

Nr.8

(Mai 1976)

Lothar P. Schardt:

Ansätze zu einer arbeitsorientierten

Eignungsdiagnostik

Zusammenfassung

Ausgehend von dem Nachweis, daß die herkömmliche berufliche Eignungsdiagnostik aufgrund des konservativen und kapitalorientierten Charakters der ihr zugrundeliegenden klassischen Testtheorie in bedeutsamem Ausmaß Anpassungsergebnisse und -potentiale erfaßt, wird die Notwendigkeit einer alternativen, arbeitsorientierten Eignungsdiagnostik abgeleitet und unter wissenschaftstheoretischen und gesellschaftspolitischen Gesichtspunkten diskutiert. Auf dieser Grundlage werden die psychometrische Einzelfalldiagnostik, lernzielorientierte Tests und die probabilistischen Testmodelle vorgestellt und daraufhin überprüft, ob und inwieweit diese neueren testtheoretischen Ansätze eine Veränderung der eignungsdiagnostischen Praxis implizieren und für die Durchsetzung arbeitsorientierter Interessen als relevant anzusehen sind. Abschließend wird versucht, die sich aus der Diskussion dieser Ansätze sowie einer früheren kritischen Analyse (SCHARDT 1976) ergebenden Aufgaben für Praxis und Forschung zu konkretisieren.

(Diese Arbeit erscheint in: TRIEBE, J.K. & ULICH, E. (Hrsg.), Beiträge zur Eignungsdiagnostik. (Schriften zur Arbeitspsychologie; Hrsg. v. E. ULICH) Bern: Huber, 1976.)

1. Kapitalorientierte vs. arbeitsorientierte Eignungsdiagnostik

Für die Begründung der Notwendigkeit, alternative testdiagnostische Ansätze zu konzipieren, ist der Nachweis konstitutiv, daß die Eignungsdiagnostik in ihrer derzeit betriebenen Form nicht dazu in der Lage ist, die ihr zugewiesenen oder zuzuweisenden Funktionen zu erfüllen. Dieser Nachweis war Gegenstand einer Untersuchung von SCHARDT (1976) über die Funktion der derzeit praktizierten Eignungsdiagnostik im allgemeinen und der psychometrischen Testdiagnostik im besonderen, wobei die methodologischen und wissenschaftstheoretischen Grundlagen, wie sie in dem Modell der Klassischen Testtheorie (KTT) impliziert sind (FISCHER 1968, LORD & NOVICK 1968) im Vordergrund standen. Diese Analyse führte zu dem Ergebnis, daß die KTT und die darauf gründende psychometrische Eignungsdiagnostik in bedeutsamem, wenn auch kaum quantifizierbarem Ausmaß Anpassungspotentiale und -ergebnisse erfaßt und sich somit in ihrem solchermaßen ausgewiesenen konservativen Charakter als weitgehend kapitalorientiert erweist. Daraus wurde, neben einer Reihe von "flankierenden" Maßnahmen, die Notwendigkeit der Konzipierung alternativer Testmodelle sowie kurz- und mittelfristig praktikabler Strategien zur Eindämmung und Beseitigung dieser konservativen Tendenzen abgeleitet.

Sofern man bereit und in der Lage ist, diese Einschätzung der Testdiagnostik - zumindest in wesentlichen Teilen - zu akzeptieren und unterstellt des weiteren eine in dem Lohnarbeitsverhältnis begründete, antagonistische Interessenstruktur von Kapitaleignern auf der einen und Lohnabhängigen auf der anderen Seite, so stellt sich für jeden Eignungsdiagnostiker die Frage, in wessen Interesse er tätig werden will und - sofern man den Betriebspsychologen nicht als bloß "abstrakten", in seinen scheinbar nur "fachlichen" Entscheidungen freien Lohnarbeiter unterstellt - tätig werden kann! Stellt man in Beantwortung dieser Frage den Anspruch, die Interessen der Betroffenen zum Maßstab der eigenen Tätigkeit zu machen, insofern also eine "Psychologie von unten" zu betreiben, so ist die Schlußfolgerung nicht mehr von der Hand zu weisen: Unumgänglich ist die Einschränkung und schließlich auch Beseitigung der konservativen qua kapitalorientierten Tendenzen der Testdiagnostik und deren Ersetzung durch alternative, somit arbeitsorientierte eignungsdiagnostische Testmodelle und -instrumentarien. Diese haben sich dadurch auszuzeichnen, daß sie von der Prämisse einer lebenslangen Entwicklung personaler und sozialer individueller Fähigkeiten via beruflicher Sozialisation ausgehen - entweder im negativen Sinne, nämlich der Verhinderung von "Sozialisationsschäden" oder im positiven

Sinne, nämlich der Höherentwicklung dieser Fähigkeiten (SCHARDT 1976).

Der einer solchen Position regelmäßig entgegengehaltene, der bekannten sozialpartnerschaftlich-pluralistischen Denkschablone entspringende Einwand, der Betriebspsychologe müsse sowohl den Kapitaleignern als auch den "Mitarbeitern" gleichermaßen zu Diensten stehen - ein Anspruch, der schon durch die tägliche betriebspsychologische Praxis ad absurdum geführt wird - sei hier nur mit dem Faktum konfrontiert, daß im medizinischen (einschließlich werksärztlichen) und klinisch-psychologischen Bereich das "Wohl" der Betroffenen (und nicht der Auftraggeber!) als primäres berufsständisches Kriterium - zumindest dem Anspruch nach - völlig unbestritten ist. Allein schon die potentiellen psychopathologischen, psychosozialen u.a. Folgewirkungen der Eignungsdiagnostik (SCHARDT 1976) lassen unter diesem Einwand die ansonsten völlig übereinstimmend vertretenen psychologisch-berufsständischen Normen zur bloßen Farce werden.

