteilung der Noten in der Laufbahnprüfung wird in Tabelle 2 hinsichtlich des Erfolges/Misserfolges beim ersten Versuch (ohne Wiederholung) dargestellt. Es zeigt sich eine breite Verteilung über die gesamte Noten- bzw. Punkteskala mit – erwarteter – positiver Tendenz.

**Tabelle 2: Ergebnis der Laufbahnprüfung (Kriterium),  
Bewährungskontrolle 1**

Note	Häufigkeit	%
Sehr gut	1	1,7
Gut	14	24,1
Befriedigend	35	60,3
Ausreichend	5	8,6
Durchgefallen	3	5,2

### 2.3 Beziehung zwischen DGP- Empfehlung und Erfolg in der Laufbahnprüfung

Von einer validen Eignungsuntersuchung ist eine positive Beziehung zwischen dem Ergebnis der Eignungsuntersuchung (Prädiktor) und dem Prüfungserfolg (Kriterium) zu erwarten, was simpel ausgedrückt heißt, dass Bewerber mit einem guten DGP-Empfehlungsgrad auch erfolgreich die Prüfung bestehen und dass umgekehrt Personen mit einem schwachen Ergebnis in der Eignungsuntersuchung auch in der Ausbildung und Prüfung eher Schwierigkeiten bekommen. Eine eindeutige Beziehung ist natürlich nicht zu erwarten, da u.a. nicht nur Personenvariablen, sondern auch situative Bedingungen sowie Interaktionen zwischen Person und Umwelt Einfluss auf den Erfolg in Ausbildung und Prüfung haben. Prognosen haben nur Wahrscheinlichkeitscharakter.

Diese Beziehungen zwischen den Prädiktoren und den Kriterien können in Form von Kreuztabellen dargestellt werden. In allen drei berechneten Kreuztabellen (Beziehung zwischen den Kriterien Test, Verhalten, Gesamtempfehlung und dem Kriterium Erfolg in der Laufbahnprüfung) zeigen sich erwartungsgemäße Beziehungen. Aus Platzgründen soll hier nur exemplarisch die Beziehung des Prädiktors Gesamtempfehlungsgrad mit der Laufbahnprüfung abgebildet werden (Tabelle 3).

#### Lesebeispiele:

1. Von den 10 Bewerbern/innen mit der besten Empfehlung „5 = zu empfehlen“ erzielten 6 = 60% ein Gut und 4 = 40% ein Befriedigend. Keine Person aus dieser Gruppe der laut DGP-Verfahren hervorragend beurteilten Bewerberinnen und Bewerber fällt durch die Laufbahnprüfung oder schneidet nur mit „ausreichend“ ab.
2. Bewerber/innen mit dem EG 4 (N = 24): 1 Person = 4,2% erzielt „sehr gut“, 2 = 8,3% bekommen „gut“. Insgesamt haben also drei von 24 Personen (12,5%), die mit „4“ im Auswahlverfahren beurteilt wurden, die Laufbahnprüfung mit „sehr gut“ oder „gut“ abgeschlossen. Demgegenüber konnten sechs von zehn (60%) der Bewerber(innen), die im Verfahren einen EG von „5“ erzielt hatten, in der Laufbahnprüfung mit einer „guten“ Leistung überzeugen. 66,7% der mit „4“ beurteilten Bewerber erzielten in der Laufbahnprüfung ein „Befriedigend“. Aber vier Personen = 16,7% werden „ausreichend“ bewertet und eine Person = 4,2% ist durchgefallen. Fasst man „ausreichend“ und „durchgefallen“ zusammen

men, sind 21% aller mit „4“ im Verfahren beurteilten Personen von schlechten Leistungen in der Laufbahnprüfung betroffen. Demgegenüber gab es keine einzige Person, die im Verfahren mit der besten Empfehlung „5“ ausgezeichnet wurde und in der Laufbahnprüfung versagt hat!

3. Die 31 Personen mit „befriedigend“ in der Laufbahnprüfung (senkrecht) rekrutieren sich aus 4 Personen = 12,9% mit dem DGP-Empfehlungsgrad 5, 11 Personen = 35,5% mit dem EG 4,5 und 16 Personen = 51,6% mit dem EG 4.

**Tabelle 3: Kreuztabelle DGP-Gesamtempfehlungsgrad und Ergebnis der Laufbahnprüfung, Bewährungskontrolle 1**

DGP-Gesamt-EG	Ergebnis der Laufbahnprüfung					Notendurchschnitt	Gesamt
	Sehr gut	Out	Befriedigend	Ausreichend	Durchgefallen		
5,0		6	4			2,4	10
		60,0 %	40,0 %				100,0 %
		46,2 %	12,9 %				19,2 %
4,5		6	11	1		2,8	17
		29,4 %	64,7 %	5,9 %			100,0 %
		39,5 %	35,5 %	20,0 %			32,7 %
4,0	1	2	10	4	1	3,2	24
	4,2 %	8,3 %	66,7 %	16,7 %	4,2 %		100,0 %
	100,0 %	15,4 %	51,6 %	80,0 %	50,0 %		46,2 %
3,5					1	5,0	1
					100,0 %		100,0 %
					100,0 %		1,9 %
Gesamt	1	12	31	5	2		52
	1,9 %	25,0 %	59,6 %	9,6 %	3,8 %		100,0 %
	100,0 %	100,0 %	100,0 %	100,0 %	100,0 %		100,0 %

