

bildungshilfe für die Dritte Welt vieles, was an deutschen Erfahrungen vorhanden ist, an die Entwicklungsländer weitergegeben werden kann. Die BBF-Forschungsergebnisse liegen im übrigen im reichen Maße auf dem Tisch. Ich habe inzwischen heute fast 300 Mitarbeiter, es liegen inzwischen 26 Bände unsere Schriftenreihe vor und 11 Folgen unserer Zeitschrift. Es gibt unzählige Gutachten, die laufend auf den verschiedensten Gebieten erstellt werden, wir stehen in Konsultationen und in permanenter Politikberatung. Und wir haben heute auch schon einige unserer wissenschaftlichen Mitarbeiter in das Bildungsministerium abgeordnet, um dort zu helfen, Probleme aufzuarbeiten. Wenn Sie mich nach den Bereichen fragen — das war ja der zweite Teil Ihrer Fragestellung —, dann muß ich das Feld der Berufsbildung im weitesten Sinne aufführen. Dazu gehören zum Beispiel Grundlagen, Informationssystem, Statistik des beruflichen Bildungswesens. Dazu gehören Finanzierungsfragen, insbesondere also Ordnungsvorstellungen und Strukturvorstellungen. Dazu gehört die vorhin schon behandelte Curriculumforschung ebenso wie die Ausbildungsordnungsforschung, aber wir beschäftigen uns auch mit Fragen der Erwachsenenbildung, der Qualifizierung Erwachsener, der Umschulung, der Rehabilitation, der Reaktivierung und der Andragogik. Wir beschäftigen uns darüber hinaus mit der Ordnung des Fernunterrichtswesens, ein für die Zukunft sehr wesentliches Gebiet, schließlich auf dem Gebiet der Medienforschung mit Fragen der Bildungstechnologie, der Ausbildungsmittel, der

offenen Lehrsysteme und der Entwicklung von Übungsreihen für die praxisorientierte Arbeit an der Front.

Huffzky: Wenn das Berufsbildungsgesetz nun noch in dieser Legislaturperiode von Bildungsminister Rohde novelliert werden soll, welchen Anteil hat dann das BBF an diesem Gesetz? Wo liegen die konkreten Vorarbeiten, die Hilfestellungen?

Rosenthal: Wir haben zunächst einmal das bisher noch geltende Gesetz im Hinblick auf seine Auswirkungen untersucht und dabei feststellen können, daß das derzeit geltende BBiG so schlecht gar nicht ist, denn die Ergebnisse, die wir jetzt vorlegen, zeigen, daß seit 1969 auf dem Gebiet der Reformierung des beruflichen Bildungswesens eine Menge geschehen konnte, dank dieses Gesetzes. Trotzdem stehen noch viele Fragen an, die geregelt und die gesetzgeberisch bearbeitet werden müssen. Die Gesetzgebung ist natürlich eine Frage der Legislative und insofern beschränken wir uns auf die Politikberatung, das heißt auf die Beratung der Bildungspolitiker, soweit sie an Fragen der beruflichen Bildung besonders interessiert sind. Wir wissen und haben uns immer so eingeschätzt, daß es auf diesem Gebiet nicht nur reine Forschung gibt, sondern, daß man hier auch Forschung als Prinzip verfolgen muß, daß man der Anwendungsorientierung in jedem Fall im Sinne von Entwicklung und Erprobung Vorkauf geben muß. Von daher tragen wir indirekt zu den Überlegungen bei, die sich mit Novellierungsfragen des Berufsbildungsgesetzes beschäftigen.

Wolfgang Mönikes

Zur Möglichkeit der Berufsberatung über Berufsgruppenprofile

I.

Zur Praxis

Es gehört zur üblichen Praxis der Berufsberatung und der Eignungsauslese in der Bundesrepublik Deutschland und in anderen Ländern, daß individuelle Profile mit Gruppenprofilen verglichen werden. Dabei handelt es sich meist um Profile aus Verfahren, die für sich in Anspruch nehmen, Teilbereiche der „Intelligenz“ oder sonstige berufsbezogene Leistungen zu erfassen. Man geht dabei von der Annahme aus, daß durch einen solchen Vergleich beurteilt werden kann, inwieweit ein Proband mit seiner Leistungsstruktur der erforderlichen Struktur eines bestimmten Berufstypus entspricht. Zu diesem Zweck wird üblicherweise das individuelle Leistungsprofil ermittelt und mit vorliegenden Berufsprofilen verglichen. Das Maß der Übereinstimmung zwischen diesen beiden Profilen wird als Hinweis für die Eignung bzw. Nichteignung für einen bestimmten Beruf bzw. für eine bestimmte berufliche Position verstanden. Oftmals wird dabei rein intuitiv vorgegangen, das heißt, der Berater orientiert sich an den optisch erkennbaren Unterschieden. Er vernachlässigt kleine Unterschiede, interpretiert mittlere Unterschiede mit Vorbehalt und große Unterschiede mit Nachdruck.

Ein solches Vorgehen wird durch die Handanweisungen zu verschiedenen im deutschen Sprachbereich vorliegenden Verfahren angeregt; beispielsweise beim Berufseignungstest von SCHMALE und SCHMIDTKE (BET), beim Intelligenz-Struktur-Test von AMTHAUER (I-S-T 70) und beim Leistungsprüfsystem von HORN (LPS).

Soweit psychologische Dienste in oder im Auftrag von Unternehmen eignungsdiagnostische Untersuchungen durchführen, wird dort mit diesen oder anderen Verfahren oft entsprechend vorgegangen.

Im folgenden soll unter unmittelbarer Bezugnahme auf den Intelligenz-Struktur-Test von AMTHAUER bei Berücksichtigung der übrigen genannten Verfahren dargestellt werden, welche theoretischen und methodischen Bedenken bezüglich dieses Vorgehens zu erheben sind und welche Folgerungen für die Berufsberatung daraus gezogen werden sollten.

II.