Mit der Wahl des Begriffes "arbeitsorientiert" soll nicht nur der als Antithese zu der derzeitigen kapitalorientierten Praxis der Eignungsdiagnostik vertretene, durch antagonistische Interessen ausgezeichnete Ansatzpunkt hervorgehoben, sondern auch auf eine, wissenschaftstheoretisch und -soziologisch analoge Entwicklung in einer anderen Disziplin hingewiesen werden: Von einer Projektgruppe des Wirtschafts- und Sozialwissenschaftlichen Instituts (WSI) des DGB wurde die Konzeption einer "Arbeitsorientierten Einzelwirtschaftslehre" vorgelegt (WSI-Forum 1973, Projektgruppe im WSI 1974), mit der in explizitem Gegensatz zu der "Kapitalorientierten Betriebswirtschaftslehre" der Anspruch verknüpft wird, die Durchsetzung "arbeitsorientierter Interessen" (MARKMANN) eher zu ermöglichen. In gleicher Weise gilt auch für die hier vorgestellten Überlegungen zur Entwicklung einer arbeitsorientierten Eignungsdiagnostik, "... die Ziel- und Interessenkonflikte zu verdeutlichen, von welchen praktisches Handeln gerade für die Arbeitnehmer und ihre Vertreter in hohem Maße bestimmt wird. Somit wird zumindest ein Teil der zu erwartenden Kritik...durch unterschiedliche gesellschaftliche Interessen und Standorte geprägt sein. Wir, die wir im gewerkschaftlichen Bereich tätig sind, haben dies immer wieder erfahren, doch meinen wir, daß das Aufzeigen und das Offenlegen dieser Art von gesellschaftlichen Bezügen, ihre Konfliktbetontheit und das darauf beruhende Selbstverständnis ein wesentlicher Bestandteil von Wissenschaft ist, die nicht als l'art pour l'art betrieben wird, sondern als eine Tätigkeit, die sich im Vorfeld politischer und wirtschaftlicher Entscheidungen und Handlungen vollzieht" (MARKMANN im Geleitwort zu: Projektgruppe im WSI 1975, S.4).

Um Mißverständnissen vorzubeugen, sind noch einige Relativierungen und Präzisierungen des Terminus "arbeitsorientiert" im hier gebrauchten Sinne erforderlich. Indem die Interessen des "Faktors Arbeit" als Ausgangspunkt und Ziel der Bemühungen ausgewiesen werden, wird keineswegs unterstellt, daß unter der der-

zeitigen Organisationsform gesellschaftlicher Arbeit - deren wesentlichstes Kennzeichen der Privatbesitz an Produktionsmitteln ist - eine ausschließlich die Interessen der abhängig Beschäftigten reflektierende Eignungsdiagnostik praktikabel wäre. Vielmehr muß der untergeordnete (konsultative) Stellenwert und die vielfach vermittelte Abhängigkeit der Eignungsdiagnostik von anderen Parametern der Unternehmensplanung sowie den industriellen Beziehungen auf betrieblicher und gesellschaftlicher Ebene (SCHARDT 1976) in Rechnung gestellt werden, so daß auch im eignungsdiagnostischen Bereich immer nur eine partielle, schrittweise auszuweitende Durchsetzung arbeitsorientierter Interessen möglich sein wird.

Des weiteren schließt das hier nicht abgeleitete Postulat antagonistischer Interessen als zentrales Charakteristikum gesellschaftlicher Verhältnisse keineswegs partielle Interessenübereinstimmungen bei konkreten betrieblichen Problemen aus - eine Chance, die im Interesse der Betroffenen unbedingt genutzt werden sollte. D.h. konkret: Wenn die Möglichkeit besteht, mit Hilfe der Eignungsdiagnostik Über- oder Unterforderungen am Arbeitsplatz zu vermeiden, so ist dies nicht nur der physischen, psychischen und sozialen Befindlichkeit der Arbeitenden zuträglich, sondern führt auch in der Regel - zumindest längerfristig - zu einer ertragssteigernden Erhöhung der Arbeitsproduktivität. In dieser Widersprüchlichkeit stellt die Eignungsdiagnostik keineswegs einen Sonderfall dar; diese ist vielmehr geradezu ein Charakteristikum vieler arbeitswissenschaftlicher Teilbereiche (z.B. der Ergonomie). Für die "Neuen Formen der Arbeitsgestaltung" wurde diese Problematik in fast schon exemplarischer Form von VOLPERT (1974) und GROSKURTH (in GROSKURTH & VOLPERT 1975) und für den werksärztlichen Dienst von REGUS (1973) auf der Basis der Untersuchung von DIEHR, FLAKE, HÖTTEMEISTER & STUMPFIG (1970) herausgearbeitet. Anstelle weiterer Ausführungen zu diesem Aspekt sei auf die genannten Arbeiten verwiesen.

Auf der Grundlage dieser Relativierungen kann eine terminologische Präzisierung vorgenommen werden: Mit "arbeitsorientiert" sollen nicht nur solche Ansätze gekennzeichnet werden, in die ausschließlich die Interessen der Betroffenen eingehen, sondern auch solche, mit deren Hilfe "lediglich" eine partielle Durchsetzung dieser Interessen ermöglicht wird. M.a.W., eine arbeitsorientierte Eignungsdiagnostik ist auch gerade auf solche kurz- und mittelfristig realisierbaren Strategien angewiesen, mit deren Hilfe eine Einschränkung der konservativen Tendenzen der kapitalorientierten Eignungsdiagnostik ermöglicht wird.

Ziel der folgenden Ausführungen ist es, bislang vorliegende testdiagnostische Ansätze daraufhin zu untersuchen, ob und wie, d.h. unter welchen Bedingungen mit

ihrer Hilfe eine arbeitsorientierte Eignungsdiagnostik im oben definierten Sinne realisiert werden kann. Dabei sollen kurz- und mittelfristig realisierbare Ansätze im Vordergrund stehen. Zumindest dem Fachmann wird es nicht verborgen bleiben, daß die folgenden Überlegungen über weite Strecken stark programmatischen Charakter tragen. Dieser Eindruck ist insofern willkommen, als damit die Hoffnung verbunden wird, zu einer Weiterverfolgung der vorgestellten Ansätze anzuregen, um so in absehbarer Zeit sowohl deren Realisierbarkeit als auch Brauchbarkeit für eine arbeitsorientierte Eignungsdiagnostik auf empirisch-wissenschaftlicher und empirisch-praktischer Grundlage besser beurteilen zu können.

2. Neue Ansätze in der Testtheorie

Die im folgenden darzustellenden neueren Entwicklungen in der Testtheorie werden jeweils unter drei Gesichtspunkten diskutiert:

- (1) den methodologischen und ggf. wissenschaftstheoretischen Grundlagen,
- (2) den Konsequenzen für die eignungsdiagnostische Praxis, insbesondere im Hinblick auf die mit ihrer Realisierung verbundenen Probleme,
- (3) der Kritik dieser Ansätze im Hinblick auf die Beurteilung ihrer Brauchbarkeit für eine arbeitsorientierte Eignungsdiagnostik.