4. Die beiden durchgefallenen Personen wurden von der DGP insgesamt mit 4 (Test 4, Verhalten 4) bzw. 3,5 (Test 3,5, Verhalten 4) beurteilt. Die Verwaltung hätte sich und der Person mit dem Testergebnis 3,5 das Ausbildungsversagen (und die auf beiden Seiten getätigten Investitionen von Zeit und Geld) ersparen können. „3er“-Kandidatinnen und -Kandidaten gelten in der DGP-Terminologie als Risiko-Kandidat(inn)en (siehe oben, Abschnitt 1.4). Dennoch entscheiden sich viele Verwaltungen aus pragmatischen Gründen, auch auf Personen mit einem Testergebnis von 3,5 zurückzugreifen.
5. Die Bewerberinnen und Bewerber mit der DGP-Gesamtempfehlung 5 erzielen in der Laufbahnprüfung eine Durchschnittsnote von 2,4, mit dem EG 4,5 erreichen sie durchschnittlich 2,8. Beim EG 4,0 ergibt sich ein Notendurchschnitt von 3,2 und beim EG 3,5 von 5,0. D.h., die Durchschnittsnote in der Laufbahnprüfung steht in eindeutiger Beziehung zum DGP-Gesamtempfehlungsgrad.

Die in dieser Kreuztabelle enthaltenen Ergebnisse entsprechen damit gut den Erwartungen an ein prognostisch valides Personalauswahlinstrument. Dieses positive Ergebnis erfährt eine Bestätigung in den berechneten Korrelationen zwi-



schen den DGP-Testergebnissen und Empfehlungsgraden und dem Ergebnis in der Laufbahnprüfung (Tabelle 4). Als Maß für den Zusammenhang zwischen Prädiktor und Kriterium werden Korrelationen berechnet, wobei ein Korrelationskoeffizient variiert zwischen  $r = -1.0$  (vollständig negativer Zusammenhang) über  $r = 0.0$  (kein Zusammenhang) bis zum anderen Extrem  $r = +1.0$  (vollständig positiver Zusammenhang). Die gegensätzliche „Polung“ der DGP-Skala und der Noten in der Laufbahnprüfung wurde zum Zwecke des leichteren Verständnisses eliminiert: Dargestellt sind logisch positive Korrelationen. Wegen der Nicht-Normalität der Verteilung wurde hier Spearman Rho berechnet.

**Tabelle 4: Korrelationen zwischen DGP-Prädiktoren und dem Kriterium Laufbahnprüfung, Bewährungskontrolle 1**

Testgesamtpunktwert (anforderungsspezifisch gewichtet, N = 58)	$r = .43^{**}$ ( $r = .50$ )
Empfehlungsgrad Test (N = 58)	$r = .43^{**}$ ( $r = .77$ )
Empfehlungsgrad Verhalten (N = 52)	$r = .43^{**}$ ( $r = .58$ )
Empfehlungsgrad Gesamt (N = 52)	$r = .46^{**}$ ( $r = .64$ )

\*\* Die Rangkorrelationen sind auf dem Niveau von 0,01 signifikant (1-seitig).

( ) Korrektur gegen „restriction of range“ wegen der abgeschnittenen Verteilung.

Die Korrelationen für die drei DGP-Empfehlungsgrade („Test“, „Verhalten“ und „Gesamt“) sind hoch signifikant und insbesondere unter Berücksichtigung der stark abgeschnittenen Verteilung voll befriedigend. Die Werte in Klammern beschreiben die hypothetischen Ergebnisse ohne Streuungsreduktion wegen der abgeschnittenen Verteilung.

Eine genaue Analyse der Ergebnisse zeigt, dass sich einzelne Dimensionen des Leistungstests als prognostisch bedeutsam erweisen, andere hingegen nicht. Auf diese differenzierten Ergebnisse wird weiter unten (nachdem auch die Befunde für die Bewährungskontrolle Schleswig-Holstein dargestellt wurden) eingegangen.

Resümiert kann gesagt werden, dass die DGP-Eignungsuntersuchung für den gehobenen Dienst ein valides Instrumentarium darstellt, um den Erfolg in der Laufbahnprüfung des gehobenen Dienstes der allgemeinen Verwaltung treffsicher vorauszusagen.