Theoretische Überlegungen

Wenn man die statistischen Daten der auf der Grundlage der klassischen Testtheorie konstruierten eignungsdiagnostischen Verfahren anschaut — was man vor dem Einsatz jeden Verfahrens tun sollte! —, fällt auf, daß die auf die Möglichkeit zur Berufsberatung sich beziehenden Daten meist sehr unbefriedigend ausfallen. Das gilt insbesondere dann, wenn die Gültigkeit des Verfahrens über ein äußeres Kriterium gemessen wurde. Hinsichtlich der hier behandelten Fragestellung handelt es sich dabei meist um den erfolgreichen Abschluß einer Lehre bzw. die bei den im jeweiligen Beruf Tätigen ermittelten Meßwerte. In diesem Zusammenhang ist zu bedenken, daß dabei oft allzu leichtfertig die Gültigkeit des äußeren Kriteriums unterstellt wird. Tatsächlich aber wird in zunehmendem Maße deutlich, daß im beruflichen Bereich bislang weder eine verlässliche Definition der jeweils

entscheidenden Anforderungsmerkmale erfolgt ist noch unter Berücksichtigung des technischen Wandels in absehbarer Zeit erfolgen kann. Wenn aber Anforderungsmerkmale beruflicher Tätigkeit nicht eindeutig festlegbar sind, können entsprechende Eignungsmerkmale nicht oder doch nur sehr unzulänglich untersucht werden. Es muß beim derzeitigen Stand der Forschung eindeutig festgestellt werden, daß es der Arbeitswissenschaft und mit ihr kooperierenden Disziplinen bislang nicht in ausreichendem Maße gelungen ist, berufsspezifische Funktionen so eindeutig festzulegen, daß sie als äußere Kriterien bei der Konstruktion, statistischen Analyse und Verwendung eignungsdiagnostischer Verfahren zugrunde gelegt werden können. Solange dies nicht der Fall ist, muß die Anwendung sogenannter Berufsgruppenprofile allein aus diesem Grunde fragwürdig bleiben und kann allenfalls mit erheblichen Vorbehalten erfolgen.

Hinzu kommt, daß in zunehmendem Maße erkannt worden ist, daß für eine erfolgreiche Berufstätigkeit nicht allein das Vorhandensein bestimmter Fähigkeiten und Fertigkeiten notwendige Voraussetzung ist, sondern daß daneben der Eigenschafts- und Motivationsstruktur des einzelnen entscheidende Bedeutung zukommt. Ohne daß es an dieser Stelle möglich ist, auf Besonderheiten dieses Teilproblems einzugehen, muß unter theoretischen Rücksichten unterstrichen werden, daß nach allen zur Verfügung stehenden Erfahrungswerten berufliche Tätigkeit ebenso entscheidend bestimmt ist durch bestimmte Eignungsmerkmale im Sinne tätigkeitsspezifischer Notwendigkeiten wie durch die Eigenschaften und Einstellungen eines Menschen, der einen bestimmten Beruf ausübt. Es scheint in jedem Falle ein zu einfaches Denkmodell zu sein, wenn in der Berufseignungsdiagnostik von der Annahme ausgegangen wird, daß das Vorhandensein bestimmter Fertigkeiten für das Ausüben bestimmter Funktionen **allein** über den beruflichen Erfolg des einzelnen entscheidet. Ob die Berücksichtigung weiterer Persönlichkeitsdimensionen wie beispielsweise die der erwähnten Eigenschafts- und Motivationsstruktur ausreichen, muß zum gegenwärtigen Zeitpunkt dahingestellt bleiben. Immerhin kann beim derzeitigen Stand der Forschung davon ausgegangen werden, daß neben spezifischen Begabungen „die Befriedigung im Beruf größtenteils vom Temperament (im Sinne von Grundeigenschaften, der Verf.) und von der Motivation einer Person abhängt“ (CATTELL, 1973, S. 302). Es deutet sich damit in diesem Zusammenhang die Notwendigkeit der Erstellung ganz anderer als der bisher verwendeten Berufsgruppenprofile an.

In diesem Zusammenhang ist letztlich darauf hinzuweisen, daß von guten eignungsdiagnostischen Verfahren eine formulierbare theoretische Fundierung erwartet werden muß. Diese Forderung weist auf einen entscheidenden Mangel der genannten und anderer üblichen Verfahren hin. Man tut ihnen kaum Unrecht, wenn man feststellt, daß sie eine theoretische Begründung vermissen lassen und oftmals nur für einen bestimmten, engumgrenzten Zweck entwickelt worden sind (Ad-hoc-Skalen). Die Entwicklung derartiger Skalen ist in der heutigen Testdiagnostik zwar weitgehend üblich, führt erwiesenermaßen aber häufig zu Fehlschlüssen, weil die isolierte Betrachtung einzelner Funktionen des Menschen bei Vernachlässigung des Gesamts seiner Persönlichkeit allzu leicht zu einer Fehlschätzung seiner Berufstätigkeit führen kann. Gesellschaftspolitische Überlegungen brauchen in diesem Zusammenhang noch nicht einmal bemüht zu werden, obwohl offensichtlich ist, daß im Rahmen der propagierten Humanisierung der Arbeitswelt eine derartige Betrachtung menschlicher Berufstätigkeit völlig unangemessen ist. Zwar wird von einzelnen Testautoren die Notwendigkeit der Verwendung von Testbatterien in der Berufseignungsdiagnostik immer mit Eifer unterstrichen, doch sollte nicht verkannt werden, daß dieser berechtigten Forderung tatsächlich nur selten entsprochen wird und damit den ermittelten Meßwerten aus einzelnen Verfahren eine weitaus geringere Bedeutung

zukommt, als sich aus den berichteten statistischen Daten derselben Verfahren ableiten läßt.

Wenn man den Intelligenz-Struktur-Test von AMTHAUER unter den bisher genannten Gesichtspunkten überprüft, stellt man fest, daß die theoretische Fundierung dieses in der Praxis gut eingeführten Verfahrens nach wie vor unklar ist. AMTHAUER nimmt eine Struktur von allen intellektuellen Leistungen „zugrunde liegend gedachten Fähigkeiten“ an, die „in einem gefugehaften Zusammenhang stehen“ und eine Struktur aufweisen, „in der die Glieder hierarchisch geordnet sind“ (AMTHAUER, 1970, S. 7). In diesen Definitionsversuchen gehen Ansätze der bekannteren Testtheorien ein, ohne daß eine klare theoretische Positionsbestimmung erfolgt. Es kommt hinzu, daß offenbar von der Annahme ausgegangen wird, daß das Gesamt dieser seelisch-geistigen Fähigkeiten in Leistungen wirksam wird, ohne daß andere Gegebenheiten der Persönlichkeit in diesem Zusammenhang bedeutsamen Einfluß gewinnen könnten. Mit diesem Vorgehen werden Einsichten vernachlässigt, die seit Mitte der sechziger Jahre unübersehbar vorliegen: So zeigte sich beispielsweise neben den von CATTELL dargestellten Überlegungen in dem bei 450 000 amerikanischen Schülern in zweitägigen Untersuchungen über zehn Jahre hinweg durchgeführten „Projekt Talent“, daß die Meßwerte für Interessen und Grundeigenschaften eine bessere Voraussage für den individuellen berufsbildungsmäßigen Werdegang darstellen als die entsprechenden Meßwerte für die überprüften Fähigkeiten (COLLEY, 1966, S. 204).