Bei dieser Darstellung steht immer der Bezug zu der KTT und der darauf gründenden eignungsdiagnostischen Praxis im Vordergrund. M.a.W., die methodologischen, methodischen und praktischen Aspekte werden hier nur insoweit behandelt, als sie

- eine Veränderung der derzeitigen eignungsdiagnostischen Praxis implizieren, und
- für eine Durchsetzung arbeitsorientierter Interessen als potentiell relevant anzusehen sind.

Insofern soll und kann also kein Anspruch auf Vollständigkeit erhoben werden (für eine umfassende Darstellung dieser Ansätze sei stattdessen auf die Originalliteratur verwiesen). Folgende Ansätze werden behandelt:

- Psychometrische Einzelfalldiagnostik
- Lehrzielorientierte Tests
- Probabilistische Testmodelle.

Es sei darauf hingewiesen, daß - trotz des notwendigerweise fragmentarischen Charakters ihrer Darstellung - jedes Begründungselement für diese Ansätze implizit eine Kritik an der KTT bzw. der von ihr bestimmten eignungsdiagnostischen Praxis enthält.

Im Anschluß daran wird versucht, die sich aus der Diskussion dieser Ansätze sowie der Kritik von SCHARDT (1976) ergebenden Forschungsaufgaben zu konkretisieren.

2.1 Psychometrische Einzelfalldiagnostik

2.1.1 Methodologische Aspekte

Das zentrale Anliegen der Psychometrischen Einzelfalldiagnostik (PED) formuliert HUBER im Vorwort zu seiner gleichnamigen Monographie folgendermaßen: "Die klassische Testtheorie wurde als Grundlage der Testkonstruktion häufig diskutiert und kritisiert. Dagegen wurde die Frage, ob und in welchem Ausmaß das klassische Reliabilitätskonzept eine theoretische Basis für die zufallskritische Auswertung von Testbefunden vorsieht, in auffallender Weise vernachlässigt. Vielleicht ist dies einer der Gründe für die fortschreitende Dissoziation zwischen Testtheorie und Testdiagnostik... (Unter dem Gesichtspunkt einer) rationalen Auswertung psychometrischer Einzelfallbefunde...wurde hier der Versuch unternommen, jene Informationen, die in der Kenntnis um die Verteilung der Eichstichprobe, der Testinterkorrelationen oder der Skalenreliabilitäten enthalten sind, zum Zwecke einer differenzierten und zufallskritisch orientierten Einzelfalldiagnostik anzuwenden" (1973, S.V; Hervorhebungen im Original).

Mit Hilfe der PED soll die Feststellung ermöglicht werden, ob ein bestimmtes Individuum über eine bestimmte Merkmalsausprägung (Kriterium; z.B. Therapieerfolg) verfügt oder nicht, indem versucht wird, "...bestimmte diagnostische Hypothesen experimentell zu verifizieren und nach Möglichkeit zufallskritisch abzusichern". Als Anwendungsgebiet bieten sich neben therapeutischen Kontrolluntersuchungen und Forschungsexperimenten vor allem diagnostische Untersuchungen an, die im vorliegenden Zusammenhang wesentlich sind (HUBER 1973, S.34).

HUBER beschreibt sechs verschiedene klinisch-psychologische Anwendungsbereiche solcher experimenteller Einzelfallstudien (1973, S.37 ff.). Für die betriebspsychologische Praxis wesentlich sind dabei solche diagnostischen Einzelfallexperimente, die als psychometrische Einzelfalluntersuchungen konzipiert sind und eine Sonderform darstellen. "Wir meinen damit Fallstudien manipulativer oder nicht-manipulativer Art (vgl. HUBER 1973, S.34 ff; L.P.S.), die unter kontrollierten Bedingungen mit quantitativ auswertbaren Testverfahren durchgeführt werden, deren Normierungs- und Standardisierungsdaten (z.B. Erwartungswert und Standardabweichung in der Eichpopulation, Reliabilität, Testinterkorrelation etc.) bekannt sind. Das Kardinalproblem einer zufallskritischen Beurteilung von Einzelfallbefunden besteht im allgemeinen in der erwartungstreuen Schätzung der Fehlervarianz zeitabhängiger Beobachtungswerte. Bei psychometrischen Einzelfalluntersuchungen ist die inferenzstatistische Problematik insofern anders gelagert, als die Varianz der Beobachtungsfehler nicht am untersuchten Probanden selbst

geschätzt werden muß, sondern unter bestimmten Voraussetzungen aus den Normierungs- und Standardisierungsdaten hergeleitet werden kann" (HUBER 1973, S. 42; Hervorhebung von mir, L.P.S.).

Ergebnis der PED ist eine zufallskritisch abgesicherte Diagnose, die auf der Testleistung eines Individuums einerseits und der zu erwartenden Leistung der Referenzpopulation andererseits beruht. Zwar beinhaltet auch die konventionelle Testdiagnostik einen solchen Vergleich, indem nämlich Abweichungen über einen bestimmten Bereich hinaus als normdeviant angesehen werden (bei standardisierten Tests wird dieser Bereich durch den Erwartungswert - 100 im Falle der IQ-Transformation - fixiert und beträgt in der Regel zwei Standardabweichungen). Der Unterschied zur PED liegt jedoch darin, daß die zufallskritische Diagnose auf die wahren Werte inferiert. Letztere sind aber "...nach den Modellvorstellungen der klassischen Testtheorie...nicht beobachtbar. Umso mehr sind wir daran interessiert, auf der Grundlage der beobachteten Testleistungen und der Kenntnis der stochastischen Eigenschaften eines Tests Informationen über einen Skalenbereich zu erhalten, innerhalb dessen die wahre Testleistung eines Probanden mit hoher Wahrscheinlichkeit zu erwarten ist" (HUBER 1973, S.116). Daraus resultieren zwei praktische Aufgaben, nämlich die Festlegung der Vertrauenswahrscheinlichkeit (unter Berücksichtigung der jeweiligen Bedeutung des Fehlers 1. und 2.Art) und die Wahl des Verfahrens zur Schätzung des Vertrauensintervalls (per Äquivalenz- oder Regressionshypothese; HUBER 1973, S.116 ff.).

