

Entwicklung eines neuen Leistungstests zur Personalauswahl von Bewerber(innen) für die Laufbahn des gehobenen Dienstes¹

Martin Kersting

Die *Deutsche Gesellschaft für Personalwesen* hat sich in Hinblick auf die Qualität ihrer eignungsdiagnostischen Untersuchungen stets hohe Standards gesetzt. Um diesem Qualitätsniveau dauerhaft gerecht werden zu können, müssen die Verfahren einerseits immer wieder einer kritischen Prüfung unterzogen und andererseits - auf Grundlage der Prüfungsergebnisse und unter Einbeziehung der jeweils aktuellen theoretischen und methodischen Kenntnisse - kontinuierlich weiterentwickelt werden. In der Vergangenheit haben sich solche Entwicklungsarbeiten der *DGP* auf dem Gebiet der psychologischen Diagnostik oft weit über die *DGP* hinaus ausgewirkt. So wurde etwa der mit mittlerweile über vier Millionen Anwendungen weitverbreitete „*Intelligenz-Struktur-Test (IST)*“ von AMTHAUER ebenso in der *DGP* entwickelt und erprobt wie der in der Berufsberatung weitverbreitete „*Berufs-Interessen Test (BIT)*“ von IRLE. Der „*Wilde Intelligenz-Test (WIT)*“ von JÄGER und ALTHOFF (1993), der lange Zeit die Grundlage der *DGP* Eignungsuntersuchungen darstellte, wurde 1983 mit großem Erfolg veröffentlicht, nachdem für die eigene Arbeit neue - unveröffentlichte - Verfahren entwickelt worden waren.

Zur Zeit wird in der *DGP* auf der Basis des Berliner Intelligenzstrukturmodells (BIS) von JÄGER (1982, 1984) und des entsprechenden Testverfahrens ein neuer Intelligenztest für die von der *DGP* durchgeführten Auswahlverfahren für Bewerber(innen) des gehobenen Dienstes entwickelt. PROF. DR. A.O. JÄGER war von 1955-1968 Leiter der *DGP* und hat dort sein -von der *DGP* und der Deutschen Forschungsgemeinschaft gefördertes - Projekt „Dimensionen der Intelligenz“ (JÄGER, 1967) verwirklicht. Die in der *DGP* begonnenen und in Berlin weiterverfolgten Forschungen haben zur Entwicklung des national und international (z.B. CAROLL, 1993, S. 64 f.; GUSTAFSSON, 1988, S. 52 f.) respektierten Berliner Intelligenzstrukturmodells (BIS) geführt. Im folgenden wird eine knappe Darstellung des Modells und seiner Bewährung gegeben. Diese Darstellung orientiert sich an der vorhandenen Literatur zum Berliner Intelligenzstrukturmodell (die für die an einer Vertiefung interessierten Leser(innen) auch genannt wird) und an der bislang unveröffentlichten Handanweisung zum Berliner Intelligenzstruktur-Test (JÄGER, SÜSS & BEAUDUCEL in Druck). Unter anderem auf der Grundlage des Modells werden die Ziele und die Möglichkeiten des neu zu entwickelnden Testes sowie einige Rahmenbedingungen der zukünftigen Leistungsdiagnostik - wie z.B. die Testdauer - erläutert. Der Projektbericht endet mit der Darstellung des Entwicklungsprozesses und dem Bericht einiger Ergebnisse zur Akzeptanz des neuen Tests bei den Bewerber(inne)n.

¹ In das - unter der Leitung des Autors stehende - Projekt haben sich fast alle Mitarbeiter(innen) der *DGP* in der ein oder anderen Form aktiv eingebracht. In bezug auf die bisherigen Projektphasen ist besonders FRAU S. OLBRICH für die umfangreichen Dateneingaben und vor allem FRAU C. WITKIEWICZ für die Realisierung und Gestaltung der Texte und Graphiken des neuen Tests sowie für die Durchführung etlicher Untersuchungen zu danken.

1. Das Berliner Intelligenzstrukturmodell (BIS) nach Jäger

Über die Frage, welche Verhaltensweisen eine „intelligente“ Person auszeichnen, läßt sich trefflich streiten. Nicht nur Laien, sondern auch Wissenschaftler(innen) haben unterschiedliche - und kontroverse - Vorstellungen von der Intelligenz. Das Berliner Intelligenzstrukturmodell wurde mit dem Ziel entwickelt, verschiedene - konkurrierende - Ansätze der Intelligenzforschung *in einem Modell zu integrieren* und somit bestehende Differenzen zu überwinden. Die mannigfaltigen Vorstellungen über intelligenzspezifische Leistungen finden u.a. in der Vielzahl unterschiedlicher Intelligenztestaufgaben ihren Ausdruck. Ausgangspunkt der Berliner Modellentwicklung war die Inventarisierung aller in der Literatur zur Intelligenz- und Kreativitätsforschung vorfindbaren 2.000 Aufgaben. Durch dieses Fundament wurde gewährleistet, daß das Modell alle Intelligenzaufgaben und -leistungen repräsentiert. Von den 2.000 Aufgaben wurden - unter Beibehaltung der Vielfalt der Anforderungen und der Markiervariablen für verschiedene Intelligenzstrukturmodelle - 191 typische Aufgaben, die sich 98 Aufgabenkategorien zuordnen ließen, ausgewählt und in ersten Studien eingesetzt. Die Analysen der Untersuchungsergebnisse stellten den empirischen Ausgangspunkt für die Modellentwicklung (Jäger, 1982, 1984) dar. Das aktuelle Modell ist in der Abbildung 1 dargestellt.

Intelligenz läßt sich demnach auf unterschiedlichen Generalitätsebenen beschreiben (Hierarchieannahme). Auf der generellsten Ebene steht die *Allgemeine Intelligenz* als das Integral aller Leistungskomponenten, auf der Ebene darunter stehen sieben intellektuelle Fähigkeiten, die sich unter zwei Aspekten oder Modalitäten, nämlich nach Operationen einerseits und Inhalten andererseits, klassifizieren lassen (Bimodalitätsannahme). Bei den operativen Fähigkeiten handelt es sich um die Verarbeitungskapazität („K“), den Einfallsreichtum („E“), die Bearbeitungsgeschwindigkeit („B“) und die Merkfähigkeit („M“). In der Modalität Inhalte werden das sprachgebundene („V“), das zahlengebundene („N“) und das anschauungsgebundene, figural-bildhafte („F“) Denken unterschieden. Eine ergänzende Skizzierung dieser Fähigkeiten wird in der Tabelle 1 gegeben.

Abbildung 1:
Das Berliner Intelligenzstrukturmodell (Jäger, 1984)

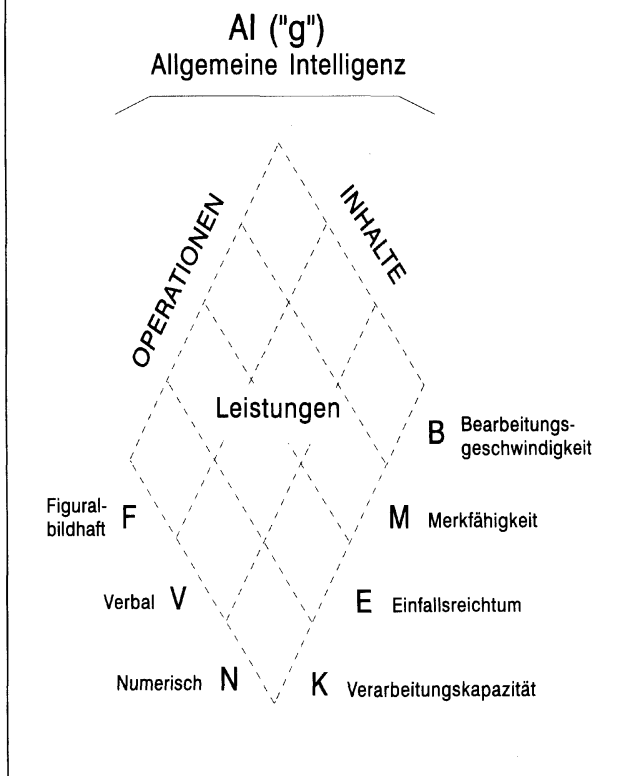


Tabelle 1: Beschreibung der sieben intellektuellen Fähigkeiten nach JÄGER
(Beschreibungen aus JÄGER, 1984, S. 30 f.)

