

## **Der neue DGP-Leistungstest auf der Basis des Berliner Intelligenzstrukturmodells: Informationen zu ausgewählten Testgütekriterien und zur Normierung<sup>1</sup>**

Martin Kersting & André Beauducel<sup>2</sup>

### **1. Einführung und Überblick**

Seit dem August 1996 setzt die DGP zur Personalauswahl einen neuen Leistungstest ein, welcher auf der Basis des Berliner Intelligenzstrukturmodells ("BIS") von Jäger (1982, 1984) und des entsprechenden Tests (BIS-4-Test, Jäger, Süß und Beauducel, 1997) entwickelt wurde.

Über die Rahmenbedingungen und die Ziele der Testentwicklung wurde im Heft 54 der *DGP-Informationen* berichtet (Kersting, 1995). Das vorliegende Heft skizziert mit zwei Artikeln die bislang mit dem neuen Verfahren gesammelten Erfahrungen: Während Althoff (1997) die praktische Bewährung des Verfahrens in den Vordergrund seines Artikels stellt, geht es an dieser Stelle um den Bericht einiger eher technischer Gütekriterien (Ergebnisse zur Konstruktvalidität, Trennschärfen der Aufgaben, interne Konsistenzen der Skalen) sowie um die Beschreibung der Datenbasis für die Testanalyse. Der beschriebene Datensatz ist zugleich Ausgangspunkt der ab dem August 1997 gültigen (teilweise altersspezifischen) Normierung der Denkaufgaben, über die ebenfalls kurz berichtet wird.

### **2. Kurzbeschreibung des Tests:**

Der sogenannte BIS-r-DGP Test wurde von den Autoren des vorliegenden Beitrags mit dem Ziel entwickelt, im Rahmen der gegebenen testökonomischen Grenzen einer Abbildung des Berliner Intelligenzstrukturmodells ("BIS", Jäger, 1982, 1984) möglichst nahe zu kommen. Der Test umfaßt 38 Denkaufgaben. Tabelle 1 enthält eine Zuordnung der Aufgaben zu den BIS-Zellen sowie eine Erläuterung der Aufgabenabkürzungen (Kurzbeschreibungen der Aufgaben sowie Beispielaufgaben können bei der DGP angefordert werden). Die BIS-Zelle zum "numerischen Einfallsreichtum (EN)" wurde nicht durch Aufgaben repräsentiert, da diese Kombination von Fähigkeiten noch nicht ausreichend etabliert ist.

Dem Großteil der 38 Aufgaben lagen entsprechende Aufgaben des Berliner Intelligenzstruktur-Tests (Form 4, Jäger, Süß und Beauducel, 1997) als Vorlage der Testentwicklung zu Grunde. Allerdings sind die Aufgaben des BIS-r-DGP Tests nicht identisch mit den Aufgaben des BIS-4 Tests. Der neue DGP Test

---

1 In das unter der Leitung des Erstautors stehende Projekt haben sich fast alle Mitarbeiter(innen) der DGP in der einen oder anderen Form aktiv eingebracht. In bezug auf die bisherigen Projektphasen haben sich besonders (in alphabetischer Reihenfolge) Herr Th. Königer, Frau M. Leicher und Frau C. Witkiewicz verdient gemacht.

2 Dr. André Beauducel ist Hochschulassistent am Institut für Klinische, Diagnostische und Differentielle Psychologie der Technischen Universität Dresden, Hohe Str. 53; 01067 Dresden.

umfasst ausschließlich unveröffentlichtes, exklusiv für die Auftraggeber(innen) der *DGP* zur Verfügung gestelltes, Aufgabenmaterial. Insbesondere bei den Aufgaben zur Bearbeitungsgeschwindigkeit unterscheiden sich die Aufgaben des *BIS-r-DGP* Tests von ihren "Vorbildern" im *BIS-4* Test: Während die Testand(inn)en im *BIS-4* Test direkt in die Aufgabenhefte schreiben, erfolgt das Markieren der Lösungen im *BIS-r-DGP* Test auf maschinenlesbaren Antwortbogen. Neun der 38 Aufgaben des neuen Tests haben ihren Ursprung nicht im *BIS-4* Test, sondern in vorherigen *DGP*-Tests: Bei den Aufgaben "DF", "KL", "SL", "AG", "TA", "ZM", "ES", "TX" und "TS" (Abkürzungen siehe Tabelle 1) handelt es sich um Weiterentwicklungen vielfach bewährter *DGP*-Aufgaben. Zwei Merkfähigkeitsaufgaben ("FP" und "OG") wurden vollständig neu konstruiert.