Der wahre Wert ist dabei definiert als der Erwartungswert einer test- und probandenspezifischen, d.h. intraindividuell definierten Zufallsvariablen - dies ist der axiomatisch zentrale Unterschied zur KTT - wobei die intraindividuelle Varianz der Testleistungen als testspezifische Fehlervarianz bezeichnet wird (analog dem "Fehleraxiom" der KTT; HUBER 1973, S.47 ff.). Diese Varianz der "Fehlervariablen" ist gleich der Varianz der Beobachtungsvariablen, da sich der Meßfehler aus der Differenz zwischen beobachteten und wahren Werten ergibt und der Erwartungswert des Fehlers als Null postuliert wird. Insofern kann die Reliabilität eines Tests als Maß für die Zuverlässigkeit konzipiert werden, mit der aufgrund eines einzelnen Beobachtungswertes der wahre Testwert vorhergesagt werden kann.

Dies setzt jedoch voraus, daß die Varianz der Beobachtungsvariablen bekannt ist. Als einzige Information liegt hierzu in der Regel lediglich der Reliabilitätskoeffizient vor. Da HUBER zeigen kann, "...daß die Varianz der Beobachtungsfehler in einer definierten Probandenpopulation gleich dem Erwartungswert der test-

spezifischen Fehlervarianz ist" (1973, S.60) - nach HUBER das "Fundament" der PED -, kann im Falle der (annähernden) Gleichheit der spezifischen Fehlervarianzen innerhalb einer Population die testspezifische Gruppenvarianz als guter Näherungswert für die testspezifischen Fehlervarianzen betrachtet werden. Die testspezifische Gruppenfehlervarianz, die einen bestimmten Test kennzeichnet, kann ihrerseits - neben der direkten Schätzung auf der Grundlage der testspezifischen Fehlervarianzen - indirekt über die Reliabilität des jeweiligen Tests (als Quadrat des Standardmeßfehlers) errechnet werden.

Sollen jedoch mehrere Testwerte miteinander verglichen werden, so stellt sich das Problem der Vergleichbarkeit, nämlich das der Normierung. Wird eine Normierung einer interindividuellen Beobachtungsvariablen angestrebt und dabei Gleichheit oder zumindest Ähnlichkeit der zugrundeliegenden, die jeweiligen Personen und den jeweiligen Test kennzeichnenden Zufallsvariablen unterstellt, spricht HUBER von einer x-Normierung (1973, S.68 ff.). Die x-Normierung beruht also auf der Verteilung der Beobachtungsvariablen (Mittelwert und Streuung) und zielt auf einen Vergleich der Beobachtungswerte. Ein solcher Vergleich von Beobachtungswerten impliziert aber notwendigerweise den Rekurs auf die entsprechenden wahren Werte insofern, als eine Äquivalenz von wahren und Beobachtungswerten unterstellt wird.

"Die Bedingungen, unter welchen die Äquivalenz der x-normierten wahren Testwerte ...gilt, haben für die Testdiagnostik schwerwiegende Restriktionen zur Folge. So kann beispielsweise bei der zufallskritischen Beurteilung einer Testwertdifferenz zwischen zwei Subtests...eines Testprofils nur dann eine sinnvolle, d.h. der diagnostischen Fragestellung entsprechende Nullhypothese formuliert werden, wenn gleichzeitig die Gleichheit der Reliabilitäten der betreffenden Tests vorausgesetzt wird...eine Alternative zur generell üblichen x-Normierung ist daher dringend erforderlich" (HUBER 1973, S.75). Umgekehrt formuliert ist ein zufallskritischer Vergleich von x-normierten Testergebnissen - was praktisch höchst bedeutsam ist - nur dann möglich, wenn die zugehörigen spezifischen Gruppenfehlervarianzen als homogen vorausgesetzt werden können.

Eine solche Alternative schlägt HUBER mit der tau-Normierung vor: Diese beruht auf der Verteilung der wahren Testwerte in der Eichpopulation, wobei unterstellt wird, daß die Verteilungen verschiedener Individuen bezüglich der Ergebnisse in einem bestimmten Test gleich oder zumindest ähnlich sind. Der entscheidende Vorteil der tau- gegenüber der x-Normierung liegt darin, daß aus der Äquivalenz der nichtnormierten wahren Testwerte auch die Äquivalenz der tau-normierten wahren Testwerte folgt, ohne daß weitere Zusatzannahmen erforderlich wären (HUBER 1973, S.74 ff.).

Eine individuelle Testleistung wird im Rahmen der PED unter zwei Aspekten zufallskritisch beurteilt:

(1) Zuerst werden die vorliegenden Statistiken (Testwerte, -differenzen, -profile) unter dem Aspekt der Reliabilität miteinander verglichen, d.h. es wird die Frage gestellt, ob die Reliabilitäten ihrerseits so reliabel sind (praktische Invarianz), daß eine beobachtete Abweichung mit vorher festgesetzter Wahrscheinlichkeit realiter gegeben ist (bzw. die beobachteten Abweichungen nicht Meßfehlern zu schulden sind).

(2) Können diese Abweichungen als statistisch bedeutsam, d.h. reell angesehen werden, so stellt sich die Frage, mit welcher Wahrscheinlichkeit eine Abweichung in der beobachteten Größe in einer Referenzpopulation zu erwarten ist. Dies ist die Frage nach der diagnostischen Valenz. Erst wenn realen Abweichungen zusätzlich auch diagnostische Valenz zukommt, d.h. eine Abweichung der vorliegenden Größenordnung in Bezug auf die Referenzpopulation statistisch bedeutsam ist, lassen sich Testresultate als "normdeviant" im herkömmlichen Sinne der Psychodiagnostik interpretieren.

2.1.2 Praktische Konsequenzen

Die Reliabilitäten dienen in der PED - im Unterschied zur deskriptiven Verwendung in der KTT - über die Beurteilung der praktischen Invarianz als "Präkursorgrößen" zur Bestimmung des Stichprobenfehlers individueller Testwerte. HUBER zeigt, daß der Stichprobenfehler geschätzter Reliabilitätskoeffizienten dann vernachlässigt werden kann, wenn der der Reliabilitätsermittlung zugrundegelegte Stichprobenumfang mindestens $N = 400$ beträgt. Die Kenntnis dieser Statistik ist also zur Beurteilung der Präzision einer Reliabilitätsschätzung unumgänglich. Für kleinere Stichprobenumfänge gibt HUBER statistische Verfahren zur Überprüfung der praktischen Invarianz an (1973; S.87 ff.). Eine zufallskritische Diagnose ist nur unter der Voraussetzung der praktischen Invarianz der Reliabilitäten möglich.