1. Operative Fähigkeiten		
K	Verarbeitungskapazität	Verarbeitung komplexer Informationen bei Aufgaben, die nicht auf Anhieb zu lösen sind, sondern Heranziehen, Verfügbarhalten, vielfältiges Beziehungsstiften, formallogisch exaktes Denken und sachgerechtes Beurteilen von Informationen erfordern.
E	Einfallsreichtum	Flüssige, flexible und auch originelle Ideenproduktion, die Verfügbarkeit vielfältiger Informationen, Reichtum an Vorstellungen und das Sehen vieler verschiedener Seiten, Varianten, Gründe und Möglichkeiten von Gegenständen und Problemen voraussetzt, wobei es um problemorientierte Lösungen geht, nicht um ein ungesteuertes Luxurieren der Phantasie.
M	Merkfähigkeit	Aktives Einprägen und kurz- oder mittelfristiges Wiedererkennen oder Reproduzieren von verbalem, numerischem oder figural-bildhaftem Material.
B	Bearbeitungsgeschwindigkeit	Arbeitstempo, Auffassungsleichtigkeit und Konzentrationskraft beim Lösen einfacher strukturierter Aufgaben von geringem Schwierigkeitsniveau.
2. Inhaltsgebundene Fähigkeiten		
V	Sprachgebundenes Denken	Einheitsstiftendes Merkmal ist hier das Beziehungssystem Sprache. Ein dem Grad seiner Aneignung und Verfügbarkeit entsprechendes Fähigkeitsbündel scheint bei allen sprachgebundenen Operationen mitbestimmend zu sein.
N	Zahlengebundenes Denken	Analog zu Verbal kann hier der Grad der Aneignung und Verfügbarkeit des Beziehungssystems Zahlen als einheitsstiftendes Merkmal interpretiert (...) werden.
F	Figural-bildhaftes Denken	Diese Einheit ist zunächst durch die benannte Gemeinsamkeit des Aufgabematerials charakterisiert.

Entscheidend ist, daß an einer Intelligenzleistung (z.B. dem Lösen einer Aufgabe) zwar alle genannten Fähigkeitskomponenten beteiligt sind, aber für bestimmte Leistungen spezifischen (und zwar jeweils einer operationsspezifischen und einer inhaltspezifischen) Fähigkeit(en) ein besonderes hohes Gewicht beikommt. Jede Intelligenzleistung ist somit ein Indikator für drei Fähigkeiten, für die „Allgemeine Intelligenz“, für eine operative und für eine inhaltsgebundene intellektuelle Fähigkeit.

Die Ebene der Leistungen ist *unterhalb* der Konstruktebene angesiedelt. Die Leistungen sind Fähigkeitsindikatoren, die sich den zwölf Modellzellen in Abbildung 1 zuordnen lassen.

1.1 Geltungsbereiche

The proof of the pudding is in the eating, lautet ein englisches Sprichwort. Entscheidend für die Qualität eines Modells ist seine empirische Bewährung. Werden von dem Modell alle in den Bedeutungshof des wissenschaftlichen Begriffs „Intelligenz“ fallenden Fähigkeiten - operationalisiert über die Leistungen in unterschiedlichsten Intelligenzaufgaben - erfaßt? Für welche Personengruppen gilt das Modell? Ist die postulierte Intelligenzstruktur zeitlich stabil? Ist die Struktur womöglich nur ein Artefakt, das aus der Anwendung bestimmter statistischer Strukturanalysetechniken resultiert? Über zahlreiche unabhängige Untersuchungen zur empirischen Klärung dieser Fragen nach den Geltungsbereichen des Berliner Intelligenzstrukturmodells berichten SÜSS, OBERAUER und JÄGER (1988).

Demnach konnte mehrfach der empirische Nachweis erbracht werden, daß die im Berliner Modell formulierte Intelligenzstruktur nicht nur bei der Verwendung von spezifischen (z.B. die in den eigenen Untersuchungen eingesetzten), sondern auch beim Einsatz zahlreicher weiterer (bei der Modellkonstruktion nicht verwandter) Intelligenzaufgaben gilt, das Modell also eine hochgradige *Generalität* aufweist. Das bedeutet, daß sich andere Intelligenzaufgaben in das Modell klassifizieren lassen (siehe z.B. JÄGER & TESCH-RÖMER, 1988; SCHMIDT, 1993). Dies gilt nicht für die zunächst explizit ausgesparten Bereiche der praktischen und sozialen Intelligenz.

In einer beachtlichen Anzahl von Studien wurde die *Universalität* des Modells ausgelotet, d.h. es wurde untersucht, für welche Personengruppen das Modell Geltung beanspruchen kann. Neben zahlreichen Untersuchungen mit Oberstufenschüler(inne)n (z.B. SCHMIDT, BROCKE, JÄGER, DOLL & KÖNIG, 1986) konnte das Modell u.a. auch bei Berufsfachschüler(inne)n (JÄGER et al., 1988), bei Azubis (Schmidt, 1983) und bei den von der DGP geprüften Bewerber(inne)n für eine Ausbildung zum gehobenen nicht-technischen Verwaltungsdienst repliziert werden (SCHMIDT, 1986). Es liegen außerdem Modellbestätigungen für zwei andere Kulturräume, nämlich für brasilianische Schüler(innen) und Student(inn)en (KLEINE & JÄGER, 1987) sowie für chilenische Student(inn)en (ROSAS, 1990) vor.

Auch die *zeitliche Stabilität* des Modells kann als gesichert gelten. Bei Wiederholungsuntersuchungen konnte das Modell nach vier Wochen (SPERBER, WÖRPEL, JÄGER & PFISTERS, 1985), nach einem Jahr (SÜSS, KERSTING und OBERAUER, 1991) und nach vier Jahren (JÄGER, 1982) repliziert werden.

Methodenstudien ergaben außerdem, daß der Nachweis der im Modell postulierten Struktur nicht von der Anwendung einer spezifischen statistischen Analysemethode abhängt (z.B. WITTMANN, 1982; SCHMIDT, 1984; PFISTER & JÄGER, 1992).

1.2 Zur Vorhersagekraft der modellkonformen Tests

Dem Bericht von Süss et al. (1988) zufolge konnten mit Tests zum Berliner Intelligenzstrukturmodell Schulleistungen (z.B. WITTMANN und MATT, 1986), Studienleistungen (JÄGER & KLEINE, 1989; ROSAS, 1990), die Steigerung der Schachleistungen² (DOLL & MAYR, 1987) und Steuerungsleistungen bei komplexen Problemen (z.B. HÖRMANN & THOMAS, 1989; SÜSS, OBERAUER & KERSTING, 1993) treffsicher vorhergesagt werden. Dabei kam den einzelnen Fähigkeiten eine differenzierte Vorhersagekraft zu, wobei sich insbesondere die Verarbeitungskapazität immer wieder als ein bedeutsamer Vorhersagefaktor erwies.

Besonders interessant für die *DGP* ist eine Studie zur Vorhersagbarkeit des Ausbildungserfolges im ersten Studienabschnitt an einer Fachhochschule für öffentliche Verwaltung durch die Leistungen in einem Test zum Berliner Intelligenzstrukturmodell, über die SCHMIDT (1987) in den *DGP-Informationen* berichtete. Auch hier erwies sich insbesondere die Verarbeitungskapazität als prognostisch bedeutsam. Darüber hinaus zeigte sich aber, daß auch andere Komponenten des Berliner Intelligenzstrukturmodells Beiträge zur Vorhersage des Studienerfolgs leisten konnten (und sei es als Suppressoren).

2. Entwicklung eines neuen DGP-Leistungstests auf Grundlage des Berliner Intelligenzstrukturmodells

2.1 Zielsetzungen

Betrachtet man die gegenwärtige Leistungsdiagnostik der *DGP* vor dem Hintergrund des Berliner Intelligenzstrukturmodells, zeigen sich insbesondere zwei Ansatzpunkte für eine qualitätssteigernde Neu- und Weiterentwicklung der *DGP* Testverfahren. Erstens werden zur Zeit einige intellektuelle Fähigkeiten nicht oder nicht ausreichend durch Testaufgaben repräsentiert und zweitens wird die Analyse der Leistungen der Bimodalitätsannahme nicht gerecht.