## 2.4 Ergebnisse der Zwischenprüfung

Ein wesentlicher Anlass für die Durchführung dieser Bewährungskontrolle war - wie oben beschrieben - der Anfang 2001 berichtete Anstieg der Durchfallerquote an der Fachhochschule. Zu erwarten sind deshalb für die hier untersuchte Stichprobe weniger positive Beziehungen zwischen den DGP-Empfehlungen und den tatsächlichen Zwischenprüfungsergebnissen. Die Untersuchungen der prognostischen Qualität des DGP-Verfahrens leiden unter der Tatsache, dass das Erfolgskriterium „Zwischenprüfung“ zum Zeitpunkt der Betrachtung weder

mathematisch aus den Vor- und Prüfungsleistungen zu errechnen ist (siehe Anhang: Berechnung der Zwischenprüfungsergebnisse gem. § 6 APVOgehD) noch in Form einer Verteilung bzw. eines Zahlenwertes vorliegt, sondern nur dichotom in der Aussage „bestanden/nicht bestanden“ besteht.

**Tabelle 5: Ergebnisse der Zwischenprüfung, Bewährungskontrolle 1**

	Häufigkeit	%
Bestanden	63	85,1
Nicht bestanden	11	14,9
<b>Gesamt</b>	<b>74</b>	<b>100,0</b>

## 2.5 Beziehung zwischen DGP- Empfehlung und Zwischenprüfung

Auch hier soll eine Beschränkung auf die Besprechung der Kreuztabelle „Gesamtempfehlungsgrad der DGP und Ergebnis der Zwischenprüfung“ vorgenommen werden, weil die Beziehungen zwischen dem DGP-Test bzw. der Verhaltenseinschätzung seitens der DGP zum Zwischenprüfungserfolg ähnliche Ergebnisse zeigen.

**Tabelle 6: Kreuztabelle DGP-Gesamtempfehlungsgrad und Ergebnis der Zwischenprüfung, Bewährungskontrolle 1**

Gesamt EG	Bestanden	Nicht bestanden	Gesamt
<b>5,0</b>	10	3	13
	<b>76,9 %</b>	<b>23,1 %</b>	<b>100,0 %</b>
	17,9 %	27,3 %	19,4 %
<b>4,5</b>	19	2	21
	<b>90,5 %</b>	<b>9,5 %</b>	<b>100,0 %</b>
	33,9 %	18,2 %	31,3 %
<b>4,0</b>	25	4	29
	<b>86,2 %</b>	<b>13,8 %</b>	<b>100,0 %</b>
	44,6 %	36,4 %	43,3 %
<b>3,5</b>	2	2	4
	<b>50,0 %</b>	<b>50,0 %</b>	<b>100,0 %</b>
	3,8 %	18,2 %	6,0 %
<b>Gesamt</b>	56	11	67
	<b>83,6 %</b>	<b>16,4 %</b>	<b>100,0 %</b>
	100,0 %	100,0 %	100,0 %

Trotz einer uneingeschränkten DGP-Empfehlung (EG 5) bestehen wider Erwarten 23,1% die Zwischenprüfung nicht. Erwartungsgemäß ist der Erfolg in der Zwischenprüfung bei den Empfehlungsgraden „4,5“ (90,5 %) und EG „4“ (86,2%) höher als bei dem Empfehlungsgrad „3,5“. Die Korrelationen zwischen den DGP-Prädiktoren und dem dichotomen Kriterium der Zwischenprüfung sind enttäuschend, sie liegen auf Höhe von Nullkorrelationen. Das bedeutet, dass bei dieser Untersuchungsgruppe mit Hilfe der DGP-Eignungsuntersuchung keine erfolgreiche Prognose hinsichtlich des Abschneidens in der Zwischenprüfung möglich war.

Zusammengefasst heißt das: Die DGP-Eignungsuntersuchung ist sehr wohl dazu in der Lage, den Gesamtausbildungserfolg, dokumentiert in der Laufbahnprüfung, valide vorherzusagen; aber bei der hier untersuchten Stichprobe konnte das Ergebnis in der – vor der Laufbahnprüfung stattfindenden – Zwischenprüfung nicht sicher vorhergesagt werden! In verschiedenen Gesprächsrunden, u. a. mit Personalverantwortlichen der betroffenen Verwaltungen, mit Vertretern der beteiligten Fachhochschule und im Beirat der DGP, wurden mögliche Ursachen für diese Diskrepanz diskutiert. Mit Fokus auf den ersten Teil der Ausbildung und die Zwischenprüfung können – komprimiert dargestellt – folgende Gründe in Erwägung gezogen werden:

- trotz guter intellektueller Kapazität fallen Fachhochschulstudenten/innen durch die Zwischenprüfung, weil die Einstellung zur Ausbildung, die Leistungsmotivation, der Einsatzwille und das Engagement, die Selbstmotiviertheit, das Selbstmanagement gegen Widerstand anzuarbeiten etc. nicht angemessen entwickelt sind bzw. nicht adäquat aktiviert werden. Aber nach „dem Schuss vor den Bug“ (Durchfallen) gelingt zumeist die Wiederholungsprüfung,
- das Studium ist in Teilen nicht zielführend auf die Prüfung hin aufgebaut,
- die Studierenden bereiten sich häufig nicht genügend methodisch auf die Prüfungen vor,
- in den Verwaltungen der Großstädte, die höhere Durchfallquoten als kleinere Städte aufweisen, ist erstens vermutlich das Ablenkungspotenzial enorm (Konzentrationsproblematik) und zweitens können sich die Anwärter/innen leicht während der praktischen Ausbildungszeiten in der Organisation „verlieren“,
- das ausschließlich dichotome Prüfungsergebnis (bestanden - nicht bestanden) und die nicht-mathematische Bestimmung dieses Ergebnisses lassen keine hohen Korrelationen erwarten,
- viele Gründe für das Ausscheiden aus der Ausbildung bzw. für Misserfolge in der Zwischenprüfung sind nicht leistungs-, sondern situationsbedingt,
- da von anderen Fachhochschulen keine hohe Durchfallerquote berichtet wurde, scheinen auch Spezifika regionaler Art ursächlich zu sein. Der Prozentsatz der Durchfaller schwankt über die Jahre sehr stark (in den letzten 13 Jahren bei einer großen niedersächsischen FH für die Zwischenprüfung von 7,87% bis 23,98%, für die Laufbahnprüfung von 5,14% bis 22,70%). Höchststände können aber – eindeutig belegbar – nicht in Zusammenhang mit Änderungen der DGP-Eignungsuntersuchungen gebracht werden.

### 3. Bewährungskontrolle für die Landes- und Kommunalverwaltungen in Schleswig-Holstein (Bewährungskontrolle 2)

Die Intention seitens der DGP Hannover, nach vielen Jahren der Zusammenarbeit speziell für das Bundesland Schleswig-Holstein eine Überprüfung des Prognoseerfolgs des Auswahlverfahrens für den gehobenen Dienst vorzunehmen, traf auf die gleichartigen Wünsche einiger Kommunalverwaltungen und insbesondere das starke Interesse des Innenministeriums an einer Methodenüberprüfung. Es gab keine speziellen Anlässe z.B. in Form von gravierenden Ausbildungsmisserfolgen, aber natürlich war über die Jahre nicht jede/r von der DGP positiv in der Eignungsuntersuchung bewertete Bewerber/in in der Ausbildung erfolgreich, sondern es gab schon Ausfälle während der Ausbildung und auch Prüfungsversager. Ähnlich wie beim Analyseverfahren in der Braunschweiger Untersuchung wurde wegen der Beachtung des Datenschutzes für das Zusammentragen der Daten eine starke Mitarbeit der beteiligten Verwaltungen notwendig. Unter Einbindung der jeweiligen Personalmitarbeiter/innen wurden in den Verwaltungen die Datensätze sowohl der DGP-Eignungsuntersuchungen als auch der Ausbildung und Prüfungen vollständig erstellt und dann anonymisiert an die DGP geschickt. Hier fanden – ohne Identifikationsmöglichkeit von einzelnen Personen – die statistischen Berechnungen statt. Insgesamt konnten die Daten von 91 Personen (Empfehlungsgrad-Test bzw. 88 Personen für den Empfehlungsgrad Verhalten sowie den Empfehlungsgrad Gesamt) in die Analyse einbezogen werden.

#### 3.1 Ergebnisse der Eignungsuntersuchung: Prädiktoren

Die Schleswig-Holstein-Stichprobe der an der Laufbahnprüfung teilnehmenden Bewerberinnen und Bewerber erzielte zum Zeitpunkt der Eignungsuntersuchungen die in Tabelle 7 dargestellte Verteilung der Empfehlungsgrade.

**Tabelle 7: Verteilung der Empfehlungsgrade im Test und bei der Gesamtempfehlung, Bewährungskontrolle 2**

DGP-Empfehlungsgrad	Testergebnis (N = 91)		Gesamtergebnis (N = 88)	
	Häufigkeit	%	Häufigkeit	%
Erfüllt alle Anforderungen 5	63	69,2 %	42	47,4 %
4,5	9	9,9 %	35	39,8 %
Erfüllt weitgehend alle Anforderungen 4	13	14,3 %	9	10,2 %
3,5	4	4,4 %	1	1,1 %
Erfüllt nur teilweise alle Anforderungen 3	1	1,1 %	1	1,1 %
2,5	1	1,1 %		
Erfüllt weitgehend nicht alle Anforderungen 2				
1,5				
Erfüllt nicht alle Anforderungen 1				

Anmerkung: Für die Personen liegen die Daten zum Gesamt-Empfehlungsgrad nicht vor.

Auch hier handelt es sich um eine stark abgeschnittene Verteilung, bei der aber auch die schwächeren Empfehlungsgrade „3 = entspricht nur teilweise den Anforderungen“ und (hinsichtlich des testbasierten Empfehlungsgrades) sogar „2,5“ vertreten sind. Die Schiefe der Verteilung ist – erwartungsgemäß, da die meisten Verwaltungen sich stark an den DGP-Empfehlungen orientieren – extrem: 69,2 % (!) der Personen, welche die Laufbahnprüfung absolviert haben, hatten im Test den besten Empfehlungsgrad „5“ erzielt, nur 6,6% schnitten schlechter ab als mit dem EG 4,0. Wegen der für die Verwaltungen günstigeren Relation von guten DGP-Ergebnissen der Bewerberinnen und Bewerber und begrenzter Anzahl der Ausbildungsplätze war es den Entscheidern möglich, sich bei der Einstellungsentscheidung auf bessere Bewerberinnen und Bewerber mit den besten DGP-Empfehlungsgraden zu konzentrieren. Diese für die Verwaltung günstige Situation ist für die Bestimmung der Aussagekraft des Auswahlverfahrens sehr ungünstig, da aufgrund der extrem abgeschnittenen Verteilung kaum noch damit gerechnet werden kann, die Prognosekraft des Verfahrens nachzuweisen.