Im einfachsten Fall wird bei gegebener Invarianz eine statistische Inferenz über den wahren Testwert durchgeführt, wobei das Signifikanzniveau und das Schätzverfahren für das Vertrauensintervall festzulegen ist (HUBER 1973, S.113 ff.). Auf dieser Grundlage kann die nunmehr zufallskritisch abgesicherte Aussage getroffen werden, daß der wahre Wert des Individuums in dem so errechneten Konfidenzbereich liegt.

Analog dazu kann die Differenz zwischen zwei Subtestwerten zufallskritisch beurteilt werden (HUBER 1973, S.127 ff.). Hierzu müssen zuerst die Subtestreliabilitäten auf praktische Invarianz und Homogenität überprüft werden. Ist beides gegeben, so kann mit den üblichen x-Normierungen gearbeitet werden; hingegen ist bei heterogenen Reliabilitäten eine tau-Normierung der Subtestwerte und eine getrennte Überprüfung ihrer praktischen Invarianz erforderlich. Auf dieser Grundlage kann dann die Differenz der Subtestwerte unter dem Aspekt der Reliabilität zufallskritisch beurteilt werden; diese ist dann als realiter gegeben anzusehen,

wenn die beobachtete Differenz die bei definierter Wahrscheinlichkeit vorliegende kritische Distanz bzw. kritische Prüfgröße übersteigt. Ist dies der Fall, so kann in einem zweiten Schritt die Testwertdifferenz unter dem Aspekt der diagnostischen Valenz beurteilt werden, die eine Überprüfung der praktischen Invarianz des Korrelationskoeffizienten beider Subtests voraussetzt. Die Überprüfung der Signifikanz der Testwertdifferenz kann ebenfalls sowohl anhand einer kritischen Distanz als auch einer kritischen Prüfgröße vorgenommen werden und führt zu der Aussage über die Wahrscheinlichkeit, mit der eine Testwertdifferenz der beobachteten Größenordnung in der Eichpopulation anzutreffen ist. Erst wenn auch die so überprüfte diagnostische Valenz als gegeben angesehen werden kann, ist die Testbeurteilung diagnostisch verwertbar!

Über diese beiden genannten wichtigsten Anwendungsmöglichkeiten hinaus hat HUBER (1973) auch Verfahrensvorschriften für die folgenden eignungsdiagnostisch relevanten Fragestellungen ausgearbeitet:

- Analyse individueller Testprofile mit Hilfe statistischer Inferenzen hinsichtlich Niveau, Streuung und Verlaufsgestalt (S.151 ff.)
- Gezielte Linearvergleiche innerhalb eines Individualprofils (S.174 ff.)
- Vergleich zweier Testprofile (einschl. Referenzprofil)(S.201 ff.)
- Gezielte Vergleiche zweier Testprofile hinsichtlich bestimmter Subtestgruppen (S.218 ff.).

Aus dieser kurzen Skizzierung der praktischen Konsequenzen der PED dürfte deutlich geworden sein, daß bei einer zufallskritischen Absicherung von Testbeurteilungen die Zahl der diagnostisch relevanten und damit gutachtlich verwertbaren Testergebnisse in der eignungsdiagnostischen Praxis erheblich vermindert wird. Diese aufgrund der erheblich verschärften Kriterien einleuchtende Vermutung wird auch durch eigene Erfahrungen mit der PED gestützt.

2.1.3 Kritik der Psychometrischen Einzelfalldiagnostik

Für die Kritik und die Beurteilung der Brauchbarkeit der PED für eine arbeitsorientierte Eignungsdiagnostik ist die Tatsache entscheidend, daß dieses Modell insofern auf der KTT beruht,

- als die PED sich ausschließlich mit der Auswertung von Testbefunden unter dem Aspekt der Reliabilität befaßt, aber deren Zustandekommen, nämlich die Testkonstruktion unhinterfragt läßt; die Statistiken also, die in der PED zufalls-kritisch beurteilt werden, sind die gleichen wie in der konventionellen Testdiagnostik;
- als die PED, ohne daß dies im theoretischen Modell notwendigerweise impliziert ist, praktisch nur unter Inkaufnahme der Populationsabhängigkeit angewendet werden kann, da die Varianz der Meßfehler nicht am Probanden selbst, sondern

nur über die Normierungs- und Standardisierungsdaten, also interindividuell definierte Variablen hergeleitet werden kann. Dies gilt analog für die Beurteilung der diagnostischen Valenz.

Damit kann der konservative Charakter der KTT und der darauf beruhenden Testdiagnostik mit Hilfe der PED nicht aufgehoben werden. Aufgrund der Tatsache, daß die PED den durch die KTT definierten Rahmen nicht verläßt, nun aber zu schlußfolgern, daß die Kritik an der KTT (SCHARDT 1976) auch auf die PED in toto und mit all ihren Konsequenzen zuträfe, wäre jedoch voreilig. Hier sind vielmehr einige substntielle Relativierungen erforderlich:

- Die beiden theoretisch entscheidenden Einwände gegen die KTT beinhalten zum einen die Populationsabhängigkeit des Modells bzw. seiner Parameter und zum anderen die "Auspartialisierung" intrapsychischer Varianz als bloße Fehlervarianz. Da der wahre Wert in der PED als eine individuum- und testspezifische Konstante definiert wird, ist der letztenannte Einwand in Bezug auf die PED nicht aufrechtzuerhalten.
- Auch hinsichtlich der Populationsabhängigkeit stellt sich bei der PED das Problem in differenzierterer Form: Zwar ist von der theoretischen Grundlage her die zufallskritisch orientierte Einzelfalldiagnostik ein populationsunabhängiges Modell; indem dieses jedoch nur unter Zugrundelegung von interindividuell definierten Normierungs- und Standardisierungsdaten praktisch angewendet werden kann, wird die Populationsabhängigkeit durch die "Hintertür", wenn auch - im Vergleich zur KTT - in vermittelterer Form, wieder eingeführt.
- Schließlich werden im Unterschied zur konventionellen Testdiagnostik die Kriterien für die diagnostische Verwertbarkeit von Testbeurteilungen ganz erheblich verschärft, indem die Beurteilung auf methodisch ausweisbare Grundlagen gestellt wird. Dabei sollte nicht übersehen werden, daß der Verzicht auf eine zufallskritische Absicherung von Testergebnissen notwendigerweise zur Konsequenz hat, daß keinerlei empirisch abgesicherte Aussagen über deren diagnostische Verwertbarkeit gemacht werden können, der Testdiagnostiker also trotz des enormen psychometrischen Aufwands spätestens in seinem Gutachten wieder bei der alten idiographischen Methode angelangt ist.