Ad 1: Repräsentativere Erfassung der intellektuellen Fähigkeiten

Einige intellektuelle Fähigkeiten - wie z.B. der Einfallsreichtum als eine zentrale Komponente der Kreativität - werden mit dem aktuellen Testverfahren der *DGP* nicht erfaßt. Andere Komponenten - wie z.B. das figural-bildhafte Denken oder die Merkfähigkeit - sind in den zur Zeit angewandten Aufgaben unterrepräsentiert. Schließlich sind die von den Bewerber(inne)n bei den *DGP* - Aufgaben zum Arbeitsverhalten (z.B. die Postaufgabe) erzielten Leistungen vermutlich multipel bedingt (Bearbeitungsgeschwindigkeit und Verarbeitungskapazität, siehe z.B. die Analysen von Schmidt, 1986, S. 12 f.) und erlauben somit keine differenzierte Erfassung der einzelnen Determinanten. Die separate Erfassung der Bearbeitungsgeschwindigkeit ist dadurch nicht gewährleistet. Ein Verzicht auf die Erhebung einzelner intellektueller Fähigkeiten im Rahmen der Eignungsdiagnostik ist aber nur in bestimmten Fällen sinnvoll, beispielsweise wenn sowohl die Anforderungsanalyse als auch die Bewährungskontrollen übereinstimmend eine Irrelevanz einer bestimmten Fähigkeit für den Ausbildungs-

² Schachinteressierten Leser(inne)n sei mitgeteilt, daß auch Garri Kasparows intellektuelle Leistungen einmal mit einem Test zum Berliner Intelligenzstrukturmodell geprüft wurden (siehe den ausführlichen Bericht im „SPIEGEL“, Nr. 52 aus 1987).

und Berufserfolg indizieren. Sowohl die bisherigen Analysen als auch die Einschätzungen von Expert(inn)en über den Wandel der Berufsanforderungen (wobei z.B. der Kreativität und dem (teilweise davon abhängigen) Problemlösen eine hohe Bedeutung zuerkannt werden) sprechen dafür, zunächst eine möglichst generelle Erfassung aller intellektuellen Fähigkeiten anzustreben. Die Bedeutung der einzelnen Komponenten (z.B. Einfallsreichtum) für die Erfolgsprognose muß dann sukzessive ausgelotet werden, wobei auch die Möglichkeit von Suppressionseffekten³ zu berücksichtigen ist. Ein Ziel der neuen Testentwicklung ist daher eine möglichst repräsentative Erfassung der intellektuellen Fähigkeiten, um somit eine Basis für eine differentielle Diagnostik der Leistungsvoraussetzungen einzelner Berufsbilder zu etablieren.

Ad 2: Differenziertere Auswertung der erhobenen Informationen

Der Bimodalitätsannahme zufolge sind an einer Intelligenzleistung immer mindestens zwei Arten von spezifischen Intelligenzkomponenten maßgeblich beteiligt: eine operative und eine inhaltsgebundene Komponente. Die Lösung der *DGP* Aufgabe „Textanalyse“ erfordert beispielsweise sowohl Verarbeitungskapazität als auch die Fähigkeit zur Verarbeitung *sprachlicher* Informationen. Die Leistung bei diesem Aufgabentyp wird im psychologischen Untersuchungsbefund der *DGP* der Kategorie „*Intellektuelle Leistungen bei sprachlichem Material*“ zugeordnet. Um die *DGP* Aufgabe „Tabellen und Statistiken“ lösen zu können, benötigt man ebenfalls „Verarbeitungskapazität“, inhaltlich ist diese Aufgabe aber einer anderen Modalität, nämlich der Welt der „Zahlen“, zugeordnet. Die Leistung bei der Aufgabe „Tabellen und Statistiken“ wird in der *DGP* einer anderen Kategorie zugeordnet als die Leistung bei der Aufgabe „Textanalyse“, nämlich der Kategorie „*Intellektuelle Leistungen bei numerischem Material*“. Diese unterschiedliche Zuordnung der Leistungen in diesen beiden Aufgaben hat ihr Recht an den verschiedenen Inhalten (Sprache versus Zahlen) der beiden Aufgaben, vergreift sich aber hinsichtlich der gemeinsamen gedanklichen Operation, der Verarbeitungskapazität, die *beiden* Aufgaben zugrundeliegt. Im Zuge der Forschungen zum Berliner Intelligenzstrukturmodell ist eine Analysemethode entwickelt worden, die eine differenziertere Ausschöpfung der erhobenen Informationen erlaubt. Mit Bezug auf die strukturelle Testtheorie nach CATTELL und TSUJIOKA (1964) werden durch die sogenannte Bündelungstechnik die Aufgaben theoriegeleitet bimodal klassifiziert und so aggregiert, daß die erwünschte Information maximiert wird, während die nicht intendierten Variationen durch Ausbalancierung minimiert werden (siehe HUMPHREYS, 1962). Um einen Indikator für die Verarbeitungskapazität zu erhalten, werden beispielsweise entsprechende operationshomogene, aber inhalts heterogene Aufgaben aggregiert (auf die *DGP*- Aufgaben bezogen würden also beispielsweise u.a. die Aufgaben „Textanalyse“ und „Tabellen und Statistiken“ zusammengefaßt, um einen Indikator für die Verarbeitungskapazität zu erhalten, da ja beide Aufgaben Verarbeitungskapazität erfordern, aber unterschiedlichen Inhaltsmodalitäten - Sprache und Zahlen - zugeordnet sind). Durch die theoriegeleitete Bündelung der Aufgaben wird das interessierende Merkmal (im Beispiel die Verarbeitungskapazität) verstärkt, während die nicht interessierenden Merkmale (im Beispiel die Inhalte) durch die diesbezügliche Heterogenität der aggregierten Aufgaben unterdrückt werden. So wird für jede der sieben Fähigkeitskomponenten und für die Allgemeine Intelligenz ein spezifisches

³ Auch ein Subtest, der wenig mit dem Kriterium (z.B. dem Ausbildungs- oder Berufserfolg) zu tun hat, kann zu einer Erhöhung der Vorhersageleistung des Gesamttests beitragen, wenn er hoch mit anderen -prognostisch bedeutsamen- Subtests korreliert ist. Dieser *Suppressionseffekt* kommt zustande, wenn der irrelevant erscheinende Subtest die Vorhersagequalität der relevanten Subtests verbessert. Dies kann dadurch geschehen, daß der prognostisch unbedeutsame Subtest diejenigen Anteile der systematischen Varianz des prognostisch bedeutsamen Subtests unterdrückt, welche nicht mit dem Kriterium korrelieren.

Aufgabenbündel geschnürt. Benötigt man z.B. einen Indikator für die intellektuelle Leistungsfähigkeit beim gedanklichen Umgang mit sprachlichem Material, wird man zweckmäßig möglichst viele operationsheterogene Aufgaben mit homogen sprachlichem Inhalt aggregieren (z.B. Merkfähigkeits-, Einfallsreichtums-, Bearbeitungsgeschwindigkeitsaufgaben und Aufgaben zur Verarbeitungskapazität, denen allen gemeinsam ist, *sprachliches* Aufgabenmaterial zu verwenden). Bei dieser Technik wird die Information einer Aufgabenbearbeitung stets zweifach genutzt: einmal als Indikator für eine operative und einmal als Indikator für eine inhaltliche gedankliche Leistung. Die *DGP* beschränkt sich hingegen zur Zeit pro Aufgabe auf die Verwertung *einer* Information. Ein weiteres Ziel der Testmodifikation ist daher die Anwendung dieser modernen Analysemethoden zum Zwecke einer differenzierteren Auswertung der erhobenen Informationen. Da aggregierte Messungen reliabler sind als Einzelmessungen (z.B. WITTMANN, 1988), wird durch die Bündelungstechnik auch die Meßqualität erhöht.

Unabhängigere Leistungsindikatoren durch die Verwendung von Faktorwerten

In Verbindung mit der Verwendung von Faktorwerten anstelle der bislang bei der *DGP* genutzten Skalenleistungen kann durch das Vorgehen schließlich auch die Voraussetzung für eine differenzierte Interpretation und Prognose der unterschiedlichen intellektuellen Leistungen geschaffen werden, da die Faktorwerte unabhängiger voneinander sind als die Skalenleistungen. Auf der Grundlage der Faktorwerte für operative und inhaltsgebundene Fähigkeiten könnten berufsspezifische Profilanalysen erstellt und erprobt werden. Die Nutzung von Profilverinformationen wurde gerade in der Berufseignungsdiagnostik oft angestrebt, konnte bislang aber nicht überzeugend geleistet werden. Die Profilverinterpretationen der Leistungen im Intelligenz-Struktur-Test (IST) wurden durch die mangelnde Unabhängigkeit und Spezifität der einzelnen Profilverkomponenten beispielsweise längst als ungedeckte Schecks entlarvt (siehe z.B. SCHMIDT-ATZERT, HOMMERS & HESS, 1995).