Obwohl es unter theoretischen Gesichtspunkten kaum verantwortbar erscheint, diese Einsichten zu vernachlässigen, soll im folgenden in unmittelbarer Bezugnahme auf die Vorgehensweise von AMTHAUER untersucht werden, inwieweit dieses Vorgehen methodischen Ansprüchen zu genügen vermag oder inwieweit neben den dargestellten theoretischen Bedenken auch methodische Überlegungen den Untersuchungsansatz von AMTHAUER in Frage stellen.

III.

Methodische Überlegungen

Nach LIENERT werden an ein Testprofil vom teststatistischen Gesichtspunkt aus „keine allzu strengen Forderungen gestellt“ (LIENERT, 1969, S. 366). Demgemäß wird erwartet, „daß jeder Einzeltest einigermaßen reliabel ist und daß er mit jedem anderen möglichst niedrig korreliert. Desgleichen erwartet man, daß jeder Test unabhängig vom anderen eine ausreichende logische oder empirische Validität nachweisen kann“ (a. a. O., S. 366f).

Es versteht sich von selbst, daß der Einsatz von Testprofilen nur dann sinnvoll sein kann, wenn sich ein Testprofil aus der Zusammenstellung verschiedener Einzeltests ergibt, die möglichst genau das messen, was sie zu messen beanspruchen, und zwar ohne gemeinsam dasselbe zu messen. Nur wenn diese Voraussetzungen zu einem Mindestmaß erfüllt sind, läßt ein Testprofil methodisch vertretbare Aussagen zu. Es wird also zu prüfen sein, inwieweit die Berufsgruppenprofile, die AMTHAUER aufgrund seines Intelligenz-Struktur-Tests erstellt, diesen Anforderungen zu genügen vermögen.

Die von AMTHAUER berichteten statistischen Daten für die Genauigkeit (Reliabilität, Zuverlässigkeit) des Intelligenz-Struktur-Tests können sowohl für das Gesamtverfahren wie für die einzelnen Aufgabengruppen insgesamt zufriedenstellen. Es soll allerdings nicht die Verwunderung verschwiegen werden, die sich aus der Tatsache ergibt, daß AMTHAUER in der 3. Auflage seines Verfahrens aus dem Jahre 1970 dieselben Reliabilitätskoeffizienten berichtet wie in der Auflage von 1953. Die entsprechenden Daten beziehen sich auf eine Stichprobe von 100 Probanden. Das muß um so mehr verwundern, als AMTHAUER in seinem Vorwort zur 3. Auflage von 1970 darauf hinweist, daß in die Normierung des Verfahrens mehr als 30 000 Ergebnisse eingehen. Es kommt hinzu, daß die durch Testhalbierungsmethode gewonnenen und zu-

nächst zufriedenstellenden Koeffizienten numerisch weniger zufriedenstellend ausfallen, wenn sie durch Testwiederholung ermittelt werden. Diese Diskrepanz weist darauf hin, daß allein wegen der nicht voll überzeugenden Genauigkeit dieses Verfahrens der Interpretation von Berufsgruppenprofilen kein absolutes Vertrauen entgegengebracht werden kann. Darauf wird im folgenden noch näher eingegangen.

Diese Bedenken verstärken sich erheblich, wenn man die Korrelationsmatrix der neun verschiedenen Untertests des Verfahrens anschaut und feststellt, daß die Meßwerte von insgesamt 799 Personen, die bei der Ermittlung dieser Matrix zugrunde gelegt wurden, nicht unerheblich miteinander zusammenhängen. Diese Tatsache führt unmittelbar zu der Vermutung, daß die einzelnen Untertests nicht unabhängig voneinander das messen, was sie zu messen beanspruchen, sondern daß sie vielmehr in unterschiedlichem Maße mehr oder minder dasselbe messen. Diese Tatsache hat in den vergangenen Jahren verschiedene Forscher dazu veranlaßt, der Faktorenstruktur des Intelligenz-Struktur-Tests von AMTHAUER besondere Aufmerksamkeit zuzuwenden. In verschiedenen Untersuchungen (BÄUMLER, G. und WEISS, R., 1966; FISCHER, H., 1958; GREIF, 1972; LANGNER, E., OLBRICHT, E. und M., 1972; KURY, 1973, und LAPPE, H. und MÖNIKES, W. 1974) konnte nachgewiesen werden, daß der Intelligenz-Struktur-Test nicht den selbstgesetzten Anforderungen entspricht, Intelligenzstruktur von neun verschiedenen Teilbereichen der Intelligenz angemessen zu erfassen. Die genannten Untersuchungen kommen übereinstimmend zu der zuletzt von LAPPE und MÖNIKES getroffenen Feststellung, daß davon auszugehen ist, daß methodisch verantwortbar nur die Dimensionen der rechnerischen, figural-vorstellungsgebundenen und sprachlichen Intelligenz interpretiert werden können und daß daneben Aussagen über die Merkfähigkeit möglich scheinen (LAPPE und MÖNIKES, 1974, vgl. Tab. 1). AMTHAUER hat diese ihm wenigstens aus jüngeren Untersuchungen bekannte Tatsache auch in der dritten veränderten und erweiterten Auflage seines Intelligenz-Struktur-Tests nicht berücksichtigt, obwohl sie seinem ursprünglichen Ansatz weitestgehend entspricht, nach dem die Sonderstruktur der Intelligenz „einen Aufbau mit bestimmten einflußstarken Schwerpunkten“ zeigt, „welcher sich uns – in Leistungen wirksam geworden – vorerst als sprachliche, rechnerische Intelligenz, räumliches Vorstellen und Gedächtnisleistungen kundtun“ (AMTHAUER, 1953, S. 106).