Wenn auch die PED nicht als die Alternative zur konventionellen Testdiagnostik angesehen werden kann, so ist ihre Verwendung im Sinne der Durchsetzung arbeitsorientierter Interessen jedoch insofern möglich und - kurzfristig - sogar notwendig, als psychodiagnostische Beurteilungen oftmals entscheidende Bedeutung für den zukünftigen Lebensweg der davon Betroffenen haben. Notwendig ist die zufallskritische Absicherung von Testergebnissen deshalb, weil damit zumindest solche negativen Begutachtungen verhindert werden können, die auf nur scheinbar diagnostisch relevanten Beurteilungen beruhen, realiter jedoch aufgrund mangelnder Güte des Testverfahrens zustandegekommen (praktische Invarianz) bzw. im Vergleich zur Referenzpopulation keineswegs als "normdeviant" (diagnostische Valenz) anzusehen sind. Statistisch gesehen gilt es also, einen Fehler 1. Art nach Möglichkeit auszuschließen (deshalb sollte das Signifikanzniveau bei der zufallskritischen Absicherung möglichst hoch angesetzt werden). Damit würde allerdings in gleicher Weise auch eine Bestauslese erschwert, da diese sich ja ebenfalls

auf signifikante - allerdings positive - Abweichungen von der Referenzpopulation stützt. Dies ist jedoch im Sinne einer arbeitsorientierten Eignungsdiagnostik keineswegs unerwünscht, da davon ausgegangen werden muß, daß eine solche positive Beurteilung in erheblichem Maße auf Anpassungsergebnissen und -potentialen beruht, und zwar im Sinne einer Anpassung an die herrschenden Normen - nämlich die der Herrschenden (SCHARDT 1976).

Damit ist gleichzeitig auch eine in der Anwendung der PED liegenden Gefahr angesprochen: Der apologetische, durch großen statistischen Aufwand suggerierte Schein der Objektivität eignungsdiagnostischer Begutachtungen wird dadurch eher noch verstärkt! Hingegen wird der interessenverhaftete, weil konservative Charakter der Testbeurteilungen selbst im Rahmen der PED erst gar nicht thematisiert.

Zusammenfassend bleibt festzuhalten, daß auch die zufallskritische Absicherung von Testresultaten mit Hilfe der Methoden der PED als allgemein verbindliches Kriterium eignungsdiagnostischer Praxis eine arbeitsorientierte Eignungsdiagnostik nicht der Aufgabe entheben kann, Alternativen zur KTT als Modell der empirisch-quantitativen Erfassung individueller Fähigkeiten zu entwickeln. Andererseits kann mangels praktikabler, arbeitsorientierter Alternativen kurzfristig auf eine solche zufallskritisch orientierte Einzelfalldiagnostik nicht verzichtet werden, weil in dem Maße, in dem testdiagnostisch begründete, der Prämisse eines konservativen Menschenbildes unterliegende Personalauslesemaßnahmen eingeschränkt werden, der Verkauf der Arbeitskraft sich für die Betroffenen wieder als das herausstellt, was er ist: eine letztlich in gesellschaftlichen Kräfteverhältnissen begründete, von den Marktmechanismen von Angebot und Nachfrage regulierte Tauschbeziehung.

2.2 Lehrzielorientierte Tests

2.2.1 Methodologische Aspekte

Mit dem Konzept der Lehrzielorientierten Tests (LOT), wie es von KLAUER, FRICKE, HERBIG, RUPPRECHT & SCHOTT (1972; im folgenden zit.n.d.2.Aufl.1974) vorgelegt wurde, wird der Anspruch verbunden, eine zufallskritisch abgesicherte Entscheidung über individuelle Testresultate zu ermöglichen. (Eine ebenso kurze wie informative Darstellung dieses Konzepts findet sich bei KLAUER 1975.) In dieser allgemeinen Zielsetzung stimmen die LOTs noch mit der PED überein. Der Unterschied zwischen beiden Ansätzen liegt darin, daß die PED dieses Ziel über die Formulierung methodologischer Standards für die Auswertung von Testresultaten

auf der Basis des Reliabilitätskonzepts zu realisieren versucht, während im Rahmen der LOTs die zufallskritische Auswertung erst durch eine alternative Testtheorie und - daraus resultierend - eine alternative Testkonstruktion ermöglicht wird.

Da das Konzept der LOTs im Rahmen der Pädagogischen Psychologie entwickelt wurde, insofern also von der Intention her kein Modell der eignungsdiagnostischen Begutachtung darstellt und m.W. bislang auch noch nicht der Versuch einer solchen Übertragung unternommen worden ist, konzentrieren sich die folgenden Überlegungen auf eine Überprüfung der Anwendbarkeit der LOTs in der eignungsdiagnostischen Praxis sowie die damit verbundenen theoretischen und praktischen Probleme.

Die Logik der konventionellen psychometrischen Eignungsdiagnostik reduziert sich letztlich darauf, von der Ausprägung eines Persönlichkeitsmerkmals, das durch das jeweilige Testverfahren operationalisiert und quantifiziert wird, auf die Fähigkeit des betreffenden Individuums zu schließen, ein bestimmtes Verhalten (Tätigkeit) zu zeigen (auszuüben). Dieser Schluß ist aber nur dann haltbar, wenn davon ausgegangen werden kann, daß das betreffende Persönlichkeitsmerkmal realiter für die Ausübung der in Frage stehenden (Arbeits-)Tätigkeit relevant ist. Im Rahmen der KTT wird dieser Nachweis über die Validitätsstudien zu erbringen versucht - ein Bemühen allerdings, das so, wie es sich heute praktisch darstellt, Anlaß zu massiver Kritik gibt (vgl. SCHARDT 1976).