Modularer Testaufbau / Baukastenprinzip

Die oben beschriebene Bündelungstechnik erlaubt in Kombination mit dem Baukastenprinzip spezifische Skalenbildungen, beispielsweise sieht der Test zum Berliner Intelligenzstrukturmodell (Form 4) auch eine Kurzform zur Erfassung der Allgemeinen Intelligenz und der Verarbeitungskapazität vor. Eine vergleichbare Kurzform eines *DGP*-Tests könnte im Rahmen einer reduzierten Leistungsdiagnostik bei Assessment-Centern für Führungskräfte eingesetzt werden. Eine solche Baukastenform des Tests ist somit ein weiteres Ziel der Testmodifikation in der *DGP*.

Theoretische Fundierung

Die konsequente Orientierung an einem bewährten Intelligenzstrukturmodell und die Entwicklung eines Paralleltests zu einem etablierten Verfahren stellt schon an sich ein erstrebenswertes Ziel der Testentwicklung dar. Der Grad der Bewährung einer Theorie bestimmt den Grad der möglichen Bewährung eines anhand dieser Theorie konstruierten diagnostischen Verfahrens. Die theoretische Ausrichtung und die Parallelität des Meßinstrumentes zu anderen Meßinstrumenten erlauben es außerdem, an den Ergebnissen anderer Forschungsgruppen direkt zu partizipieren und die eigene Arbeit auch an fremden Studien orientieren zu können, da diese von den Methoden und Ergebnissen her übertragbar werden.

Die Realisierung der hier genannten Ziele stellt nicht nur eine Verbesserung des neuen Verfahrens gegenüber dem herkömmlichen *DGP* Verfahren dar, sondern schafft einen

Test, der allgemein die Voraussetzungen mitbringt, konkurrierenden, mit den gleichen - hier für das herkömmliche *DGP* Verfahren aufgelisteten - Mängeln behafteten Tests überlegen zu sein. Außer dem Originaltest zum Berliner Intelligenzstrukturmodell - der aufgrund der mit der Veröffentlichung einhergehenden allgemeinen Zugänglichkeit nur beschränkt für eignungsdiagnostische Zwecke in Frage kommt - wird das neue Verfahren der *DGP* den einzigen Test darstellen, der die im Rahmen der Forschung zum Berliner Intelligenzstrukturmodell entwickelten theoretischen und methodischen Standards einer modernen Intelligenzdiagnostik erfüllt.

Drastische Reduktion der Testzeit und Erweiterung der Verhaltensbeobachtung

Weitere Ziele der Testentwicklung können weitgehend unabhängig von dem Berliner Intelligenzstrukturmodell formuliert werden. Ein zentrales Ziel der Testentwicklung ist es, die für die schriftliche Leistungsprüfung benötigte Zeit drastisch zu reduzieren, ohne daß diese zeitliche Reduktion zu Lasten der Meßqualität geht. Die freigesetzte Zeit soll für eine Erweiterung des Verhaltensteils des Auswahlverfahrens genutzt werden. Denn so wichtig die „akademische“ Intelligenz auch ist: sie ist nur ein Teilsystem der Gesamtpersönlichkeit. Gerade dem - von Intelligenztests nicht erfaßten - kommunikativen Verhalten kommt eine hohe Bedeutung für den beruflichen Erfolg zu. Aus den jahrelangen und umfangreichen Erfahrungen - u.a. im Assessment-Center Bereich - verfügt die *DGP* sowohl über die situativen Aufgaben, die notwendig sind, um bestimmte Verhaltensweisen zu provozieren, als auch über das Know-how der Verhaltensbeobachtung und -beurteilung. Zur Zeit wird bei den Standardauswahlverfahren für den gehobenen Dienst aber nur eine einzige Gruppen- und eine Einzelaufgabe zur Provokation von relevanten Verhaltensweisen angewandt. Grund für diese Bescheidung ist schlicht und einfach die mangelnde Zeit, ein Großteil der Zeit des Auswahlverfahrens wird durch schriftliche Tests beansprucht. Allein für die Applizierung von 13 Denkaufgaben im engeren Sinn werden zur Zeit vier Zeitstunden benötigt, hinzu kommt ein Zeitbedarf von mehreren Stunden für die Erfassung des Arbeitsverhaltens, der Kenntnisse und Fertigkeiten. Demgegenüber benötigt man für die Applizierung aller 45 Aufgaben des Tests zum Berliner Intelligenzstrukturmodell zwei Stunden und 10 Minuten. Der vielfach höhere Zeitbedarf für die *DGP*-Tests ist in der universelleren Anwendbarkeit des Tests begründet. Die *DGP* verwendet zur Zeit ein- und den gleichen Test zur Personalauswahl bei ganz unterschiedlichen Ausbildungs- und Berufsbildern. Die Anforderungsspezifität wird in der Regel erst bei der Aus- und Bewertung der Aufgaben, z.B. durch unterschiedliche Norm- und Anforderungswerte, hergestellt. Die Anweisungen und die Aufgaben der Tests mußten daher so konstruiert werden, daß sie sowohl bei Verwaltungsfachangestellten und gewerblichen Lehrlingen als auch bei Aufstiegsbewerber(inne)n vom gehobenen zum höheren Dienst anwendbar waren. Diese Vorgabe führte dazu, daß die Gestaltung der Anweisungen sich notwendigerweise an der schwächsten Zielgruppe orientierte und die Instruktionen dementsprechend ausführlich ausfielen. Dies erklärt, warum für die Anweisung der identischen Aufgabentypen „Tabellen und Statistiken“ bzw. „Tabellenlesen“ beim *DGP* Test ca. fünf Minuten, beim BIS-Test hingegen zwanzig Sekunden zu veranschlagen sind. Die Vorgabe der möglichst breiten Verwendbarkeit wirkt sich auch auf die Anzahl der für eine gute Messung notwendigen Fragen oder Items aus. Eine Aufgabe, die einige wenige leichte und schwierige Fragen sowie eine größere Anzahl von Fragen mit einer mittleren Schwierigkeit umfaßt, hat die besten Voraussetzungen, für die meisten diagnostischen Zwecke eine befriedigende psychometrische Qualität zu erzielen. Die Schwierigkeit einer Frage / eines Items wird nicht vollständig von der Anforderung selbst determiniert, sondern variiert u.a. in Abhängigkeit von der Leistungsfähigkeit der Personen, die den Test bearbeiten. Die *DGP*-Aufgaben umfassen sehr viele Fragen/Items, um für verschiedene Leistungsgruppen geeignete Fragen/Items bereitzuhalten. Während z.B. die Aufgabe der „sprachlichen Analogien“ im BIS-Test acht Items umfaßt, sind es im

DGP-Test dreiundzwanzig Items, und entsprechend höher ist der Zeitbedarf für die Aufgabenbearbeitung. Zahlreiche Items sind „gut“ für die Bewerber(innen) des mittleren Dienstes, erweisen sich aber für die Bewerber(innen) des gehobenen Dienstes als viel zu leicht. Eine Frage, die von 98 % der Bewerber(innen) des gehobenen Dienstes richtig beantwortet wird, trägt zur Differenzierung zwischen den leistungsstarken und leistungsschwachen Personen dieser Gruppe nichts bei. Eigentlich benötigt man ein solches Item bei einer Testung von Bewerber(inne)n des gehobenen Dienstes also nicht. Das Item mußte bislang aber „mitgeschleppt“ werden, weil es zur Differenzierung von Personen anderer Gruppen benötigt wurde.

Die beschriebene Problematik stellt keine Schwäche des Tests dar, sondern ist die Konsequenz aus den Zielvorgaben der damaligen Testkonstruktion. Ein Auto, das verschiedenen Ansprüchen (Geräumigkeit, hohe Fahrleistungen, Komfort usw.) zugleich genügen soll, wird in spezifischer Hinsicht im Vergleich mit einem „reinen“ Transporter, einem „reinen“ Sportwagen und einem „reinen“ Straßenkreuzer immer schlecht abschneiden. Um das Ziel der drastischen Verkürzung der Testbearbeitungszeit bei gleichbleibender oder erhöhter Meßqualität erreichen zu können, hat die DGP beschlossen, für die Bewerber(innen) des gehobenen Dienstes einen eigenen, *niveauspezifischen* Test zu konstruieren. Durch diese Beschränkung auf eine Zielgruppe wird es möglich, die Anweisungszeiten und die Anzahl der Fragen pro Aufgabe deutlich zu reduzieren. Der Preis dieser spezifischen Testung ist ein deutlich erhöhter organisatorischer Aufwand (z.B. Verdoppelung sämtlicher Testmaterialien, der notwendigen Zeitinvestition für die „Testpflege“, Vereitlung organisatorischer Erleichterungen wie z.B. der gemeinsamen Testung von Bewerber(inne)n des mittleren und gehobenen Dienstes sowie der nachträglichen „Umnormierung“ einzelner Ergebnisse usw.). Dieser erhöhte Aufwand wird sich auf die Kosten der Testung auswirken. Andererseits erlaubt die Verkürzung eine Applizierung aller Tests an einem Tag, so daß die bislang notwendige - und organisatorisch für alle Beteiligten aufwendige - schriftliche Testung an beiden Tagen der Eignungsuntersuchung (Vor- und Haupttest) voraussichtlich entfallen kann. Der zweite Tag des Auswahlverfahrens für Bewerber(innen) des gehobenen Dienstes würde in Zukunft dann allein für die Verhaltensbeobachtung und -bewertung reserviert.