Die Tatsache der Vernachlässigung dieser Forschungsergebnisse stellt die Berufsgruppenprofile von einer anderen Seite in Frage. Wenn richtig ist, was von verschiedenen Forschern in verschiedenen Forschungsansätzen zu verschiedenen Zeitpunkten mit verschiedenen Forschungsmethoden hinsichtlich des Zusammenhangs der Ergebnisse der neuen Untertests des Intelligenz-Struktur-Tests ermittelt worden ist, dann ist die Berücksichtigung der von AMTHAUER vorgegebenen und zur Interpretation vorgeschlagenen Berufsgruppenprofile insofern müßig, als nicht davon ausgegangen

Tabelle 1

Rotierte Faktorenlösungen für 2, 3, 4 und 5 Faktoren für die untersuchte Gesamtstichprobe

TEST	FAKTORENLÖSUNG			
	2	3	4	5
SE	54 —	50 — —	51 — — —	— — — — 53
WA	60 —	58 — —	52 — — (41)	57 — — —
AN	70 —	71 — —	69 — — —	60 — — — (39)
GE	65 —	64 — —	66 — — —	(32) — — — 59
ME	— —	— — 38	— — — 47	— — — 50 —
RA	51 (44)	(36) — 60	(38) — 61 —	— — 63 — (33)
ZR	(36) 61	— — 72	— — 62 (35)	— — 63 (34) —
FA	— 64	— 66 —	— 66 — —	— 66 — — —
WÜ	— 64	— 65 —	— 65 — —	— 65 — — —

werden kann, daß die das Profil ausmachenden Untertests tatsächlich alle Verschiedenes messen.

Damit ist aber auch bis zum Nachweis des Gegenteils in Frage gestellt, ob jedem einzelnen Untertest unabhängig vom anderen eine ausreichende logische und empirische Validität zukommt, was als entscheidende Voraussetzung für die Möglichkeit der Erstellung eines Testprofils eingangs festgehalten worden war.

Weiterhin muß bei der Analyse der statistischen Daten auch darauf hingewiesen werden, daß nach AMTHAUERS eigenen Angaben den einzelnen ermittelten Durchschnittswerten Standardabweichungen zuzuordnen sind, die im Mittel für die einzelnen Untertests drei Standardwertpunkte ausmachen. Das führt unter Berücksichtigung der entsprechenden Genauigkeitskoeffizienten zu einer weitestgehenden Egalisierung der von AMTHAUER vorgegebenen Berufsgruppenprofile. Berücksichtigt man diese Tatsache, so kann man mit AMTHAUERS Profilen eine Abiturientin nicht von einer Oberprüferin, einen Biologen nicht von einem Psychologen und einen Chemielaboranten praktisch nicht von einem Meister in der Chemie-Industrie zuverlässig unterscheiden.

In noch stärkerem Maße wird die Möglichkeit zur Interpretation von Berufsgruppenprofilen eingeschränkt, wenn man bei der Interpretation die Vertrauensintervalle der ermittelten Werte berücksichtigt. Es muß außerordentlich verwundern, daß AMTHAUER auch in der neuesten Auflage seines Verfahrens unter Hintanstellung der von ihm selbst berichteten statistischen Daten auf Interpretationsmöglichkeiten hinweist, die den Anforderungen der klassischen Testtheorie nicht genügen. So behauptet er, einen durch Zufall nicht mehr zu erklärenden Unterschied zwischen einzelnen Untertestergebnissen bei drei Standardwertpunkten, obwohl LIENERT im Jahre 1969 eine Profildifferenzenmatrix vorgelegt hat (LIENERT, S. 462), aus der eindeutig hervorgeht, daß bereits bei einer Sicherung auf dem 5-Prozent-Niveau erheblich höhere Werte notwendig sind, um von einem durch Zufall nicht mehr zu erklärenden Unterschied sprechen zu können (vgl. Tabelle 2). Legt man diese Profildifferenzenmatrix bei der Interpretation von Berufsgruppenprofilen unter Hintanstellung aller sonstigen bereits genannten Probleme zugrunde, so ergibt sich, daß in kaum einem der von AMTHAUER vorgeschlagenen Berufsgruppenprofile ein auffälliger Unterschied zwischen den Ergebnissen in verschiedenen Untertests sich zeigt. Von einem Profil kann in echtem Sinne nicht mehr gesprochen werden. Wenn die von AMTHAUER vorgeschlagenen Profile interpretiert werden, werden somit Scheinprofile interpretiert. Die Ergebnisse solcher Interpretationen

Tabelle 2

Profildifferenzenmatrix nach LIENERT (S. 462) *) **)

TEST	SE	WA	AN	GE	ME	RA	ZR	FA	WÜ	TOT
SE	8.3	9.0	9.4	7.8	8.5	8.1	7.1	9.8	8.8	6.8
WA		9.6	10.0	8.5	9.2	8.8	7.8	10.4	9.4	7.6
AN			10.4	9.0	9.6	9.2	8.3	10.7	9.8	8.1
GE				7.3	8.1	7.6	6.5	9.4	8.3	6.5
ME					8.8	8.3	7.3	10.0	9.0	7.1
RA						7.8	6.8	9.6	8.5	6.5
ZR							5.5	8.8	7.6	5.8
FA								11.1	10.2	8.5
WÜ									9.2	7.3
TOT										3.3

*) Eine entsprechende Profildifferenzenmatrix ist für das Leistungsprüfungssystem von HORN berechenbar, nicht aber für den Berufseignungstest von SCHMALE und SCHMIDTKE, weil dazu die notwendigen Statistiken fehlen

**) AMTHAUERS eher polemische Auseinandersetzung mit dieser Interpretationshilfe von LIENERT zeigt im Ende der fünfziger Jahre nicht auf der Höhe methodischer Anforderungen an Tests und trifft nach seinen eigenen Ausführungen (AMTHAUER, 1957, S. 168) nicht für die Interpretation interindividueller Vergleiche zu

müssen mit hoher Wahrscheinlichkeit zu falscher Beratung führen. Denn ein Germanist unterscheidet sich praktisch nicht von einem Diplomingenieur, ein Chemiker nicht vom Theologen und eine Stenotypistin nicht von einem im Bergbau Tätigen.