Insoweit stimmt dieser Ansatz der LOTs noch mit diesem Rationale überein: "...das Lehrziel (muß) als Persönlichkeitsmerkmal definiert werden, das sich in bestimmtem Verhalten äußert. Das Verhalten ist dann nicht das Lehrziel, sondern zeigt die Erreichung des Lehrziels nur an, der Test ist nicht das Lehrziel, er repräsentiert es nur, und es ist denkbar, daß das Lehrziel von zwei Tests unterschiedlich gut repräsentiert wird" (KLAUER et al. 1974, S.18). M.a.W., mit Hilfe eines LOT kann überprüft werden, ob ein bestimmtes Lehrziel erreicht wurde, oder nicht (eine bestimmte Tätigkeit ausgeführt werden kann, oder nicht), wobei das Lehrziel (Fähigkeit im o.g. Sinne) durch die Lösung derjenigen Aufgaben (Ausführung derjenigen Tätigkeitselemente) definiert ist, zu deren Lösung der Unterrichtete befähigt werden soll (durch deren Ausführung die jeweilige Stelle definiert ist). LOTs beruhen demzufolge auf dem Binomialmodell und liefern "lediglich" die qualitativ-dichotome Information "bestanden-nicht bestanden" bzw. "geeignet-nicht geeignet". Das wiederum bedeutet, daß - im Gegensatz zur KTT - kein relatives, sondern ein absolutes Urteil abgegeben wird oder, anders formuliert, daß die Referenzpopulation als (Vergleichs-)Norm abgelöst und durch das Lehrziel als Kriterium ersetzt wird. Hierin liegt der methodologische Kern des Modells der LOTs, nämlich in der

Ersetzung der Realnormen als Vergleichsmaßstab durch ausschließlich sachimmanent zu definierende Idealnormen (Kriterien).

Abgesehen davon, daß das Binomialmodell für die lehrzielorientierte Messung nicht konstitutiv ist, d.h. auch differenziertere, multikategoriale oder ordinale Beurteilungen (im Sinne einer GUTTMAN-Skala) prinzipiell möglich scheinen (KLAUER et al. 1974; STRITTMATTER 1973), ist im vorliegenden Zusammenhang die Lehrzieldefinition und die Bestimmung der Kontentvalidität von zentraler Bedeutung, da sich hierin der theoretische und methodologische Kern dieses Testmodells konkretisiert: "Die Lehrzieldefinition ist...ein entscheidender Punkt für die sogenannte Kontentvalidität eines lehrzielorientierten Tests. Man meint damit die Eigenschaft eines Testes, daß er prüft, ob genau dieses fragliche Lehrziel statt irgendeines anderen erreicht worden ist. Um das zu gewährleisten, ist das Lehrziel in Form einer Aufgabenmenge zu definieren, nämlich in Form der Aufgabenmenge, zu deren Lösung der Schüler befähigt werden soll" (KLAUER 1975, S.37).

Die Testaufgaben selbst sind "...durch mindestens zwei Komponenten bestimmt, die Inhaltskomponente und die Verhaltenskomponente (TYLER), mitunter auch noch durch eine Produktkomponente (GUILFORD). Demnach ist also bei der Definition der Aufgabenmenge immer mindestens der Inhalt anzugeben, der dem Prüfling angeboten wird, sowie das daran zu vollziehende Verhalten" (KLAUER 1975, S.37 f.). Die Inhaltskomponente bezieht sich auf das, was vermittelt werden, und die Verhaltenskomponente auf das Verhalten, das der Lernende ausführen soll, wenn das Lehrziel erreicht worden ist; hierbei kann zwischen kognitiven, affektiven und psychomotorischen, hierarchisch geordneten Verhaltensaspekten unterschieden werden (BLOOMsche Taxonomie). Eine quantitative Beschreibung der Lehrziele ist mit Hilfe einer Lehrzielmatrix möglich, wobei die Gesamtbandbreite angibt, wieviele verschiedene Aufgabenklassen (Teilziele) das Lehrziel bestimmen. Die Inhaltsbandbreite enthält die Anzahl der Verhaltensweisen, die von dem jeweiligen Lehrziel betroffen sind bzw. gefordert werden. Mit der Zielakzentuierung ist eine quantitative Angabe darüber möglich, ob von dem Lehrziel eher der Inhalts- oder der Verhaltensaspekt gefordert wird (KLAUER et al. 1974; Beitrag von SCHOTT).

Die Definition der Menge aller Aufgaben bzw. Tätigkeitselemente kann entweder durch eine vollständige Aufzählung aller Elemente der Menge oder durch die Angabe der "mengenstiftenden Merkmale" vorgenommen werden. Im letzteren Fall müssen dann die Testaufgaben erst noch generiert werden. Jede der auf dieser Grundlage konzipierten Aufgaben ist auf ihre Kontentvalidität hin zu überprüfen. Diese ist dann gegeben, wenn die Aufgabe die Definitionsvorschrift für das Lehrziel erfüllt.

Da das Lehrziel das einzige Kriterium von LOTs darstellt, sind auch die Gütekriterien der KTT nicht mehr anwendbar (bei 100% richtigen Lösungen wird die Varianz und somit auch die Reliabilität und Validität Null; Items von LOTs weisen nur geringe Trennschärfe auf, da ihre Schwierigkeit aufgrund der Kon-

struktionsprinzipien nahe Eins liegt). Es wurde daher vorgeschlagen, einen Übereinstimmungskoeffizienten zu berechnen, der auch in den Extrembereichen noch interpretierbare Werte liefert und sich zur Quantifizierung aller Gütekriterien eignet (KLAUER et al. 1974; Beitrag von HERBIG).

Eine informative Gegenüberstellung der schon genannten und weiterer Unterschiede zwischen herkömmlicher und lehrzielorientierter Messung gibt KLAUER in Form einer Tabelle, die hier nur wiedergegeben, aber nicht mehr weiter diskutiert werden soll:

herkömmliche Tests

normorientiert
relativ
quantitativ
dimensional
komparatistisch
populationsabhängig
summativ
psychologisch

lehrzielorientierte Tests

kriteriumsorientiert
absolut
qualitativ
dichotomisch
dezisionistisch
populationsunabhängig
formativ
pädagogisch

(nach Tabelle 1 in KLAUER et al. 1974, S.21)

2.2.2 Praktische Konsequenzen

Die Konstruktion eines LOT läuft nach KLAUER in folgenden Schritten ab (1975, S.38):

- (1) Definition des Lehrziels in qualitativer Hinsicht, d.h. Bestimmung derjenigen Aufgabenklassen, die der Lernende hinterher lösen können soll;
- (2) Definition des Lehrziels in quantitativer Hinsicht, d.h. Festlegung des Grades der Beherrschung, die der Lernende erreichen soll;
- (3) Wahl der geeigneten Aufgabenform;
- (4) Bestimmung der Mindestzahl notwendiger Testaufgaben aufgrund von Schritt 2 und 3;
- (5) Formulierung von Testaufgaben aufgrund von Schritt 1 und 3;
- (6) Überprüfung der Aufgaben der Kontenvalidität, d.h. auf Zugehörigkeit zu der in Schritt 1 definierten Aufgabenmenge;
- (7) Überprüfung der Testaufgaben auf Repräsentativität, d.h. darauf, ob auch alle wichtigen Teilmengen der in Schritt 1 definierten Aufgabenmenge in der Teststichprobe angemessen wertreten sind;
- (8) Bestimmung des kritischen Punktwertes aufgrund von Schritt 2 und der im Test gegebenen Aufgaben.