Abgesehen von dem grundsätzlich verschiedenen Aufgabenmaterial unterscheidet sich der *BIS-r-DGP* Test vom *BIS-4* Test durch folgende Merkmale: (1.) Wegfall der Aufgaben zur *BIS*-Zelle "EN", (2.) eine geringere Anzahl von Einfallsreichtumsaufgaben in den übrigen beiden Zellen, (3.) Beschränkung der Auswertung der Aufgaben zum Einfallsreichtum auf den sogenannten "U-Modus<sup>3</sup>", (4.) Verwendung maschinenlesbarer Antwortbogen.

---

3 Im sogenannten "U-Modus" wird lediglich die Anzahl aller instruktionsgemäßen Lösungen gezählt ("U" für Umfang). Die so ausgewerteten Testleistungen sind ein Indikator der *Ideenflüssigkeit*. Mit einer anderen Auswertungstechnik (dem sogenannten "X-Modus") wird ein Indikator für die *Ideenflexibilität* bestimmt, indem zunächst jede Lösung mit Hilfe vorgegebener Kategorienlisten einer Kategorie zugeordnet wird. Anschließend wird als Aufgabenrohwert die Anzahl *unterschiedlicher* Kategorien gezählt. Die *DGP* beschränkt sich aus pragmatischen Gründen zunächst auf eine Auswertung der Testdaten der Dimension "Einfallsreichtum" im "U-Modus". Ob in Zukunft eine Umstellung auf den "X-Modus" erfolgt, wird aufgrund der Ergebnisse der Studien zur Kriteriumsvalidität entschieden.

	<b>F</b>	<b>V</b>	<b>N</b>
<b>B</b>	BD Buchstaben Durchstreichen FI Figuren Vergleichen MG Marigold	KW Wörter Klassifizieren TG Teil-Ganzes UW Unvollständige Wörter	RZ Rechen-Zeichen SE Sechs-Teilbar XG X-Größer
<b>M</b>	FM Firmen-Zeichen FP Figuren-Paare OG Orientierungs- Gedächtnis	PS Phantasiesprache ST Sinnvoller-Text WM Worte-Merken	ZP Zahlen-Paare ZW Zahlen Wieder- erkennen ZZ Zweistellige Zahlen
<b>E</b>	OJ Objekt-Gestaltung ZF Zeichen-Fortsetzen	AM Anwendungs- Möglichkeiten EF Eigenschaften -Fähigkeiten MA Masselon	
<b>K</b>	AN Analogien (figural) AW Abwicklungen BG Bongard DF Drehfiguren FA Figuren-Auswahl	AG Analogien (verbal) KL Wortklassifikation. SL Schlüsse SV Schlüsse- Vergleichen TA Textanalyse	BR Buchstabenreihen ES Ergebnisse Schätzen TS Tabellen/Statistiken TX Textrechnen ZM Zahlenmatrizen

Tabelle 1: Verteilung der Aufgaben auf die Skalen und Zellen

Randspalten und -zeilen: Die BIS-Fähigkeiten:

"K" Verarbeitungskapazität "V" Sprachgebundenes Denken  
 "E" Einfallsreichtum "N" Zahlengebundenes Denken  
 "M" Merkfähigkeit "F" Figural-bildhaftes Denken  
 "B" Bearbeitungsgeschwindigkeit

(Erläuterungen zu den Fähigkeitskomponenten in Jäger (1984), siehe z.B. Kersting, 1995, S. 5)

### 3. Beschreibung der Datenbasis für die Testanalyse- und Normierung

Die folgenden Analysen basieren auf den Daten von 3274 Personen, die im Zeitraum von August 1996 bis April 1997 im Rahmen ihrer Bewerbung für eine Ausbildung zum gehobenen Verwaltungsdienst den BIS-r-DGP Test (Version 1.0) bearbeitet haben. Es handelt sich um Bewerber(innen) für verschiedene allgemeine Verwaltungen sowie zu einem Teil (15,9 %) für eine Finanzverwaltung. Bewerber(innen) für eine Ausbildung beim Auswärtigen Amt fanden in der Analyse und Normierung (siehe Punkt 8) keine Berücksichtigung. Tabelle 2 enthält nähere Angaben zu der analysierten Personengruppe. (Die Datenbasis ist nicht identisch mit der Datenbasis, die Althoff (1997) seinem Bericht zugrundelegt).