An dieser Tatsache vermag auch nicht zu ändern, daß für den Intelligenz-Struktur-Test ein durchaus zufriedenstellender Profilreliabilitätskoeffizient von 0.8620 ermittelt werden kann. An dem vorliegenden konkreten Beispiel ist gut erkennbar, wie man sich der Meinung von LIENERT anschließen kann, der darauf hinweist, daß der Reliabilitätskoeffizient eines Profils „nur einen allgemeinen allgemeinerorientierenden Eindruck“ vermittelt, der aber zur Interpretation von Profildifferenzen „praktisch unbrauchbar“ ist (LIENERT, 1969, S. 458).

Noch eine weitere kritische Anmerkung muß gegenüber den Berufsgruppenprofilen von AMTHAUER gemacht werden: Er stellt Profile verschiedener Berufsgruppen dar, für die jeweils das arithmetische Mittel der bei den einzelnen Aufgaben Gruppen erreichten Standardwerte berichtet wird. Nicht berichtet werden die entsprechenden Standardabweichungen. Da die Berufsgruppen lediglich Untergruppen der gesamten Eichstichprobe zu sein scheinen, muß zunächst untersucht werden, ob für den einen wie für den anderen Fall die von AMTHAUER selbst berichteten einzelnen Mittelwerte und Standardabweichungen zutreffen. Es muß daher davon ausgegangen werden, daß in der Gesamtstichprobe die ermittelten Punktwerte ebenso stark streuen wie in den einzelnen Berufsgruppen. Für differentielle Eignungstests aber wäre zu fordern, daß „deren Punktwerte in der Berufsgruppe wenig, in der Population aber sehr streuen“ (LIENERT, 1969, S. 375). Da dies im Intelligenz-Struktur-Test nicht der Fall ist, schließt LIENERT, daß sich der IST „besser zur Intelligenzmessung als zur Erfassung der Berufseignung“ eigne (LIENERT, S. 375). Man fragt sich — oder auch nicht —, warum derartige Hinweise kompetenter Autoren schlicht nicht zur Kenntnis genommen werden.

Nachdem somit aufgrund verschiedener Ansätze (Faktorenstruktur, Reliabilität) nachgewiesen ist, daß die Bezugnahme auf einzelne Untertestergebnisse im Rahmen eines Testprofils nicht die vom Testautor vorgeschlagenen Interpretationsmöglichkeiten eröffnet, muß auch noch erwähnt werden, daß die Berechnung einer globalen kritischen Differenz im Falle des Intelligenz-Struktur-Tests nicht in Frage kommt. Die Berechnung einer solchen kritischen Differenz ist nur dann möglich, wenn sich die Reliabilitätskoeffizienten der Einzelteste nicht signifikant unterscheiden. LIENERT hat bereits 1969 nachgewiesen, daß dies in bezug auf den Intelligenz-Struktur-Test von AMTHAUER der Fall ist (LIENERT, S. 464). Sie unterscheiden sich signifikant voneinander. Es bedeutet also keinen Ausweg aus dem ausgewiesenen Dilemma, wenn man versuchen wollte, Berufsgruppenprofile auf diesem Wege zu Einzelprofilen in Beziehung zu setzen. Vielmehr würde ein solches Vorgehen gegen grundlegende Regeln der statistischen Profilinterpretation verstoßen.

Es ist an anderer Stelle bereits darauf hingewiesen worden, daß in der Beratungspraxis häufig Profile intuitiv gedeutet werden und Profilvergleiche intuitiv durchgeführt werden. Unter methodischen Gesichtspunkten ist in diesem Zusammenhang ebenfalls die Benutzung eines statistischen Kennwerts zweckmäßig. In der Literatur liegen für diesen Zweck verschiedene Vorschläge vor, von denen im vorhinein jene ausgeschlossen werden können, in denen nur der Verlaufsgestalt, aber nicht der Profilhöhe Aufmerksamkeit zugewandt wird.

OSGOOD und SUCI haben 1952 einen Ähnlichkeitsindex D entwickelt, in dem sowohl die Profilhöhe als auch die Verlaufsgestalt bei der Ähnlichkeitsbeurteilung berücksichtigt werden. Es liegt nahe, einen solchen Ähnlichkeitsindex auch bei der Interpretation der Berufsgruppenprofile nach dem Intelligenz-Struktur-Test von AMTHAUER zu berücksichtigen.

Zwar sprechen die bislang dargestellten Kritikpunkte gegen ein solches Vorgehen, doch kann seine Benutzung nicht von vornherein ausgeschlossen werden. Sofern dies geschieht, ist allerdings darauf hinzuweisen, daß neben der Höhe des ermittelten Koeffizienten auch den jeweiligen Unterschieden in den einzelnen Untertests Aufmerksamkeit zugewandt werden muß. Falls die individuellen Meßwerte nämlich über den entsprechenden Gruppennormen liegen, wird dies die Höhe des Koeffizienten ebenso beeinträchtigen, wie wenn die individuellen Meßwerte unter den entsprechenden Gruppennormen liegen. Wenn man sich allerdings dem Grundsatz von AMTHAUER anschließt, daß es in der Berufsberatung nicht um eine Bestauslese oder Konkurrenzauslese gehen dürfe, sondern daß aus der Zahl der Bewerber jene Personen herauszufinden seien, „die den beruflichen Anforderungen am besten angepaßt sind“ (AMTHAUER, 1970, S. 17), dann müßte allein die Höhe der ermittelten Koeffizienten als Entscheidungskriterium dienen. Geht man von diesem Ansatz aus, kommt man allerdings zu überraschenden Ergebnissen: Selbst bei den von AMTHAUER als übereinstimmend bezeichneten Profilen zeigen sich Koeffizienten, die in der üblichen Interpretationsweise nicht auf eine sehr große Übereinstimmung hinweisen. Das mag an den eingangs erwähnten Voraussetzungen liegen, zeigt aber deutlich, daß auch über ein Ähnlichkeitsindex eine angemessene Interpretation der Berufsgruppenprofile nicht erfolgen kann.