### 3.2 Ergebnisse der Laufbahnprüfung

Von den Personen, die in der Analyse berücksichtigt werden konnten, haben 17,6% die Laufbahnprüfung mit „gut“ bestanden, 64,8% erzielten ein „Befriedigend“ und 17,6% ein „Ausreichend“. Was bei diesen Zahlen unberücksichtigt bleibt: Fünf Personen mussten die Laufbahnprüfung wiederholen, sie hatten sie im ersten Anlauf nicht bestanden. Für die folgenden Analysen wurde bei diesen fünf Personen der Wert herangezogen, der in der Prüfungswiederholung erzielt wurde. Die Zusammenhänge zwischen dem DGP-Auswahlverfahren und der Laufbahnprüfung wären nominell höher (deutlicher), wenn stattdessen der Wert aus der ersten Laufbahnprüfung zugrunde gelegt worden wäre.

**Tabelle 8: Ergebnisse der Laufbahnprüfung: Kriterien, Bewährungskontrolle 2**

Note	N	%
Sehr gut		
Gut	16	17,6
Befriedigend	59	64,8
Ausreichend	16	17,6
Durchgefallen		

**Anmerkung:** 5 Personen haben die Laufbahnprüfung wiederholt bzw. im ersten Anlauf nicht bestanden. Für die Analysen wurde bei diesen 5 Personen der Wert herangezogen, der in der Prüfungswiederholung erzielt wurde. Die Zusammenhänge zwischen dem DGP-Auswahlverfahren und der Laufbahnprüfung sind nominell höher, wenn stattdessen der Wert aus der ersten Laufbahnprüfung zugrunde gelegt wird.

### 3.3 Beziehung zwischen DGP-Empfehlungsgrad und Laufbahnprüfungsergebnis

Im Abschnitt 2.5, weiter oben, sind Aussagen über die Erwartungen der Beziehungen und die Methodik der Ergebnisdarstellung getroffen worden, die auch bei dieser Bewährungskontrolle für Schleswig-Holstein zutreffen. Ebenso findet auch hier bei der Beschreibung der Analyseergebnisse eine Beschränkung auf die Abhängigkeit von DGP-Gesamtempfehlungsgrad und Laufbahnprüfungsergebnis statt:

Mit Ausnahme der beiden Personen mit schwächeren Empfehlungsgraden stimmen die Ergebnisse recht gut mit den Erwartungen überein: Die Prädiktoren-Kriterien-Relation ist positiv: D.h. Personen mit besseren Empfehlungsgraden erzielen auch tendenziell bessere Prüfungsergebnisse.

Auch hier sollen zum leichteren Verständnis der Kreuztabelle einige Lesebeispiele gegeben werden:

#### Lesebeispiele:

- Von 42 Bewerbern/innen mit der besten Empfehlung „5 = zu empfehlen“ erzielen 12 = 28% ein „Gut“, 24 = 57,1% ein „Befriedigend“. Nur 6 Personen, also 14,3% erzielten mit „Ausreichend“ ein kritisches Ergebnis in der Laufbahnprüfung. Personen, die im Verfahren mit „4,5“ oder „4“ beurteilt wurden, konnten demgegenüber auch in der Laufbahnprüfung häufiger nicht überzeugen.
- Die beste Note, die in der Laufbahnprüfung überhaupt erreicht wurde, die Note „gut“, wurde ausschließlich von Personen erreicht, die im DGP-Verfahren die besten Empfehlungen, nämlich „4,5“ und „5“ erzielt haben.
- Mit zunehmend schwächerem Empfehlungsgrad ist ein Absinken des Prozentsatzes der guten Noten in der Laufbahnprüfung und ein Anstieg der schwächeren Noten zu verzeichnen.

#### **Tabelle 9: Kreuztabelle DGP-Gesamtempfehlungsgrad und Ergebnis in der Laufbahnprüfung, Bewährungskontrolle 2**

*siehe Seite 16*

**Tabelle 9: Kreuztabelle DGP-Gesamtempfehlungsgrad und Ergebnis in der Laufbahnprüfung, Bewährungskontrolle 2**

DGP-Gesamt-EG	Ergebnis der Laufbahnprüfung			Gesamt
	Out	Befriedigend	Ausreichend	
5,0	12	24	8	44
	29,6 %	57,1 %	14,3 %	100,0 %
	75,0 %	42,1 %	40,0 %	47,7 %
4,5	4	24	7	35
	11,4 %	68,6 %	20,0 %	100,0 %
	25,0 %	42,1 %	46,7 %	39,8 %
4,0		7	2	9
		77,8 %	22,2 %	100,0 %
		12,3 %	13,3 %	10,2 %
3,5		1		1
		100,0 %		100,0 %
		1,8 %		1,1 %
3,0		1		1
		100,0 %		100,0 %
		1,8 %		1,1 %
Gesamt	16	57	15	88
	18,2 %	64,8 %	17,0 %	100,0 %
	100,0 %	100,0 %	100,0 %	100,0 %