Alter	weiblich		männlich		Gesamt
	(Ost)	(West)	(Ost)	(West)	
<b>17-18 Jahre</b>	867 (243)	(624)	344 (172)	(172)	1211 (37%) (415 Ost / 796 West)
<b>19-20 Jahre</b>	472 (137)	(335)	442 (136)	(306)	914 (27,9 %) (273 Ost / 641 West)
<b>21 Jahre und älter</b>	550 (221)	(329)	599 (188)	(411)	1149 (35,1%) (409 Ost / 740 West)
	(Ost: 601)	(West: 1288)	(Ost: 496)	(West: 889)	3274 (1097 Ost [33,5%] / 2177 West [66,5%])
	1889 Frauen (57,7%)		1385 Männer (42,3%)		

Tabelle 2: Informationen zur Datenbasis für die Testanalyse und Normierung

#### 4. Konstruktvalidität

Zunächst wurde geprüft, ob die von Jäger (1982, 1984) ermittelte Intelligenzstruktur sich in den mit dem neuen BIS-r-DGP Test erhobenen Daten aufzeigen läßt, ob sich insbesondere die vier Operationsklassen und die drei Inhaltsklassen nachweisen lassen. Zur Prüfung wurden die 38 Aufgaben des neuen BIS-r-DGP Tests bimodal klassifiziert und anschließend zunächst zu operationshomogenen und dann zu inhaltshomogenen Bündeln zusammengefaßt (siehe Jäger, 1982). Mit diesen Variablenbündeln wurden Faktoranalysen (Hauptachsenanalysen mit Kommunalitäteniteration und anschließender Varimax Rotation) durchgeführt. (Informationen über die Variablenbündel dieser Faktorenanalyse: siehe Tabelle 6 im Anhang.

Die in den Tabellen 3a und 3b dargestellten Varimax-rotierten Ladungsmatrizen und die Eigenwertverläufe der unrotierten Hauptachsenanalysen zeigen, daß sowohl die vier Operationsklassen als auch die drei Inhaltsklassen in einer prägnanten Einfachstruktur erscheinen. Obwohl der Faktor "Einfallsreichtum" durch den Verzicht auf Testaufgaben zum numerischen Einfallsreichtum nur schwach besetzt ist und die Einfallsreichtumsaufgaben lediglich im sogenannten "U-Modus" (siehe Fußnote 3) ausgewertet wurden, stimmt die Struktur der Testleistungen mit der Modellstruktur überein.

	FACTOR 1 K	FACTOR 2 B	FACTOR 3 M	FACTOR 4 E	Communality h <sup>2</sup>		
B1	.20	<b>.75</b>	.21	.20	.69		
B2	.26	<b>.78</b>	.19	.18	.75		
B3	.17	<b>.77</b>	.20	.20	.70		
M1	.18	.12	<b>.67</b>	.02	.50		
M2	.30	.24	<b>.66</b>	.10	.59		
M3	.18	.19	<b>.71</b>	.12	.59		
E1	.09	.17	.06	<b>.74</b>	.58		
E2	.05	.22	.11	<b>.74</b>	.62		
K1	<b>.62</b>	.16	.10	<b>.11</b>	.43		
K2	<b>.62</b>	.11	.09	.05	.41		
K3	<b>.69</b>	.14	.20	.05	.53		
K4	<b>.61</b>	.12	.15	.03	.41		
K5	<b>.63</b>	.13	.22	.03	.46		
Gesamtvarianz	17.1	16,0	12,7	9,6	55,4		
%Aufgekl.Var.	30.9	28,9	22,9	17,3	100.0		
Eigenvalue:	4.83	1.77	1.30	1.04	.61	.59	.55