Es muß zum Abschluß der methodischen Überlegung noch darauf hingewiesen werden, daß AMTHAUER zunächst mit außerordentlich kleinen Stichproben in den Berufsgruppenprofilen gearbeitet hat, die teilweise nicht einmal eine Stichprobe von 20 Personen erreichten. In der dritten veränderten und erweiterten Auflage sind die Stichproben insgesamt zwar größer geworden und ist die Zahl der berücksichtigten Berufe erweitert worden, doch muß man sich die Frage stellen, weshalb angesichts einer von AMTHAUER berichteten Zahl von 30 000 vorliegenden Ergebnissen immer noch mit Stichproben von 20 bis 50 Personen gearbeitet wird und bei bestimmten Berufsgruppen (höhere Beamte, Postbeamte im gehobenen Dienst, Refadient, Stenotypistinnen, Verwaltungsbeamte im gehobenen Dienst und anderen) die Stichproben nicht so groß gewählt worden sind, daß methodische Bedenken nicht erhoben werden müßten. Man hat an dieser wie an anderen Stellen der Handanweisung den Eindruck, daß nicht mit ausreichender Sorgfalt gearbeitet worden ist. Das ist um so bedauerlicher, als für den individuellen Fall als Folge einer durchgeführten Berufsberatung Entwicklungen eintreten können, die unter Berücksichtigung der objektiven und subjektiven Situation irreparable Schäden begründen können.

IV.

Praktische Konsequenzen

Es ist an anderem Ort nachgewiesen und hier nur kurz erwähnt worden, daß der Intelligenz-Struktur-Test von AMTHAUER seinen selbst gesetzten Ansprüchen nicht gerecht werden kann, neben dem allgemeinen Intelligenzniveau eines Menschen Einblick in die differenzierte Struktur seiner Intelligenz zu geben. Nach den vorliegenden Untersuchungen ist davon auszugehen, daß methodisch verantwortbar nur die Dimension der rechnerischen, figural vorstellungsgebundenen und sprachlichen Intelligenz interpretiert werden kann und daß daneben Aussagen über die Merkfähigkeit möglich zu sein scheinen.

Damit fällt die von AMTHAUER angedeutete Möglichkeit der Unterscheidung zwischen einer mehr sprachlich-theoretischen Begabung und einer praktischen Begabung. Die das entsprechende Profil bezeichnenden Untertests SE, WA, AN und GE markieren gemeinsam den Faktor der sprachlichen Intelligenz und können insofern zur von AMTHAUER vorgeschlagenen Differenzierung nichts Entscheidendes beitragen. Daraus folgt, daß diese im Zusammenhang mit den Berufsgruppenprofilen vorgeschlagenen Interpretationsmöglichkeiten ausfallen.

Wenn AMTHAUER für sein Verfahren in Anspruch nimmt, „daß es mit dem Intelligenz-Struktur-Test möglich ist, besser gesicherte und aufschlußreichere Aussagen als mit den bisher angewendeten Verfahren zur Bestimmung der Intelligenz zu machen“, und dann Beispiele für den Vergleich von Gruppenprofilen angibt, dann muß unter Bezugnahme auf das Dargelegte festgestellt werden, daß das Verfahren auch diesen Ansprüchen nicht zu genügen vermag (AMTHAUER, 1970, S. 16). Wenn AMTHAUER für sein Verfahren in Anspruch nimmt, daß es „Fehlansätze vermeiden kann und niemals in die Gefahr kommt, eine Bestauslese oder Konkurrenzauslese zu treiben im Sinne einer Auswahl der schlechtesten Intelligenztesten aus der Bewerbergruppe; (AMTHAUER, S. 17), dann muß festgestellt werden, daß dieser Anspruch ebenso übersteigert ist wie der Anspruch, durch dieses Verfahren Intelligenzstruktur in angestrebtem Ausmaß untersuchen zu können. Unter Berücksichtigung der Meßgenauigkeit des Verfahrens und des Zusammenhangs zwischen den einzelnen Untertests muß festgestellt werden, daß der Benutzer des Intelligenztests von AMTHAUER kein verlässliches Mittel in der Hand hat, „die Person aus der Zahl der Bewerber herauszufinden, die den beruflichen Anforderungen am besten angepaßt sind“ (AMTHAUER, 1970, S. 17).

Mit diesen Feststellungen soll nicht bestritten werden, daß die vom Testautor verfolgten Absichten durchaus begrüßenswerte Erweiterungen der berufsberatenden Möglichkeiten bedeutet hätten, wenn sein Verfahren den Gütekriterien der klassischen Testtheorie in ausreichendem Maße entsprochen hätte. Da dies unter verschiedenen Gesichtspunkten nicht der Fall ist, muß beim Stand der Entwicklung des Verfahrens deutlich davor gewarnt werden, jenen Interpretationshinweisen von AMTHAUER zu folgen, die sich auf die Möglichkeit der Interpretation von Berufsgruppenprofilen beziehen.

Wenn eingangs neben dem Intelligenz-Struktur-Test von AMTHAUER der Berufseignungstest von SCHMALE und SCHMIDTKE und das Leistungsprüfsystem von HORN genannt wurden, so deshalb, weil die Kritik, die im vorausgehenden in bezug auf den Intelligenz-Struktur-Test von AMTHAUER dargestellt wurde, sich mit nur geringen Abwandlungen auf den Berufseignungstest und das Leistungsprüfsystem übertragen läßt. Beide Verfahren nehmen wie der Intelligenz-Struktur-Test für sich in Anspruch, unter Zuhilfenahme vorgegebener Berufsgruppenprofile Aussagen über die Berufseignung von mit ihnen Untersuchten machen zu können. Nach den bislang vorliegenden Daten können auch diese Verfahren den selbstgesetzten Ansprüchen nicht in ausreichendem Maße genügen.