Berücksichtigung (kultur- und geschlechts-) spezifischer Gruppen

Mit der Neukonstruktion wird darüber hinaus das Ziel verfolgt, die mögliche Benachteiligung bestimmter - z.B. geschlechts- oder kulturspezifischer - Gruppen im Test dadurch auszuschließen, daß diese Gruppen schon in der Phase der Testkonstruktion - sozusagen von Anfang an - berücksichtigt werden. Dies schuldet die DGP - ebenso wie andere Institutionen - insbesondere den ostdeutschen Auftraggeber(inne)n und Testand(inn)en. Vor dem Mauerfall basierte die Testkonstruktion in Westdeutschland auf den an westdeutschen Personen erhobenen Daten. Nach der Vereinigung zeigten sich beim Einsatz „westdeutscher Eignungstests“ für ostdeutsche Bewerber(innen) im innerdeutschen Vergleich geringfügig niedrigere Leistungen. Gruppenunterschiede in Eignungstests konstituieren aber selbst bei *identischen* Steigungen der Regressionslinien zwischen Test und Kriterium für beide Populationen ein Problem bei der von Arbeits- und Organisationspsycholog(inn)en unterstützten Entscheidungsfindung. Mit Hilfe einer Modellrechnung (KERSTING, 1995, S. 37 f.) auf Basis der empirisch nachgewiesenen innerdeutschen Unterschiede in dem „westlichen“ Eignungstest der DGP konnte gezeigt werden, daß geeigneten Bewerber(inne)n aus den neuen Bundesländern aufgrund ihrer relativen Testdefizite vergleichsweise häufiger zu Unrecht die Zielposition verwehrt wird (falsch Negative), während ihren westlichen Landsleuten relativ häufiger der „angenehmere“ Fehler einer Überschätzung droht (Zulassung trotz mangelnder Eignung). Bei der Testneuentwicklung wird das Ziel verfolgt, Aufgaben, die in Ost- und Westdeutschland und/oder bei Frauen und Männern anders

selbst kann nicht bei der DGP angewandt werden, da dieses Verfahren 1996 veröffentlicht wird. Die DGP verwendet aus verschiedenen Gründen ausschließlich unveröffentlichte, exklusiv für die Auftraggeber(innen) der DGP zur Verfügung gestellte Testverfahren. Es ist daher notwendig, einen weiteren Test zum Berliner Intelligenzstrukturmodell zu entwickeln, der eine parallele Messung zu dem bereits existierenden Test (Form 4) ermöglicht. Der Berliner Intelligenzstruktur-Test (Form 4) wird der DGP als Vorlage der Testentwicklung zur Verfügung gestellt. Die Autoren, PROF. DR. A.O. JÄGER, DR. H.-M. SÜSS und DIPL.-PSYCH. A. BEAUDUCEL, stehen der Testentwicklung beratend bei, A. BEAUDUCEL übernimmt darüber hinaus aktiv Arbeiten der Testentwicklung und -evaluation.

„funktionieren“, nach Möglichkeit von Anfang an auszusondern.

Voraussetzungen für eine EDV-gestützte Testauswertung schaffen

Die Testentwicklung verfolgt schließlich das Ziel, eine rationelle Datenauswertung auf Basis der elektronischen Datenverarbeitung (EDV) zu unterstützen. Dies erspart nicht nur Zeit und somit Geld, sondern minimiert vor allem die Fehleranfälligkeit der Auswertung und dient der Erhöhung der Auswertungsobjektivität. Um dieses Ziel zu verwirklichen, wurde projektbegleitend eine Arbeitsgruppe zur PC-gestützten Auswertung, Aufbereitung und Ablage der Testdaten gegründet.⁴ Dabei wird zunächst eine Weiterverarbeitung der „von Hand“ ermittelten Testrohwerte angepeilt, die Tests - und respektive die Antwortbogen - sollen aber insofern „zukunftssicher“ konstruiert werden, daß auch die Möglichkeit eines maschinellen „Einlesens“ der Antwortbogen bei der Testkonstruktion berücksichtigt wird. Dazu ist es z.B. notwendig, freie Antwortformate - wie sie zur Zeit etwa bei den Aufgaben „Textrechnen“ oder „Zahlenmatrizen“ verwandt werden - zu modifizieren. Auch in dieser Hinsicht werden mit der neuen Testentwicklung der DGP Tore für die Zukunft geöffnet.

Einen zusammenfassenden Überblick über die hier dargestellten Ziele der Testentwicklung gibt die Abbildung 2.

Abb. 2: Ziele der Testentwicklung

- Repräsentative und differenzierte Erfassung der intellektuellen Fähigkeiten (inkl. des Einfallsreichtums)
- Differenzierte Auswertung der in einer Leistung enthaltenen Informationen durch theoriegeleitete bimodale Klassifikation und Aggregation der Aufgaben
- Verwendung von Faktorwerten, um vergleichsweise unabhängige Leistungsindikatoren zu erhalten
- Anwendung des Baukastenprinzips der Testgestaltung als Voraussetzung für spezifische Skalenbildungen
- Orientierung an einem klaren, anerkannten und bewährten theoretischen Konzept (u. a. als Voraussetzung für die Partizipation an externen Forschungsergebnissen)
- Drastische Verkürzung der Testzeit durch niveauspezifische Testung als Voraussetzung für eine erweiterte Verhaltensbeobachtung und -bewertung
- Testfaires durch Berücksichtigung (z. B. kultur-)spezifischer Gruppen schon in der Phase der Testkonstruktion
- Ermöglichung und Unterstützung einer EDV-gestützten Auswertung, Aufbereitung und Ablage der Testdaten

2.2 Der Entwicklungsprozeß

Die Entwicklung des neuen Testverfahrens (BIS-DGP) nach Maßgabe des Berliner Intelligenzstrukturmodells für die Eignungsdiagnostik von Bewerber(inne)n für den gehobenen Dienst erfolgt in enger Kooperation mit der Berliner Arbeitsgruppe. Der von dieser Forschergruppe entwickelte Originaltest zum Berliner Intelligenzstrukturmodell

⁴ Für das Projekt der technischen Realisierung der EDV-gestützten Auswertung, Aufbereitung und Ablage der Testdaten konnte Herr T. Königer gewonnen werden.

schwindigkeitsaufgaben des neuen BIS-DGP Tests, sondern auch die entsprechenden Originalaufgaben aus dem BIS-4 Test eingesetzt. Ergebnisse liegen zur Zeit noch nicht vor.

Konzeptuelle Probleme bereitet - neben dem schon thematisierten klärungsbedürftigen Verhältnis einzelner Intelligenzkomponenten des Berliner Modells und der von der DGP als „Arbeitsverhalten“ gemessenen Leistungen - vor allem die Erfassung und eignungsdiagnostische Bewertung des Einfallsreichtums. Diesbezüglich stellt sich die Aufgabe, die Messung gegen eine Testvorbereitung zu sichern und empirische Grundlagen für die Bewertung zu sammeln.

selbst kann nicht bei der *DGP* angewandt werden, da dieses Verfahren 1996 veröffentlicht wird. Die *DGP* verwendet aus verschiedenen Gründen ausschließlich unveröffentlichte, exklusiv für die Auftraggeber(innen) der *DGP* zur Verfügung gestellte Testverfahren. Es ist daher notwendig, einen weiteren Test zum Berliner Intelligenzstrukturmodell zu entwickeln, der eine parallele Messung zu dem bereits existierenden Test (Form 4) ermöglicht. Der Berliner Intelligenzstruktur-Test (Form 4) wird der *DGP* als Vorlage der Testentwicklung zur Verfügung gestellt. Die Autoren, Prof. Dr. A.O. JÄGER, Dr. H.-M. SÜSS und Dipl.-Psych. A. BEAUDUCEL, stehen der Testentwicklung beratend bei, A. BEAUDUCEL übernimmt darüber hinaus aktiv Arbeiten der Testentwicklung und -evaluation.