Tabelle 3a: Faktorenanalyse operationshomogener Bündel (Varimax-Lösung). Erläuterung der Variablenbündel: siehe Tabelle 6

	FACTOR 1 N	FACTOR 2 F	FACTOR 3 V	Communality h <sup>2</sup>			
F1	.22	<b>.61</b>	.26	.50			
F2	.26	<b>.66</b>	.33	.62			
F3	.25	<b>.66</b>	.26	.57			
V1	.24	.31	<b>.61</b>	.53			
V2	.32	.30	<b>.67</b>	.64			
V3	.32	.35	<b>.64</b>	.63			
N1	<b>.70</b>	.22	.26	.61			
N2	<b>.63</b>	.29	.29	.56			
N3	<b>.70</b>	.22	.21	.58			
%Gesamtvarianz	20.2	19.3	18.6	58.1			
%Aufgekl. Var.	34.8	33.2	32.0	100.0			
Eigenvalue:	4.75	.98	.76	.48	.45	.43	.41

Tabelle 3b: Faktorenanalyse inhaltshomogener Bündel (Varimax-Lösung)  
Erläuterung der Variablenbündel: siehe Tabelle 6

### **5. Potentielle Umstellung auf Faktorwerte als Leistungsindikatoren**

Das neue Testverfahren ermöglicht es, ab August 1997 alternativ zu den bisherigen Skalenleistungen Faktorwerte zur Berufseignungsdiagnose zu verwenden. Im Unterschied zu den Skalen, die jeweils nur ca. ein Viertel (operative Skalen) bzw. ein Drittel (inhaltliche Skalen) der Testaufgaben berücksichtigen, werden bei der Bildung der Faktorwerte jeweils alle 38 Aufgaben mit unterschiedlichen Gewichten einbezogen. Die Faktorwerte orthogonaler Faktoren sind unabhängiger voneinander als die Skalen und schaffen daher bessere Voraussetzungen für berufsspezifische Profilanalysen. Während die Skalen im vorliegenden Datensatz im Bereich von  $r=.20$  bis  $r=.45$  (Operationen) bzw.  $r=.60$  bis  $r=.63$  (Inhalte) miteinander korrelieren, korrelieren die aufgrund von Hauptkomponentenanalysen ermittelten Faktorwerte der orthogonalen Faktoren innerhalb der Operations- und Inhaltsklassen zu  $.00$  miteinander. Bei aller Unterschiedlichkeit zwischen den Interkorrelationen der Skalen und Faktorwerte gilt jedoch, daß die Skalen hoch ( $r=.89$  bis  $r=.96$ ) mit den entsprechenden Faktorwerten korrelieren.

Zur Prüfung der Stichprobenunabhängigkeit der Faktorwerte wurde eine Kreuzvalidierung durchgeführt. Dazu wurde der Datensatz per Zufall in zwei Hälften geteilt. Die Faktorwerte der ersten Zufalls-Teilstichprobe wurden dann mit den über die regressionsanalytisch geschätzten Gewichte der zweiten Teilstichprobe geschätzten entsprechenden Faktorwerte korreliert (und umgekehrt). Alle Korrelationen erreichten Werte größer als  $r=.994$ .

### **6. Korrelationen der Aufgaben mit den Faktorwerten (Trennschärfen)**

Als ein Indikator für die Qualität einzelner Aufgaben kann deren Korrelation mit den Faktorwerten (Trennschärfe) gelten. Der Trennschärfekoeffizient gibt an, wie gut eine Aufgabe das Gesamtergebnis in einer Leistungsdimension repräsentiert. Da im Berliner Intelligenzstrukturmodell jede Aufgabe zwei Leistungsdimensionen (nämlich eine operative und eine inhaltliche Dimension) repräsentiert, finden sich entsprechend jeweils zwei Werte für die Trennschärfe der jeweiligen Aufgabe. (In der Tabelle 4 jeweils fett gedruckt.). Die Korrelationen sind aufgrund der fehlenden part-whole Korrektur überschätzt. Berichtet werden die Korrelationen der standardisierten Leistungen in den Aufgaben (Punktwerte) mit den Faktorwerten einer Hauptkomponentenanalyse (mit anschließender Varimax Rotation). Nach Ebel (1972) kann bei einer Trennschärfe ab  $0.40$  von einer -unter diesem Aspekt- guten Testaufgabe gesprochen werden. Nur acht (10,5%) der 76 Aufgaben-Trennschärfen verfehlen (knapp) diesen Wert. Diese acht Aufgaben erzielen gleichviel noch Trennschärfen über  $r=0.30$  und gelten damit nach der Klassifikation von Ebel (ebd.) als "brauchbare" Aufgaben, bei denen eine Verbesserung möglich ist. U.a. diese Verbesserung ist mit den Modifikationen von der Testversion "1.0" zur Testversion "1.1" intendiert. Ebenso erfreulich wie die hohen Korrelationen zwischen den Aufgaben und den ihnen zugeordneten Leistungsdimensionen ist die Tatsache, daß nur eine (0,5%) der 190 möglichen nicht-intendierten Korrelationen (in der Tabelle 4 nicht fett gedruckt) eine Höhe von  $r=0.30$  erreicht.