Im Berufseignungstest von SCHMALE und SCHMIDTKE werden zwar Anforderungsnormen berichtet, doch kann deren Bedeutung allein deshalb nicht eingeschätzt werden, weil weder die Standardabweichungen der entsprechenden Statistiken noch die entsprechenden Genauigkeitskoeffizienten berichtet werden. Es muß lediglich verwundern, daß die berichteten Anforderungsnormen für die zahlenmäßig teilweise sehr schwach repräsentierten Stichproben sich fast ausschließlich im Durchschnittsbereich der benutzten Z-Skala, also zwischen Standardwerten von 90 bis 110, bewegen. Unterstellt man die bei diesen Verfahren übliche Genauigkeit der Untersuchung, so bedeutet dies, daß eine echte Differenzierung zwischen gezeigten Leistungen kaum mehr erfolgen kann. Dabei wird allerdings vorausgesetzt, daß die verschiedenen Untertests Verschiedenes unabhängig voneinander messen. Diese Voraussetzung allerdings besteht zu Unrecht. Denn SCHMALE und SCHMIDTKE haben in einer eigenen Publikation aus dem Jahre 1969 festgestellt, daß die faktorielle Struktur des Berufseignungstests sich wesentlich von der der General Aptitude Test Battery (GATB) unterscheidet, auf den sie sich bei der Entwicklung des BET bezogen hatten. In einer Faktorenanalyse konnten lediglich vier interpretierbare Faktoren ermittelt werden, auf die bestimmte Syndrome aller Untertests in unterschiedlichem Maße laden. Es handelt sich dabei um Faktoren, die im Sinne allgemeiner

geistiger Beweglichkeit, motorischen Geschicks, räumlichen Sehens und mechanischer Fertigkeit sowie Wahrnehmungsgenauigkeit umschrieben werden.

Damit aber ist eine Situation gegeben, wie sie auch für den Intelligenz-Struktur-Test als bezeichnend dargestellt wurde. Weder scheinen die Einzeltests ausreichend genau zu sein, noch hängen sie untereinander möglichst niedrig zusammen. Da damit auch die ausreichende logische oder empirische Validität der einzelnen Untertests in Frage gestellt wird, muß vom theoretischen Standpunkt her die Möglichkeit der Interpretation von Berufsgruppenprofilen durch dieses Verfahren ebenfalls grundsätzlich in Frage gestellt werden.

Die Kritik am Leistungsprüfsystem von HORN, das im Jahre 1962 erschienen ist, kommt auf einem anderen Wege zu einem ähnlichen Ergebnis. Obwohl sich das Verfahren durch seine Anlehnung an THURSTONEs Intelligenztheorie vom theoretischen Konzept her wohltuend von dem zuvor erörterten Verfahren unterscheidet, muß doch festgestellt werden, daß der Testautor bereits in diesem Ansatz seinen selbstgesetzten Ansprüchen nicht genügt. Soweit bekannt ist, hat HORN sich zwar an THURSTONEs Faktoren orientiert, sie aber nicht selbst in Eigenuntersuchungen nachgewiesen. Er hat vielmehr nur entsprechende Aufgabenreihen in sein Verfahren aufgenommen und ist offenbar davon ausgegangen, daß sich in einer deutschen Stichprobe dieselbe Faktorenstruktur wiederfinden würde wie in der ursprünglichen Eichstichprobe bei THURSTONE. Eine vom methodischen Standpunkt her sehr naive Annahme. Es mußte daher Folgeuntersuchungen überlassen bleiben, die vorausgesetzte Struktur des Leistungsprüfsystems zu überprüfen. Dies ist zwar nicht direkt geschehen, doch zeigen die vorliegenden Arbeiten, die zusammenfassend bei GREIF dargestellt werden (GREIF, 1972, S. 121 ff.), „daß die von HORN angenommene Faktorenstruktur des Tests zumindest nicht präzise bestätigt werden konnte“ (a. a. O., S. 120). In eigenen Untersuchungen konnte GREIF vielmehr nachweisen, „daß die Faktorenstruktur des Tests mit vier Hauptfaktoren beschrieben werden kann“ (a. a. O., S. 130). Diese können mit räumlichem Vorstellungsvermögen, sprachgebundenem Denken, Einfallsgeschwindigkeit und Wortflüssigkeit umschrieben werden. Somit sind HORNs Annahmen über die faktorielle Struktur seines Verfahrens nur zum Teil bestätigt. Daraus muß mit testtheoretischer Notwendigkeit abgeleitet werden, daß eine Interpretation der Ergebnisse aus diesem Verfahren nicht entsprechend der von HORN unterstellten Faktorenstruktur vorgenommen werden kann, zumal einige der durchgeführten Aufgabenreihen auf anderen Faktoren laden, als von HORN unterstellt wurde.

Wenn schon eine Interpretation der Ergebnisse im Sinne intellektueller Teilbereiche aufgrund dieser Ergebnisse unmöglich erscheint, muß um so eindeutiger die Möglichkeit von Berufsgruppenprofilvergleichen, wie sie von HORN vorgeschlagen werden, ausgeschlossen werden. Mögen auch die Genauigkeitskoeffizienten für einzelne Aufgabenreihen und das Gesamtverfahren durchaus zufriedenstellend sein, so muß bei diesem Verfahren aufgrund der somit letztlich in Frage gestellten Validität des Verfahrens vor der Berücksichtigung von Berufsgruppenprofilen eindeutig gewarnt werden. Dies um so mehr, als HORN sich bei den Berufsgruppenprofilen auf außerordentlich kleine Stichproben bezieht, die zudem nicht näher umschrieben werden und somit nicht als verlässliche Bezugsstichproben berücksichtigt werden können.

Zusammenfassend muß daher festgestellt werden, daß weder der Berufseignungstest von SCHMALE und SCHMIDTKE noch der Intelligenz-Struktur-Test von AMTHAUER, noch das Leistungsprüfsystem von HORN den selbstgesetzten Ansprüchen genügen, soweit diese die Möglichkeit der Interpretation von Berufsgruppenprofilen im Sinne von Anforderungsnormen behaupten. Die hochgeschraubten Erwartungen, die sich für den Praktiker aus den Interpretationshinweisen der Testautoren für diesen Bereich ergeben, können aufgrund

testtheoretischer und methodischer Mängel der genannten Verfahren durchweg nicht erfüllt werden. Vor der Verwendung dieser Verfahren im Bereich der Berufsberatung muß insoweit gewarnt werden, als sie benutzt werden sollen, um die Eignung von Probanden für bestimmte Berufe zuverlässig vorauszusagen.

Als annehmbare Alternative bietet sich derzeit unseres Wissens lediglich die Testserie für Hauptschulabgänger (E. U. B.) an, die im Dezember 1969 in den Dienststellen der Bundesanstalt für Arbeit eingeführt wurde. Sie ist Bestandteil des „Test-Validierungs- und Interpretations-Systems“ und scheint testtheoretischen Ansprüchen in höherem Maße zu genügen als die hier kritisch dargestellten Verfahren. Die endgültige Einschätzung dieses Verfahrens muß einem späteren Zeitpunkt vorbehalten bleiben, zu dem die statistischen Daten dieses Verfahrens umfassend dargestellt werden.