**Anmerkung:** Fünf Personen haben die Laufbahnprüfung wiederholt/im ersten Anlauf nicht bestanden. Für die Analysen wurde bei diesen fünf Personen der Wert herangezogen, der in der Prüfungswiederholung erzielt wurde. Die Zusammenhänge zwischen dem DGP-Auswahlverfahren und der Laufbahnprüfung sind nominell höher (deutlicher), wenn stattdessen der Wert aus der ersten Laufbahnprüfung zugrunde gelegt wird.

In Tabelle 10 werden die korrelativen Beziehungen zwischen dem DGP-Verfahren der Eignungsbeurteilung und dem späteren Erfolg in der Laufbahnprüfung für die Bewährungskontrolle 2, Schleswig Holstein, dargestellt. Die in Folge der - gegenüber der Braunschweiger Stichprobe noch - extremeren Vorauswahl noch stärker „abgeschnittene“ Verteilung führt zu leicht geringeren Korrelationen zwischen z. B. dem Empfehlungsgrad Test und dem Ergebnis in der Laufbahnprüfung. Der wegen Streuungseinschränkungen korrigierte Wert fällt dafür entsprechend höher aus. Überraschend ist, dass die Verhaltensbeurteilung bei dieser Bewährungskontrolle nicht im Zusammenhang mit dem Laufbahnprüfungserfolg steht.

**Tabelle 10: Korrelationen zwischen DGP-Prädiktoren und dem Kriterium Laufbahnprüfung, Bewährungskontrolle 2**

Testgesamtpunktwert (anforderungsspezifisch gewichtet, N = 91)	$r = .33^{**}$ ( $r = .41$ )
Empfehlungsgrad Test (N = 91)	$r = .35^{**}$ ( $r = .69$ )
Empfehlungsgrad Verhalten (N = 89)	$r = .07$
Empfehlungsgrad Gesamt (N = 88)	$r = .22^{**}$ ( $r = .34$ )

\*\* Die Rangkorrelationen sind auf dem Niveau von 0,01 signifikant (1-seitig).

\* Die Rangkorrelationen sind auf dem Niveau von 0,05 signifikant (1-seitig).

( ) Korrektur gegen „restriction of range“ wegen der abgeschnittenen Verteilung.

#### **4. Differenzierte Betrachtung der Zusammenhänge zwischen den einzelnen Testdimensionen und dem Prüfungserfolg bei beiden Bewährungskontrollen**

Eine genaue Analyse der Ergebnisse zeigt, dass sich in beiden Bewährungskontrollen einige Einzeldimensionen des Leistungstests als prognostisch bedeutsam erweisen, andere hingegen weniger. Entgegen den Befunden der Anforderungsanalyse war beispielsweise die Einfallsmenge der Bewerber in beiden Bewährungskontrollen vergleichsweise irrelevant für den Erfolg in der Laufbahnprüfung. Auch die Fähigkeit zum Umgang mit figuralem Aufgabenmaterial erwies sich für den Erfolg in der Laufbahnprüfung in Bewährungskontrolle 2 mehr oder minder als unbedeutend. Da sich diese Befunde auch in anderen Bewährungskontrollen mit ähnlichen Berufsbildern (zum Teil aber unter Verwendung anderer Kriterienmaße, siehe z.B. Kersting, 1999) abzeichneten, konnte der DGP-Leistungstest aufgrund der Analysen optimiert werden. Die gute Zusammenarbeit zwischen den Kunden/innen der DGP und der DGP ist somit Garant einer kontinuierlichen Qualitätsverbesserung. Die neueste Version des Testverfahrens, der sogenannte „A 1“-Test, verzichtet konsequenterweise auf Aufgaben zur Einfallsmenge sowie auf figurales Aufgabenmaterial. Dadurch wird der DGP-Test zugleich kürzer und vorhersagegenauer. An der Prüfung anderer Leistungsdimensionen wie z.B. der Merkfähigkeit wird ungeachtet der ausbleibenden prognostischen Validität festgehalten. Nach Auskunft der DGP-Kunden/innen sind diese Fähigkeiten zwar nicht für die Laufbahnprüfung, wohl aber für die berufliche Bewährung entscheidend.