Aufgaben	K	B	M	E	Faktorwerte	N	V	F	
FI	.13	<b>.69</b>	.17	.15		.25	.26	<b>.53</b>	FI
BD	-.00	<b>.65</b>	.11	.18		.11	.16	<b>.58</b>	BD
MG	.07	<b>.62</b>	.16	.15		.11	.25	<b>.54</b>	MG
TG	.17	<b>.63</b>	.13	.10		.22	<b>.50</b>	.22	TG
KW	.13	<b>.66</b>	.12	.15		.18	<b>.50</b>	.28	KW
UW	.14	<b>.64</b>	.16	.10		.29	<b>.42</b>	.25	UW
XG	.29	<b>.57</b>	.11	.06		<b>.62</b>	.24	.13	XG
SE	.26	<b>.58</b>	.14	.09		<b>.58</b>	.23	.21	SE
RZ	.10	<b>.71</b>	.12	.12		<b>.42</b>	.27	.30	RZ
FM	.19	.20	<b>.57</b>	.12		.15	.27	<b>.53</b>	FM
OG	.12	.03	<b>.49</b>	-.01		.13	.01	<b>.42</b>	OG
FP	.19	.16	<b>.57</b>	.12		.16	.21	<b>.53</b>	FP
ST	.21	.22	<b>.51</b>	.04		.26	<b>.50</b>	.11	ST
WM	.05	.05	<b>.58</b>	.03		.08	<b>.40</b>	.14	WM
PS	.11	.13	<b>.65</b>	.07		.20	<b>.45</b>	.19	PS
ZP	.12	.11	<b>.60</b>	-.01		<b>.54</b>	.08	.14	ZP
ZZ	.20	.17	<b>.51</b>	.02		<b>.51</b>	.15	.18	ZZ
ZW	.03	.08	<b>.52</b>	.07		<b>.44</b>	.11	.07	ZW
OJ	.04	.13	.05	.71		-.01	.22	<b>.40</b>	OJ
ZF	-.08	.19	.06	.72		-.04	.20	<b>.42</b>	ZF
EF	.11	.12	.03	<b>.73</b>		.01	<b>.46</b>	.19	EF
MA	.15	.16	.12	<b>.53</b>		.07	<b>.46</b>	.19	MA
AM	.09	.13	.06	<b>.64</b>		-.02	<b>.48</b>	.20	AM
AN	<b>.47</b>	.08	.16	.02		.21	.11	<b>.37</b>	AN
AW	<b>.44</b>	.07	.07	.06		.15	.08	<b>.35</b>	AW
BG	<b>.37</b>	.06	.11	.03		.13	.07	<b>.32</b>	BG
FA	<b>.38</b>	.06	.06	.02		.12	.02	<b>.33</b>	FA
DF	<b>.48</b>	.04	.11	-.01		.20	.01	<b>.36</b>	DF
KL	<b>.45</b>	.08	.04	.06		.09	<b>.40</b>	.06	KL
SV	<b>.50</b>	.19	.04	.16		.20	<b>.45</b>	.11	SV
SL	<b>.52</b>	.11	.19	.08		.22	<b>.40</b>	.18	SL
TA	<b>.48</b>	.07	.06	.05		.16	<b>.52</b>	-.07	TA
AG	<b>.53</b>	.08	.17	.04		.21	<b>.38</b>	.14	AG
BR	<b>.45</b>	.13	.21	-.00		<b>.39</b>	.14	.19	BR
ZM	<b>.53</b>	.10	.10	-.06		<b>.60</b>	.05	.05	ZM
ES	<b>.58</b>	.11	.08	-.01		<b>.54</b>	.10	.08	ES
TS	<b>.50</b>	.08	.03	.02		<b>.41</b>	.12	.07	TS
TX	<b>.62</b>	.07	.00	.02		<b>.49</b>	.15	.02	TX