Literatur

1. Amthauer, R.: Intelligenz und Beruf. Zeitschrift für experimentelle und angewandte Psychologie 1, 1953, S. 102–144.
2. Amthauer, R.: I-S-T. Intelligenz-Struktur-Test. 2. erweiterte und verbesserte Auflage. Göttingen: Verlag für Psychologie 1953
3. Amthauer, R.: Über die Prüfung der Zuverlässigkeit von Tests – erörtert am Intelligenz-Struktur-Test. Psychologische Rundschau 8, 1957, S. 165–171.
4. Amthauer, R.: Intelligenz-Struktur-Test (I-S-T 70). 3. veränderte und erweiterte Auflage. Göttingen: Verlag für Psychologie 1970.
5. Bäumler, G. und Weiß, R.: Über den Zusammenhang der Paulitestleistung mit Intelligenzleistungen (IST-AMTHAUER, CFT-CATTELL). Psychologie und Praxis 10, 1966, S. 27–36.
6. Cattell, R. B.: Die empirische Erforschung der Persönlichkeit. Weinheim und Basel: Beltz 1973.
7. Cooley, W. W.: Predicting career plan changes. In: Flanagan, J. C. und Cooley, W. W. (Hrsg.): Projekt TALENT one-year follow-up studies. Pittsburgh: University of Pittsburgh, School of Education 1966, S. 181–204.
8. Fischer, H.: Ein Vergleich zwischen dem IST von AMTHAUER und den PMA von THURSTONE. Diagnostica 4, 1958, S. 25–32.
9. Greif, S.: Gruppenintelligenztests – Untersuchungen am WIT, IST, LPS und AIT – Bern: Herbert Lang, 1972.
10. Horn, W.: Leistungsprüfsystem (LPS). Göttingen: Verlag für Psychologie 1962.
11. Kury, H.: Vergleichsuntersuchungen an Psychologiestudierenden mit dem Intelligenzstrukturtest (IST) von AMTHAUER und dem Leistungsprüfsystem (LPS) von HORN. Psychologie und Praxis 17, 1973, S. 57–67.
12. Langner, E.; Olbrich, E. und M.: Untersuchungen zur faktoriellen Struktur des IST von AMTHAUER – mit einem Vergleich der Ergebnisse der Intelligenztests der FVP. Wehrpsychologische Untersuchungen 3/72, 1972, S. 7–73.
13. Lappe, H. und Mönikes, W.: Zur Faktorenstruktur eines Intelligenzstrukturtests (I-S-T 70). Diagnostica 20, 1974.
14. Lienert, G. A.: Testaufbau und Testanalyse. Weinheim, Berlin, Basel: Beltz 1969.
15. Osgood, Ch. E. und Suci, G. J.: A Measure of Relation Determined by both Mean Differences and Profile Information. Psychological Bulletin 49, 1952, S. 251.
16. Schmale, H. und Schmidtke, H.: BET Tabellenband. Bern und Stuttgart: Hans Huber 1967.
17. Schmale, H. und Schmidtke, H.: Eignungsprognose und Ausbildungserfolg. Köln und Opladen: Westdeutscher Verlag 1969 (Forschungsberichte des Landes Nordrhein-Westfalen Nr. 2044).

Ute Laur

Forschung im Modellversuch MME

In diesem Aufsatz wird die sozialwissenschaftliche Untersuchung zum Modellversuch „Mehrmediensystem Elektrotechnik/Elektronik“ (MME) vorgestellt. Sie befaßt sich mit dem Problem der Effektivität eines mit Hilfe objektiver Lehrsysteme gestalteten Unterrichts. Dabei liegt das Schwergewicht auf der Analyse seiner „affektiven Nebenwirkungen“, also seiner zwar nicht intendierten, aber auftretenden Auswirkungen auf Einstellungen, Motive, Verhaltenserwartungen usw. der Beteiligten und der Frage nach den sozialen, persönlichkeitspezifischen und bildungstechnologischen Bedingungen für erfolgreiches Lernen mit diesen Lehrsystemen des MME.

Vorbemerkung

Das vom BBF entwickelte „Mehrmediensystem Elektrotechnik/Elektronik“ (MME)¹⁾ ist in seiner gegenwärtigen Ausbaustufe Gegenstand eines vom BMBW geförderten Modellversuchs, an dem 56 berufsbildende Schulen aus allen Bundesländern beteiligt sind. Er wird seit Februar 1973 unter Leitung des BBF realisiert und sollte – so war von Anfang an die Zielsetzung der an ihm arbeitenden Projektgruppe – auch für Forschung genutzt werden. Nach unserem Verständnis von Bildungstechnologie gehört systematischer,

empirisch begründeter Erkenntnisgewinn zu ihren Aufgabenbereichen; er stellt eine wichtige Voraussetzung für die Entwicklung und laufende Verbesserung von Mitteln und Verfahren zur erfolgreichen Gestaltung von Lehr-Lern-Processen dar. Demgemäß wurde die Durchführung einer sozialwissenschaftlichen Untersuchung im Rahmen des Modellversuchs eingeplant. Bisher haben sich an diesem Forschungsvorhaben („Feldversuch“), das über etwa zwei Jahre laufen soll, 200 Klassen, d. h. ca. 5000 Schüler und eine relational entsprechende Anzahl von Lehrern beteiligt. Der Rücklauf der erhobenen Daten aus den Versuchsklassen ist voll im Gange und mit ersten Teilauswertungen wurde begonnen. Es wäre jedoch verfrüht, schon an dieser Stelle auf Ergebnisse einzugehen; dazu sind sie noch zu fragmentarisch. Deshalb beschränken wir uns im folgenden auf die Darstellung der Anliegen und Probleme, der Konzeption und Auswertungsstrategie des „Feldversuchs“ – soweit dies auf dem begrenzten Raum möglich ist.

1. Funktion und Ziele des Feldversuchs

Die sozialwissenschaftliche Untersuchung ist Bestandteil des Modellversuchs, der ihren gesellschaftlichen und bildungspolitischen Bezugsrahmen sowie ihr Aktionsfeld bestimmt. Die Zielsetzungen dieser beiden im Kontext realisierten Maßnahmen sind jedoch nicht identisch: Mit dem

¹⁾ Siehe hierzu auch den Informationskasten auf S. 18 der ZfB. 4/73