#### **5. Fazit**

Die vorliegende Untersuchung bestätigt die Treffsicherheit und den Wert der DGP-Eignungsuntersuchungen durch die folgenden sechs Hauptbefunde:

- Von den insgesamt (beide Bewährungskontrollen) 52 Personen, die im Auswahlverfahren als „den Anforderungen entsprechend“ (Gesamtempfehlungsgrad „5“) beurteilt werden, gab es keine, die den Anforderungen der Laufbahnprüfung entgegen der Aussage des DGP-Verfahrens nicht genügte.
- 89,3 % aller Personen, welche die Laufbahnprüfung mit der Note „sehr gut“ oder „gut“ absolviert haben, hatten im DGP-Verfahren die besten Empfehlungsgrade von „4,5“ oder „5“ erzielt.
- Es ist wenig erfolgversprechend, Bewerberinnen und Bewerber, die von der DGP nicht empfohlen werden konnten, zur Ausbildung zuzulassen. Es gibt keine einzige Person mit dem relativ „schlechten“ Gesamtempfehlungsgrad „3,5“ oder „3“, die in der Laufbahnprüfung eine gute Note erzielt hat.
- Von den drei Personen mit dem schlechtesten Gesamtempfehlungsgrad der vorliegenden Untersuchungsgruppen („3“ und „3,5“) haben zwei Personen die Laufbahnprüfung mit „befriedigend“ bestanden, die dritte Person ist „durchgefallen“. Der DGP-Definition entsprechend handelt es sich bei Bewerberinnen und Bewerbern, die mit einem Empfehlungsgrad im „Dreier-Bereich“ beurteilt werden, de facto um „Risikokandidat(inn)en“. Da es mehrere hundert Bewerber(innen) mit einem Gesamtempfehlungsgrad von „3,5“ oder schlech-



- ter gab, ist davon auszugehen, dass durch das neue Auswahlverfahren Fehlinvestitionen in Millionenhöhe vermieden werden konnten (zu Kosten-/Nutzen-Berechnungen für die DGP-Verfahren siehe Kersting, 2004b).
- Trotz der extremen Vorauswahl auf der Basis der Testleistungen und der damit verbundenen Streuungseinschränkung betrug die zufallskritisch gesicherte Korrelation zwischen dem Testgesamtpunktwert und dem Erfolg in der Laufbahnprüfung  $r = .43$  (Bewährungskontrolle 1) bzw.  $r = .33$  (Bewährungskontrolle 2). Dieser Zusammenhang muss angesichts der extremen Streuungseinschränkung als überaus zufriedenstellend und substantiell eingestuft werden. Der für Streuungseinschränkungen korrigierte Vorhersagekoeffizient für die testbasierten Empfehlungsgrade betrug  $r = .50$  bzw.  $r = .41$ . Der Testgesamtpunktwert – bzw. in etwas größerem Maße der Testempfehlungsgrad – ist also eine gute Basis für die Vorauswahlentscheidung.
  - Der Gesamtempfehlungsgrad stellt eine gute Basis für die Einstellungsentscheidung für die Ausbildung zum gehobenen Dienst der allgemeinen Verwaltung dar.

Empirische Kontrollen der Aussagekraft von berufsbezogenen Eignungsbeurteilungen sind weder eine lästige Pflicht noch eine unnötige „Schönwetter“-Tätigkeit (nice to know), sondern sie sind der Garant dafür, dass Schwächen in Verfahren erkannt und beseitigt werden, dass Veränderungen in den Anforderungen zeitnah berücksichtigt werden und dass Verfahren kontinuierlich verbessert werden.

## 6. Literatur

Althoff, K. (1997). *Erste Erfahrungen mit dem neuen DGP-Eignungstest für den gehobenen Dienst (BIS-r-DGP) unter besonderer Berücksichtigung gruppen-spezifischer Unterschiede*. DGP-Informationen, 46, Heft 55, S. 59-68.

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

Heyse, H. & Kersting, M. (2004). *Anforderungen an den Prozess der Eignungsbeurteilung*. In L. F. Hornke & U. Winterfeld (Hrsg.), *Eignungsbeurteilungen auf dem Prüfstand: DIN 33430 zur Qualitätssicherung* (S. 29-41). Heidelberg: Spektrum Akademischer Verlag.

Hossiep, R. (1995). *Berufseignungsdiagnostische Entscheidungen. Zur Bewahrung eignungsdiagnostischer Ansätze*. Göttingen: Hogrefe.

Jäger, A.O. (1970). *Personalauslese*. In K. Gottschaldt, Ph. Lersch, F. Sander & H. Thomae (Hrsg.), *Handbuch der Psychologie* (Bd. 9, *Betriebspsychologie*, S. 613-667). Göttingen: Hogrefe.