Tabelle 4: Korrelationen der Aufgaben mit den Faktorwerten  
Erläuterungen der Abkürzungen: siehe Tabelle 1

## 7. Interne Konsistenz

In der Tabelle 5 werden die Maße der internen Konsistenz der sieben Skalen und der Gesamtskala berichtet. Bei der Interpretation der Werte ist zu berücksichtigen, daß die Modellannahmen die Konstruktion heterogener Skalen verlangen. Die im folgenden angewandte Methode der Bestimmung der internen Konsistenz nach Cronbach (alpha-Koeffizient) geht von eindimensionalen Tests aus. Eine Interpretation der Werte als Reliabilitätsmaße führt zu einer Unterschätzung der Reliabilität der bimodalen Testskalen. Sinnvoller interpretierbar sind die Homogenitätskoeffizienten auf der Basis der operations- bzw. inhaltshomogenen Bündel. Deren Anzahl ist mit zwei (bei "E") bis fünf allerdings sehr gering.

Skalen:	K	M	B	E	V	N	F	AI
Auf Basis d. Aufgab.	.79	.77	.85	.73	.78	.77	.73	.89
Auf Basis der Bündel	.80	.79	.88	.75	.82	.80	.80	.94

Tabelle 5: Interne Konsistenzen der Skalen; Abkürzungen: siehe Tabelle 1

## 8. (Altersspezifische) Normierung der Denkaufgaben ab August 1997

Aufgrund der Testanalysen wurden einige Instruktionen und Aufgaben geringfügig optimiert, die neue Testversion "1.1" kommt ab dem August 1997 zum Einsatz. Da die Auswirkungen der geringfügigen Modifikationen vorhersehbar erscheinen, wurde die (teilweise altersspezifische) Normierung der Testversion "1.1" ebenfalls anhand der hier dargestellten Datenbasis vorgenommen. Die Gültigkeit der so bestimmten Normen für die Version "1.1" wird zu Saisonbeginn noch überprüft.

Die folgenden Abschnitte informieren über das Vorgehen bei der Normierung: Zunächst erfolgte eine Normierung der 38 Aufgaben. Dazu wurden für alle Rohwerte der Einzelaufgaben auf der Basis der Gesamtstichprobe (N = 3274) die Prozentränge ermittelt. Die Prozentränge wurden dann in entsprechende z-Werte der Standardnormalverteilung umgelesen (T-Werte, nach Terman benannt, siehe McCall, 1939). Die z-Werte wurden mit 10 multipliziert und zum Ergebnis wurde der Wert 100 addiert, so daß Standardwerte der Aufgaben gebildet wurden, die im folgenden als Punktwerte bezeichnet werden, da die Aufgabenleistungen nicht diagnostisch interpretiert werden.

Mit den Punktwerten der Aufgaben wurden operations- und inhaltshomogene Bündel gebildet, mit denen sukzessive Hauptkomponentenanalysen mit anschließender Varimax-Rotation zur Berechnung der Faktorwerte durchgeführt wurden. Auf die ebenfalls häufig eingesetzte Hauptachsenanalyse wurde hier verzichtet, da diese auch bei orthogonaler Rotation der Faktoren zu leichten Interkorrelationen der Faktorwerte führt. Schließlich wurden die sieben Regressionsgleichungen zur Schätzung der Faktorwerte aus den Punktwerten



der jeweils 38 Aufgaben berechnet, d.h. es wurden für jede Komponente 38 B-Gewichte (Slope) und eine Konstante (Intercept) ermittelt. Anschließend wurde eine Kreuzvalidierung der B-Gewichte und der Konstanten anhand einer Zufallsteilung der Testand(innen)gruppe durchgeführt, die zu extrem hohen multiplen Korrelationen führte, was für eine entsprechende Robustheit der Regressionsgleichungen spricht (siehe oben, Punkt 5). Die B-Gewichte und Konstanten wurden in Tabellen abgelegt, die es ermöglichen, die Faktorwerte von Testand(inn)en auch ohne erneute Faktorenanalysen zu ermitteln.