Der neue Test wird voraussichtlich 40 Aufgaben umfassen, wovon sieben bis acht Aufgaben aus der bisherigen Testbatterie der *DGP* in verkürzter Form übernommen werden. Hierzu zählen besonders bewährte Aufgaben, wie z.B. die „Textanalyse“, die Aufgabe „Tabellen und Statistiken“ oder die „Analogieaufgabe“, die sich voraussichtlich problemlos in das Berliner Intelligenzstrukturmodell klassifizieren lassen. Die übrigen Aufgaben wurden analog zu den Aufgaben des BIS-4-Tests neuentwickelt. Alle modifizierten oder neuentwickelten Aufgaben liefen zunächst vereinzelt zur Erprobung bei verschiedenen Eignungsuntersuchungen mit und wurden dann aufgrund der gewonnenen Daten modifiziert, erneut erprobt usw.. Im September 1995 konnte dann die erste Untersuchung mit einer Vorform der neuen Testbatterie durchgeführt werden. Der BIS-*DGP* Test in der Version 0.1 wurde bei 125 Bewerber(inne)n im Anschluß an den „herkömmlichen“ Vortest eingesetzt. Auf Grundlage dieser Daten werden die Aufgaben weiter optimiert. Die Version 0.2 des BIS-*DGP*-Tests wird dann von Prof. H. Brandstätter in Linz zur weiteren Erprobung eingesetzt. Zusätzlich zu dem BIS-*DGP* Test wird dabei auch der BIS-4 Test eingesetzt, um so erste Parallelitätsprüfungen vornehmen zu können. Die im nächsten Schritt vorgesehene Applizierung beider Tests (BIS-*DGP* und BIS-4) bei Student(inn)en einer Verwaltungsfachhochschule wird erste Prüfungen der Kriteriumsvalidität erlauben. Das neue Testverfahren soll nach Möglichkeit mit Beginn der Saison 1996/1997 einsetzbar sein.

Der Einsatz des neuen Verfahrens wird voraussichtlich ca. zwei Stunden und 45 Minuten reine Testzeit beanspruchen. Zusätzlich zu dieser differenzierten Intelligenzdiagnostik wird die *DGP* auch in Zukunft einige berufselementare Fertigkeiten und allgemeine Kenntnisse der Bewerber(innen) sowie das Arbeitsverhalten testen. Bezüglich des Arbeitsverhaltens muß die Beziehung zu den Modellkomponenten Verarbeitungskapazität und Bearbeitungsgeschwindigkeit noch geklärt werden. Für diese weiteren Tests und für die benötigten Pausen ist ein zusätzliches Zeitbudget zu veranschlagen. Es wird angestrebt, die Gesamttestdauer auf vier Stunden zu beschränken, abschließende Aussagen hierzu können zur Zeit aber noch nicht getroffen werden.

Technische Hürden bei der Erstellung von BIS-4 analogen Testaufgaben für den BIS-*DGP*-Test stellen sich insbesondere hinsichtlich der jeweils gewählten Antwortformate. Im Originaltest zum Berliner Intelligenzstrukturmodell werden die Lösungen unmittelbar im Testheft - und nicht wie bei der *DGP* auf einem separaten Antwortbogen - fixiert. Die Übernahme dieser Vorgabe würde dem oben genannten Ziel einer testökonomischen und möglicherweise EDV-gestützten Testauswertung zuwiderlaufen. Es wurde daher versucht, analoge Varianten der Aufgaben mit Antwortbogen zu konstruieren. Insbesondere für die Aufgaben zur Bearbeitungsgeschwindigkeit muß nun geprüft werden, ob der Anforderungscharakter der Aufgaben auch bei einer Darbietung der Aufgaben mit Antwortbogen (die den Testand(inn)en eine - u.a. zeitlich belastende - Übertragungsleistung abverlangt) unverändert bleibt und ob Antwortbogeneffekte (z.B. aufgrund linkshändiger Personen) vermieden werden können. Zur Klärung dieser Fragen wurden bei der Untersuchung im September nicht nur die neun Bearbeitungsge-

schwindigkeitsaufgaben des neuen BIS-DGP Tests, sondern auch die entsprechenden Originalaufgaben aus dem BIS-4 Test eingesetzt. Ergebnisse liegen zur Zeit noch nicht vor.

Konzeptuelle Probleme bereitet - neben dem schon thematisierten klärungsbedürftigen Verhältnis einzelner Intelligenzkomponenten des Berliner Modells und der von der DGP als „Arbeitsverhalten“ gemessenen Leistungen - vor allem die Erfassung und eignungsdiagnostische Bewertung des Einfallsreichtums. Diesbezüglich stellt sich die Aufgabe, die Messung gegen eine Testvorbereitung zu sichern und empirische Grundlagen für die Bewertung zu sammeln.

3. Zur Akzeptanz der Anweisungen und der Aufgaben des neuen Tests

Bei der Konstruktion des neuen Tests wird nicht nur eine hohe psychometrische Qualität, sondern auch eine hohe soziale Qualität des neuen Verfahrens angestrebt. Bei der DGP-internen Diskussion des neuen Testkonzepts wurden Befürchtungen laut, daß die geplante Darbietung einer größeren Anzahl von Aufgaben bei einer gleichzeitigen Reduktion der Testzeit zu einer erhöhten Belastung der Bewerber(innen) und einer geringeren Akzeptanz des Verfahrens führen könnte. Dieser Befürchtung wurde entgegengehalten, daß auch zu lange Testbearbeitungszeiten frustrierend und angstausslösend wirken können und daß der beständige Wechsel der Aufgaben (und somit der Anforderungen) leistungsstimulierend wirken kann. So wurden für jedes der beiden Testkonzepte (herkömmlicher DGP-Test versus BIS-DGP Test) Akzeptanzvorteile postuliert. Letztendlich handelte die Diskussion aber von empirisch prüfbaren Punkten. Das letzte Wort in empirischen Fragen sollte der Empirie überlassen werden. Mit einer Untersuchung wurde versucht, es ihr zu entlocken. Zu diesem Zweck wurden bei der Erhebung im September - bei der ja zuerst der herkömmliche DGP-Vortest und im Anschluß der neue BIS-DGP-Test eingesetzt worden waren - die Teilnehmer(innen) zu einigen Akzeptanzgesichtspunkten der beiden Testverfahren befragt. Um eine möglichst unverzerrte Einschätzung der Personen zu erhalten, wurde die Erhebung anonymisiert durchgeführt. Der Zusammenhang der Akzeptanz zu Leistungsmerkmalen der einschätzenden Person kann deshalb nicht bestimmt werden.

Neben der allgemeinen Akzeptanzfrage sollte mit der Erhebung auch geklärt werden, ob die verkürzten Anweisungen und Beispiele des neuen Tests von den Testand(inn)en als ausreichend angesehen werden.

Von den 125 Teilnehmer(inne)n der Untersuchung füllten 124 den kurzen Fragebogen zur Akzeptanz verwertbar aus. Der Frauenanteil betrug 59 %, die Personen waren im Durchschnitt 21 Jahre alt (Median: 19). Die nachstehenden Ergebnisse müssen aufgrund verschiedener Besonderheiten der Untersuchung als vorläufig qualifiziert werden. Zunächst handelte es sich bei dem eingesetzten neuen Verfahren lediglich um eine Alpha-Version, die hinsichtlich der Akzeptanz mit dem ausgereiften DGP-Verfahren konkurrieren mußte. Auf keinen Fall können die Ergebnisse ohne weiteres als Akzeptanzwerte des - nicht eingesetzten - BIS-4-Tests interpretiert werden. Zum anderen wurde der neue Test appliziert, nachdem die Teilnehmer(innen) schon fast vier Stunden Leistungsprüfung hinter sich hatten, ein Umstand, der sich vermutlich negativ auf die Akzeptanz des später applizierten BIS-DGP-Verfahrens ausgewirkt haben dürfte.

Zunächst wurde nach der Einstellung gegenüber der unterschiedlichen Testbearbeitungszeit gefragt: 52 % der Personen bevorzugten der Befragung zufolge die im bisherigen DGP-Test gegebene Möglichkeit, sich ausführlich und lange mit vielen Fragen eines Aufgabentyps zu beschäftigen, während 31 % die Vielfalt und Abwechslung der

schnellen Abfolge unterschiedlicher Aufgabentypen in der vorläufigen Version des *BIS-DGP-Tests* präferierten. 17 % sahen diesbezüglich keinen Unterschied zwischen den beiden Verfahren.