- Kersting, M. (1997). *Personalmarketing und die soziale Akzeptanz von Auswahlverfahren*. DGP-Informationen, 46, Heft 55, S. 69-86.
- Kersting, M. (1999). *Personalauswahl mit dem BIS-r-DGP-Test: Evaluation der prognostischen Validität*. DGP-Informationen, 47, Heft 56, S. 43-60.
- Kersting, M. & Beauducel, A. (1997). *Der neue DGP-Leistungstest »BIS-r-DGP«: Informationen zu ausgewählten Testgütekriterien und zur Normierung*. DGP-Informationen, 46, Heft 55, S. 92-102.
- Kersting, M. & Beauducel, A. (2001). *BIS-r-DGP. Berliner Intelligenzstruktur-Test der Deutschen Gesellschaft für Personalwesen e. V.* In: W. Sarges & H. Wottawa (Hrsg.). *Handbuch wirtschaftspsychologischer Testverfahren* (S. 103-110). Lengerich: Pabst Science Publishers.
- Kersting, M. (2003). *Assessment Center: Erfolgsmessung und Qualitätskontrolle*. In S. Höft & B. Wolf. *Qualitätsstandards für Personalentwicklung in Wirtschaft und Verwaltung* (S. 72-93). Hamburg: Windmühle. (ISBN: 3-922789-92-7)
- Kersting, M. (2004a). *Qualitätssicherung und -verbesserung: Zur Überprüfung der Gültigkeit berufsbezogener Eignungsbeurteilungen*. In L.F. Hornke & U. Winterfeld (Hrsg.). *Eignungsbeurteilungen auf dem Prüfstand: DIN 33430 zur Qualitätssicherung* (S. 103-128). Heidelberg: Spektrum Akademischer Verlag.
- Kersting, M. (2004b). *Kosten und Nutzen beruflicher Eignungsbeurteilungen*. In L.F. Hornke & U. Winterfeld (Hrsg.). *Eignungsbeurteilungen auf dem Prüfstand: DIN 33430 zur Qualitätssicherung* (S. 55-77). Heidelberg: Spektrum Akademischer Verlag.
- Kersting, M. & Heyse (2004). *Anforderungen an die Qualität der Verfahren*. In L.F. Hornke & U. Winterfeld (Hrsg.). *Eignungsbeurteilungen auf dem Prüfstand: DIN 33430 zur Qualitätssicherung* (S. 43-54) Heidelberg: Spektrum Akademischer Verlag.
- Schuler, H. (1996). *Psychologische Personalauswahl. Einführung in die Berufseignungsdiagnostik*. Göttingen: Verlag für Angewandte Psychologie.
- Thielepape, M. (1990). *Wie können die Verwaltungen auf die zurückgehenden Bewerberzahlen reagieren?* DGP-Informationen, 41, S. 67-80.
- Thielepape, M. (2003). *Ergebnisse einer Bewährungskontrolle der DGP-Eignungstests für eine große Krankenpflegeschule in Niedersachsen*. DGP-Informationen, Heft 57, 19-23.

## Anhang

### Berechnung des Zwischenprüfungsergebnisses gem. § 13 APVOgehD (Niedersachsen)

#### 1. Leistungskontrollen aus dem Grundstudium gem. § 6 Abs. 2 und 3 APVOgehD

Studienfachgruppe R	<b>9,00</b>	davon 38 %	<b>3,42</b>
Studienfachgruppe W	<b>8,00</b>	davon 38 %	<b>3,04</b>
Studienfachgruppe SF	<b>12,00</b>	davon 24 %	<b>2,88</b>

Mittelwert der Durchschnittspunktzahlen gem. § 13 Abs. 1 Satz 1 Nr. 2	<b>9,34</b>
---	-------------

#### 2. Prüfungsklausuren gem. § 12 APVOgehD

Stoffgebiet	Fach (Beispiel)	Punkte
1. Arbeit: Studienfachgruppe R	<b>Allgemeines Verwaltungsrecht</b>	<b>5</b>
2. Arbeit: Studienfachgruppe W	<b>Öffentliche Finanzwirtschaft</b>	<b>4</b>
3. Arbeit: Studienfachgruppe R/W/SF	<b>Betriebliche Funktionen</b>	<b>4</b>
Gesamtwert der Prüfungsklausuren		<b>13</b>

#### 3. Ergebnis der Zwischenprüfung gem § 13 APVOgehD

- Die Zwischenprüfung ist bestanden, da mindestens zwei Prüfungsarbeiten mit 5,00 Punkten oder besser bewertet worden sind (§ 13 Abs. 1 Satz Nr. 1)
- Die Zwischenprüfung ist bestanden, da mindestens eine Arbeit mit 5,00 Punkten oder besser bewertet worden ist und der Gesamtwert der Prüfungsarbeiten mindestens 13,00 Punkte beträgt und die Durchschnittspunktzahl nach § 6 Abs. 3 mindestens 9,00 Punkte beträgt (§13 Abs. 1 Satz 1 Nr. 2)
- Die Zwischenprüfung ist nicht bestanden, weil mehr als eine Arbeit mit weniger als 5,00 Punkten bewertet worden ist und der Mittelwert der Leistungskontrollen weniger als 9,00 Punkte bzw. der Gesamtwert der Prüfungsarbeiten weniger als 13,00 Punkte beträgt.

#### Korrespondenzanschrift der Autoren:

Dr. Martin Kersting  
Institut für Psychologie  
der RWTH Aachen  
Jägerstr. 17/19, 52056 Aachen  
eMail: Kersting@psych.rwth-aachen.de

Dipl.-Psych. Michael Thielepape  
Deutsche Gesellschaft  
für Personalwesen e.V.  
Stammestr. 40 D, 30459 Hannover  
eMail: Hannover@dgp.de