Als nächstes wurden die Prozentränge für die Faktorwerte der sieben Komponenten (für die Gesamtgruppe) berechnet und (wie bei den Aufgaben) in entsprechende z-Werte der Standardnormalverteilung umgelesen. Dann wurden die z-Werte in Standardwerte der Faktorwerte durch Lineartransformation umgerechnet. Schließlich wurden die Zuordnungstabellen von Faktorwerten zu Standardwerten erstellt und mit Hilfe linearer Regressionen geglättet.

Zur Überprüfung eventuell notwendiger Altersnormierungen wurden die Faktorwerte zunächst nochmal anhand der geglätteten Zuordnungstabellen in Standardwerte umgelesen. Es ergaben sich bedeutsame Produkt-Moment-Korrelationen der so standardisierten Faktorwerte mit dem Alter für die "Merkfähigkeit (M)", den "Einfallsreichtum (E)" sowie für die "verbalen (V)" und "figural-anschauungsgebundenen Fähigkeiten (F)". Für "M" und "F" gab es einen negativen linearen Trend (mit zunehmendem Alter nahm die Testleistung ab). Für "E" und "V" ergab sich hingegen ein positiver Trend. Die Linearität in den Trends war so eindeutig, daß Abweichungen der Alterseffekte von der Linearität als Zufallseffekte interpretiert und bei der Altersnormierung nicht berücksichtigt wurden. Für die vier Komponenten mit einem bedeutsamen Alterstrend wurde dieser in den Standardwerten der Faktorwerte korrigiert, indem die B-Gewichte und Konstanten der linearen Regression der jeweiligen Faktorwerte über das Alter herangezogen wurden, um eine Lineartransformation herzustellen, die den Alterstrend exakt kompensiert. Die Verwendung der Lineartransformation entspricht der Verwendung von Normtabellen mit entsprechend extrapolierten Werten für jeden Altersjahrgang zwischen 17 und 32 Jahren. Abbildung 1 veranschaulicht die Effekte der Altersnormierung am Beispiel der "Merkfähigkeitsleistungen".