Bezüglich der Qualität der Instruktionen und Beispiele gaben die Bewerber(innen) mit 48 % den kurzen und knappen Einführungen des neuen *BIS-DGP-Tests* gegenüber den (von 33 % bevorzugten) ausführlichen und langen Instruktionen des herkömmlichen *DGP-Tests* den Vorzug. (19 % wählten die Antwortalternative „*Ich sehe da keinen Unterschied*“.)

Ein Problem bei den beiden dargestellten Ergebnissen ist, daß das „forced choice“ Format der Fragen eine Quantifizierung der Zustimmung oder Ablehnung vereitelt. Die Werte teilen z.B. nichts darüber mit, ob diejenigen 52 %, die den herkömmlichen *DGP-Test* hinsichtlich der für einen Aufgabentyp zur Verfügung stehenden Bearbeitungszeit dem neuen Test vorziehen, dies mit großer Entschiedenheit oder mit hauchdünner Präferenz tun. Ebenso sagt die Wahlmehrheit für die Instruktionen des neuen Tests natürlich nur aus, daß diese Instruktionen von knapp der Hälfte der Bewerber(innen) als „besser“ erachtet werden als die herkömmlichen Instruktionen. Die interessante Frage nach der *Qualität* der Instruktionen bleibt aber unbeantwortet. Der restliche Teil des Akzeptanzfragebogens sah deshalb jeweils fünf - teils positive, teils negative - qualitative Aussagen über die beiden Testverfahren vor. Die Teilnehmer(innen) wurden gebeten, ihre Zustimmung zu den einzelnen Items auf einer sechsstufigen (Likert-) Skala auszudrücken. Die einzelnen Aussagen und die Befragungsergebnisse sind in den Tabellen 1 und 2 dargestellt. Tabelle 1 zeigt die Reaktion der Teilnehmer(innen) auf die *positiven* Aussagen. Ein hoher Wert bedeutet eine hohe Zustimmung. Das Antwortformat ging von „1“ („*Die Aussage trifft nicht zu*“) über die Zwischenstufen „*trifft überwiegend nicht zu*“, „*trifft kaum zu*“, „*trifft etwas zu*“ und „*trifft überwiegend zu*“ bis „6“ („*Die Aussage trifft genau zu*“).

Inwieweit die Bewerber(innen) den *negativen* Aussagen über die Verfahren zustimmten, ist in der Tabelle 2 verzeichnet. Eine hohe Zustimmung zu einer negativen Aussage bedeutet inhaltlich eine *Ablehnung* des Testverfahrens hinsichtlich des thematisierten Aspektes.

Die Hauptergebnisse lauten: Den verkürzten Anweisungen wird von den Testand(inn)en eine hohe Qualität attestiert, die Instruktionen gewährleisten das Aufgabenverständnis. Beiden Testkonzepten (herkömmlicher *DGP-Test*, neuer *BIS-DGP-Test*) werden Stärken und Schwächen zugeschrieben. Die sich bei der ersten „forced-choice“-Frage abzeichnende Bevorzugung der längeren Bearbeitungszeiten kann auf der Ebene differenzierter Aussagen nicht bestätigt werden. Aggregiert man jeweils die fünf Aussagen zu einem Verfahren (nach vorheriger Umpolung der Antworten zu den negativen Aussagen), so ergibt sich eine geringfügig bessere Wertung für das neue Testverfahren (4.2 gegenüber 3.9), die sich aber nur auf dem 10%-Niveau als signifikant erweist. Ein solcher unmittelbarer Vergleich ist aber problematisch, da die jeweiligen Aussagen nicht unbedingt gleichwertig sind. Wichtiger ist, daß beide Verfahren offensichtlich einerseits akzeptabel sind bzw. daß andererseits kein Verfahren es allen Personen vollständig recht machen kann. Für das Akzeptanzurteil sind - dem gegenwärtigen Stand der Forschung nach - ohnehin nicht nur Verfahrensmerkmale, sondern auch Personenmerkmale bedeutsam. Die insgesamt positive Bewertung der Verfahren in einer *anonymen* Befragung im unmittelbaren Anschluß an einen mehrstündigen anstrengenden Test, ist beachtenswert. Dies deutet einmal mehr darauf hin, daß die von pseudo-professionellen Testkritiker(inne)n betriebene Schwarzmalerei der „unzumutbaren“ Testverfahren letztendlich Höhenflüge in der Thermik selbstgemachter Winde darstellen, die nur solange funktionieren, wie man selektiv düstere Einzelfälle aus „1001 Bewerbungsverfahren“ herausgreift (HESSE & SCHRADER, 1988).

Tabelle 1:

Zustimmung zu positiven Aussagen über den neuen BIS-DGP-Test			
Aussage	Mittelwert	Median	Std.Dev.¹
Mir hat es besonders gefallen, daß bei dem Test ganz unterschiedliche Aufgabentypen einander rasch abwechselten.	4.1	4	1.5
Durch die vielen unterschiedlichen Aufgaben in dem Test ist gewährleistet, daß für jeden etwas dabei ist.	4.8	5	1.0
Die Erklärungen der Aufgaben bei dem Test waren völlig ausreichend und präzise. Ich wußte genau, was bei den Aufgaben zu tun war.	5.2	5	1.0
Zustimmung zu positiven Aussagen über den herkömmlichen DGP-Test			
Aussage	Mittelwert	Median	Std.Dev.¹
Bei dem Test hat mir besonders gefallen, daß es mehr Zeit und mehrere Fragen für einen Aufgabentyp gab.	4.3	5	1.3
Bei dem Test war durch die größere Anzahl an Fragen zu einem Aufgabentyp und die längere Bearbeitungszeit pro Aufgabentyp gewährleistet, daß jede/r sich richtig auf den jeweiligen Aufgabentyp einstellen kann.	4.6	5	1.1
Ich finde es besser, wenn man sich auf einige wenige Aufgabentypen beschränkt - wie bei diesem Test - und diese ausführlich darbietet, anstatt immer von einem Aufgabentyp zum nächsten zu springen.	3.6	4	1.5

N=124; ¹ Std.Dev.: Standardabweichung**Tabelle 2:**

Zustimmung zu negativen Aussagen über den neuen BIS-DGP-Test			
Aussage	Mittelwert	Median	Std.Dev.¹
Bei dem Test war die Bearbeitungszeit so knapp bemessen, daß ich meine wirkliche Leistung gar nicht zeigen konnte.	3.8	4	1.3
Der rasche Wechsel von einer Aufgabe zur nächsten in dem Test war besonders belastend.	3.5	4	1.5
Zustimmung zu negativen Aussagen über den herkömmlichen DGP-Test			
Aussage	Mittelwert	Median	Std.Dev.¹
Die Erläuterungen und Beispiele zu den Tests waren unnötig lang. Schon bevor die Anweisung zu Ende war, wußte ich, worauf es bei der Aufgabe ankam.	3.9	4	1.4
Daß man sich so lange mit so vielen Fragen eines Aufgabentyps beschäftigen mußte, wie bei diesem Test, war besonders belastend.	2.8	3	1.4

N=124; ¹ Std.Dev.: Standardabweichung

4. Zusammenfassung

Zur Zeit wird in der *DGP* ein neues Testverfahren für die Eignungsdiagnostik von Bewerber(inne)n des gehobenen Dienstes entwickelt. Die Entwicklung des Tests orientiert sich an dem von Jäger vorgestellten Berliner Intelligenzstrukturmodell (BIS) und dem modellkonformen BIS-4 Test. Das Modell ist national und international anerkannt, seine Geltungsbereiche wurden in zahlreichen unabhängigen Untersuchungen ausgelotet. Mit dem neuen Test soll eine größere Vielfalt von intellektuellen Fähigkeiten erfaßt werden als mit dem herkömmlichen Testverfahren, wobei auch Kreativitätskomponenten in die zukünftige Diagnostik einbezogen werden können. Durch eine theoriegeleitete (Bimodalitätsannahme) Klassifikation und Aggregation der Aufgaben werden die Leistungsinformationen in Zukunft differenzierter ausgewertet als bisher. Die neue Testbatterie wird speziell für den gehobenen Dienst konzipiert. Diese niveauspezifische Testung ermöglicht eine drastische Verkürzung der für die Testung benötigten Zeit. Die Durchführung des neuen Tests wird voraussichtlich 165 Minuten benötigen. Es wird angestrebt, das schriftliche Verfahren insgesamt (inkl. Überprüfung der Kenntnisse und des Arbeitsverhaltens) auf ca. vier Stunden zu verkürzen. Die freigesetzte Zeit soll für eine Erweiterung der Verhaltensbeobachtung und -beurteilung genutzt werden. Spezifische Personenkreise - wie z.B. ostdeutsche Bewerber(innen) - werden von Anfang an bei der Testkonstruktion berücksichtigt, um ein vergleichbares „Funktionieren“ der Tests bei verschiedenen Gruppen zu gewährleisten. Mit dem neuen Test werden außerdem die Voraussetzungen für eine EDV-gestützte Testauswertung geschaffen. Das neue Verfahren soll nach Möglichkeit mit dem Beginn der Saison 1996/1997 einsetzbar sein, der Entwicklungsprozeß wird in dem Artikel skizziert. Ergebnissen einer vorläufigen Akzeptanzeinschätzung zufolge sind die Aufgaben des neuen Testverfahrens trotz der drastisch verkürzten Instruktionen gut verständlich. Das neue Testkonzept wird von den Bewerber(innen) nicht schlechter beurteilt als das herkömmliche.

5. Literatur

- Carroll, J.B. (1993). **Human cognitive abilities. A survey of factor analytic studies**
New York, NY: Cambridge University Press.
- Cattell, R.B. & Tsujioka, B. (1964) **The importance of factor-trueness and validity, versus homogeneity and orthogonality, in test scales**
Educational and Psychological Measurement, 24, 3-30
- Doll, J. & Mayr, U. (1987) **Intelligenz und Schachleistung. Eine Untersuchung an Schachexperten**
Psychologische Beiträge, 29, 270-289
- Gustafsson, J.-E. (1988) **Hierarchical models of individual differences in cognitive abilities**
In R.J. Sternberg (Hrsgb.), Advances in the psychology of human intelligence, (Vol. 4, S. 35-71)
Hillsdale, NJ: Erlbaum.
- Hesse, J. & Schrader, H.L. (1988) **Erlebnisse aus 1001 Bewerbungen**
Frankfurt/M.: Eichborn.
- Hörmann, H.J. & Thomas, M. (1989) **Zum Zusammenhang zwischen Intelligenz und komplexem Problemlösen**
Sprache & Kognition, 8, 23-31
- Humphreys, L.G. (1962) **The organisation of human abilities**
American Psychologist, 17, 475-483
- Jäger, A.O. (1967) **Dimensionen der Intelligenz**
Göttingen: Hogrefe
- 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. & Althoff, K. (1993) **Der WILDE-Intelligenz-Test** (Zweite, revidierte Auflage
Göttingen: Hogrefe.
- Jäger, A.O. & Tesch-Römer, C. (1988) **Replikation des Berliner Intelligenzstrukturmodells (BIS) in den „Kit of Reference Test for Cognitive Factors“ nach French, Ekstrom & Price (1963)**
Zeitschrift für Differentielle und Diagnostische Psychologie, 9, 77-96
- Jäger, A.O., Süß, H.-M. & Beauducel, A. (in Druck) **Berliner Intelligenzstruktur-Test, Form 4, Handanweisung**
Göttingen: Hogrefe.
- Kersting, M. (1995) **Der Einsatz „westdeutscher“ Tests zur Personalauswahl in den neuen Bundesländern und die Fairneßfrage: Auswirkungen der Testleistungsdisparität zwischen Ost und West auf die Auswahlentscheidung**
Report Psychologie, 20, 32-41

- Kleine, D. & Jäger, A.O. (1987) **Replikation des Berliner Intelligenzstrukturmodells (BIS) bei brasilianischen Schülern und Studenten**
Diagnostica, 33, 14-29
- Kleine, D. & Jäger, A.O. (1989) **Kriteriumsvalidität eines neuartigen Tests zum Berliner Intelligenzstrukturmodell. Eine Untersuchung an brasilianischen Schülern und Studenten**
Diagnostica, 35, 17-37
- Pfister, H.R., & Jäger, A.O. (1992) **Topographische Analysen zum Berliner Intelligenzstrukturmodell BIS**
Diagnostica, 38, 91-115
- Rosas, R. (1990) **Replikation des Berliner Intelligenzstrukturmodells (BIS) und Vorhersagbarkeit des Studien-erfolgs bei chilenischen Studenten**
(Unveröffentlichte Dissertation im FB Erziehungs- und Unterrichtswissenschaften der Freien Universität Berlin)
Berlin: FU Berlin
- Schmidt-Atzert, L., Hommers, W. & Heß, M. (1995) **Der I-S-T 70: Eine Analyse und Neubewertung**
Diagnostica, 41, 108-130
- Schmidt, J.U. (1983) **Berliner Lehrlingsuntersuchung 1982**
(Arbeitsbericht 3 des Forschungsprojektschwerpunktes „Produktives Denken / Intelligentes Verhalten“)
Berlin: Freie Universität Berlin.
- Schmidt, J.U. (1984) **Simultane Überprüfung der Zweimodalität im Berliner Intelligenzstrukturmodell**
Diagnostica, 30, 93-103
- Schmidt, J.U. (1986) **Analysen zum Berliner Intelligenzstrukturmodell und der Eignungstestbatterie der DGP. (DGP-Informationen)**
Hannover: Deutsche Gesellschaft für Personalwesen e.V., 46, 2-24
- Schmidt, J.U. (1987) **Zur Vorhersagbarkeit des Ausbildungserfolges im gehobenen Verwaltungsdienst durch einen auf der Grundlage des Berliner Intelligenzstrukturmodells entwickelten Test. (DGP-Informationen)**
Hannover: Deutsche Gesellschaft für Personalwesen e.V., 48, 29-45
- Schmidt, J.U. (1993) **Thurstones Primary Mental Abilities und das Berliner Intelligenzstrukturmodell - Eine empirische Gegenüberstellung**
Zeitschrift für Differentielle und Diagnostische Psychologie, 14, 87-100
- Schmidt, J.U., Brocke, B., Jäger, A.O., Doll, H. & König, F. (1986) **Entwicklung eines Tests für das Berliner Intelligenzstrukturmodell**
(Arbeitsbericht 4 des Forschungsprojektschwerpunktes „Produktives Denken / Intelligentes Verhalten“)
Berlin: Freie Universität Berlin.

- Sperber, W., Wörpel, S.,
Jäger, A.O. & Pfisters, R. (1985) **Praktische Intelligenz. Untersuchungsbericht und erste Ergebnisse**
(Arbeitsbericht 5 des Forschungsprojektschwerpunktes „Produktives Denken / Intelligentes Verhalten“)
Berlin: Freie Universität Berlin
- Süß, H.M., Oberauer, K.
& Jäger, A.O. (1988) **Geltungsbereiche des Berliner Intelligenzstrukturmodells (BIS)**
In W. Schönplugh (Hrsg.), Bericht über den 36. Kongreß der Deutschen Gesellschaft für Psychologie in Berlin 1988 (Bd. Band 1, S. 300 f.)
Göttingen: Hogrefe
- Süß, H.M., Kersting, M. &
Oberauer, K. (1991) **Intelligenz und Wissen als Prädiktoren für Leistungen bei computersimulierten komplexen Problemen**
Diagnostica, 37, 334-352
- Süß, H.-M., Oberauer, K.
& Kersting, M. (1993) **Intellektuelle Fähigkeiten und die Steuerung komplexer Systeme**
Sprache & Kognition, 12, 83-97
- Wittmann, W. (1982) **Jägers bimodale Klassifikation von Intelligenzleistungen im Spiegel der Kovarianzzerlegung**
In: A.O. Jäger (Hrsgb.) Berliner Beiträge zur Intelligenzforschung. (Arbeitsbericht 1 des Forschungsprojektschwerpunktes „Produktives Denken / Intelligentes Verhalten“)
Berlin: Freie Universität Berlin, S. 35-41
- Wittmann, W.W. (1988) **Multivariate reliability theory: Principles of symmetry and successful validation strategies**
In J.R. Nesselroade & R.B. Cattell (Hrsg.), Handbook of multivariate experimental psychology (2nd ed., S. 505-560)
New York: Plenum
- Wittmann, W. & Matt, G.E.
(1988) **Aggregation und Symmetrie. Grundlagen einer multivariaten Reliabilitäts- und Validitätstheorie, dargestellt am Beispiel der differentiellen Validität des Berliner Intelligenzstrukturmodells**
Diagnostica, 32, 309-329.