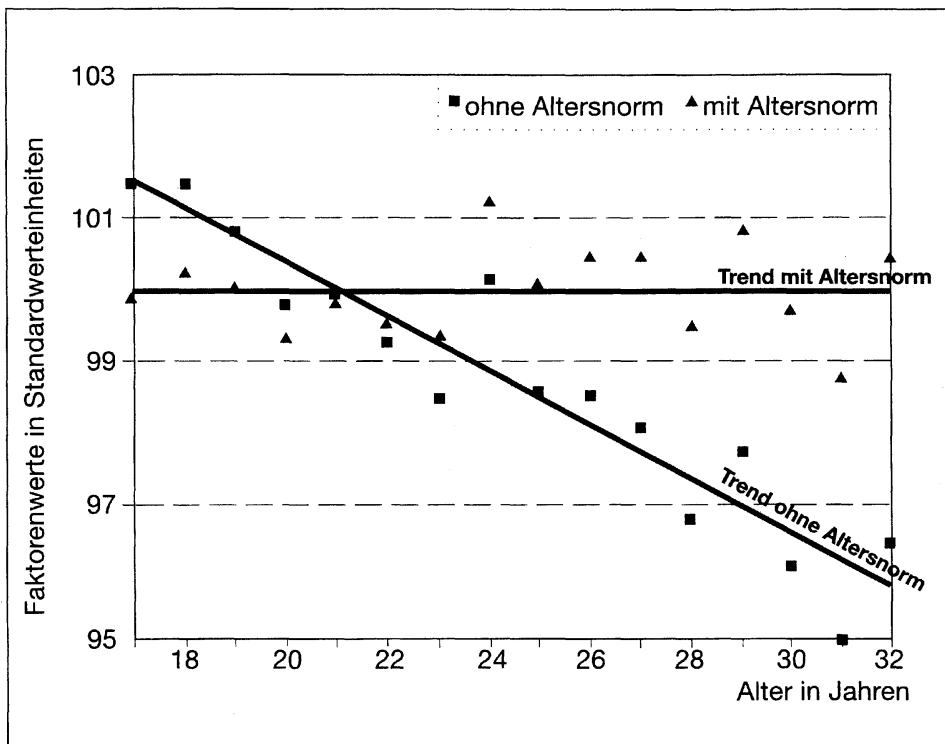


Abbildung 1: Altersspezifische Merkfähigkeitsleistungen im BIS-r-DGP Test  
 Faktorwerte in Standardwerteinheiten mit und ohne Altersnormierung; N=3274,  
 BIS-r-DGP Test Version "1.0"

### 9. Fazit

Bereits die erste Version des neuen DGP-Testverfahrens für den gehobenen Dienst hat sich -nach Maßgabe der hier berichteten Kennwerte- bewährt. Ab August 1997 setzt die DGP die geringfügig überarbeitete Version "1.1" des Tests ein. Die neue EDV-gestützte Auswertung der Tests und die damit einhergehende Datenspeicherung erlauben es der DGP, den für ihre Kund(inn)en großen Vorteil einer empirisch umfassenden Datenbasis zur kontinuierlichen Qualitätssicherung und -optimierung zu nutzen. Dieser Datenpool hat es ermöglicht, ab der nächsten Testsaision für einige Leistungskomponenten für jeden Altersjahrgang zwischen 17 und 32 Jahren Altersnormen anzuwenden. Als nächster Schritt der Testevaluation und -optimierung erfolgt u.a. die Auswertung der bereits durchgeführten Studien zur Kriteriumsvalidität des neuen BIS-r-DGP Tests.

## 10. 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). Hannover: Deutsche Gesellschaft für Personalwesen, 55, Seite 59-68.
- Ebel, R.L. (1972). Essentials of educational measurement. New Jersey: Englewood Cliffs.
- Jäger, A.O. (1982). Mehrmodale Klassifikation von Intelligenzleistungen. Experimentell kontrollierte Weiterentwicklung eines deskriptiven Intelligenzstrukturmodells. Diagnostica, 28, 195-226.
- Jäger, A.O. (1984). Intelligenzstrukturforschung: Konkurrierende Modelle, neue Entwicklungen, Perspektiven. Psychologische Rundschau, 35, 21-35.
- Jäger, A.O., Süß, H.-M. & Beauducel, A. (1997). Berliner Intelligenzstruktur-Test. Form 4. Handanweisung. Göttingen: Hogrefe.
- Kersting, M. (1995). Entwicklung eines neuen Leistungstests zur Personalauswahl von Bewerber(innen) für die Laufbahn des gehobenen Dienstes. (DGP-Informationen). Hannover: Deutsche Gesellschaft für Personalwesen, 54, 3-20.
- McCall, W.A. (1939). Measurement. New York.

## Anhang

A) Vorbündelung:	B) Operationshomogene Aufgabenbündel	C) Inhaltshomogene Aufgabenbündel
e1v= MA+AM. F1k= AW+BG. F2k= FA+DF. v1k= SL+KL. v2k= AG+SV. n1k= BR+ES. n2k= TX+TS.	b1= BD+UW+SE. b2= FI+KW+XG. b3= MG+TG+RZ.  m1= OG+WM+ZP. m2= FP+ST+ZZ. m3= FM+PS+ZW.  e1= OJ+EF. e2= ZF +e1v.  k1= AW+SV+TS. k2= FA+KL+TX. k3= AN+SL+ES. k4= BG+TA+ZM. k5= DF+AG+BR.	f1= MG+OG+ZF +F2k. f2= FI+FM+OJ +AN. f3= BD+FP+F1k.  V1= KW+WM+AM +TA. v2= TG+ST+MA +V1k. v3= UW+PS+EF+V2k.  n1= XG+ZW+N1k. n2= RZ+ZZ+N2k. n3= SE+ZP+ZM.

Tabelle 6: Informationen über die Variablenbündel für den BIS-r-DGP Test, Version 1.0 (Erläuterungen der Aufgabenkürzel: siehe Tabelle 1)