Die Frage der Anwendbarkeit dieses Testmodells in der eignungsdiagnostischen Praxis konzentriert sich auf die Definition des Lehrziels, nämlich die Bestimmung der für die jeweilige Stelle erforderlichen Tätigkeit(en). Genau dies ist

Gegenstand der Arbeitsanalyse, auf deren desolaten Zustand und die daraus resultierenden Probleme für die Eignungsdiagnostik schon an anderer Stelle hingewiesen wurde (SCHARDT 1976). Eine naheliegende, pragmatische Lösung dieses Problems bestünde darin, sich auf eine Auflistung der jeweils erforderlichen Arbeitsvollzüge zu beschränken.

Als ein besonders einfaches, weil nur aus einer Aufgabe bestehendes Beispiel kann die Maschinenschreibprobe (z.B. für Stenotypistinnen) angeführt werden. Aber schon allein an der Tatsache, daß eine derartige, scheinbar einfache Arbeitstätigkeit sowohl einer weiteren taxonomischen Ausdifferenzierung auf der Grundlage der o.g. Komponenten als auch einer psychologischen Arbeitsanalyse zugänglich ist, wird deutlich, daß eine solche "pragmatische" Lösung (im Sinne der inhaltlichen Validität nach der Definition von LIENERT 1967, S.17) auf keinen Fall befriedigen kann - und zwar deswegen, weil das Lehrziel dann nicht mehr als das Persönlichkeitsmerkmal definiert wird, das für die psychische Regulation der Tätigkeit (im Sinne von HACKER) verantwortlich ist. Vielmehr ist die Tätigkeit selbst jetzt Gegenstand des Tests und die Frage nach der Validität somit gegenstandslos (es handelt sich also streng genommen nicht mehr um einen Test im psychometrischen Sinne - vgl. die Definition bei LIENERT 1967, S.7 - sondern eher um eine Arbeitsprobe). Auf diese Weise aber das "Validitätsproblem hinauszudefinieren" ist eine "unsinnige Konsequenz" (KLAUER). Stattdessen muß als Lehrziel das definiert werden, "...was den individuellen Verhaltensäußerungen von heute, morgen und übermorgen invariant gemeinsam ist - nämlich eine bestimmte Verhaltenseigenschaft, ein bestimmtes Verhaltensmerkmal oder ein Gefüge, eine Struktur von Verhaltenseigenschaften. Das Lehrziel ist also eine Abstraktion vom vorfindlichen Verhalten, so wie alle Eigenschaften oder Merkmale Abstraktionen sind" (KLAUER et al. 1974, S.32; Hervorhebung von mir, L.P.S.).

Positiv gewendet resultiert daraus die Forderung nach der Ableitung des Lehrziels und damit der Testaufgaben qua Tätigkeitselemente auf der Grundlage einer Allgemeinen Lohnarbeitspsychologie, wie sie von VOLPERT (in GROSKURTH & VOLPERT 1975) in groben Umrissen auf der Basis der Handlungstheorie von HACKER (1973) und der Persönlichkeitstheorie von SEVE (1973) zu konzipieren versucht wurde (vgl. SCHARDT 1976). M.a.W., in der Eignungsdiagnostik muß das Lehrziel anhand einer psychologischen Arbeitsanalyse als die psychischen Invarianten definiert werden, über die die Regulation der je konkreten Tätigkeiten im Sinne von Handlungseinheiten erfolgt. Es ist zu vermuten, daß sich dabei die Taxonomie und Evaluation der Testaufgaben (Kontentvalidität) anhand der von SCHOTT im ein-

zelen beschriebenen TYLERSchen Lehrzielmatrix (KLAUER et al. 1974) als unzureichend erweist. Festzuhalten bleibt, daß zumindest auf längere Sicht die Ermittlung der Kontentvalidität anhand arbeitsanalytisch ausgewiesener Kriterien unabdingbar ist. Diese Forderung läuft letztlich darauf hinaus, das Kriterium der Kontentvalidität durch das der Konstruktvalidität zu ersetzen (vgl. dazu die expliziten Hinweise von KLAUER et al. 1974, S.18 u. 36f).

Problematisch erscheint auch das den LOTs zugrunde liegende binomiale Modell, da dieses u.a. voraussetzt, daß den Aufgaben gleiche Lösungswahrscheinlichkeit zukommt und daß es sich bei den Lösungen verschiedener Aufgaben um stochastisch unabhängige Ereignisse handelt. Während die erstgenannte Voraussetzung praktisch in den Griff zu bekommen ist - zumindest statistisch -, bleibt die zweite theoretisch genauso unbegründet, wie die Annahme der lokalen stochastischen Unabhängigkeit im RASCH-Modell (vgl. Abschnitt 2.3 in diesem Beitrag; KLAUER et al. 1974; Beitrag von FRICKE).

Ein weiteres bedeutsames Problem liegt in der Festlegung des Beherrschungsgrades (der nicht mit dem - noch Zufallseinflüsse reflektierenden - kritischen Punktwert zu verwechseln ist). "Man kann nicht gut eine 100%ige Sicherheit verlangen, weil kein Mensch und keine Maschine fehlerlos arbeiten. Im allgemeinen wählt man das Niveau von 95% oder 90%" (KLAUER 1975, S.38). Eine solche formal-pragmatische Lösung ist im vorliegenden Fall jedoch unbrauchbar. Erforderlich ist vielmehr die Festlegung eines "Mindest"-Beherrschungsgrades nach sachlichen Gesichtspunkten (z.B. Unfallsicherheit bei Maschinen- und Automatenkontrolle; Sicherstellung des gesamten, arbeitsteilig organisierten Arbeitsablaufs). Dieser könnte dann u.U. erheblich unter den von KLAUER angegebenen Quoten liegen. Denkbar wäre auch, neben einer Mindestgrenze einen "erwünschten" bzw. "vollständigen" Beherrschungsgrad festzulegen. Aufgrund dessen könnte dann für alle die Bewerber, deren Testresultate zwischen den aus diesen Festlegungen resultierenden kritischen Punktwerten liegen, die Notwendigkeit von Qualifikationsmaßnahmen zur Aktivierung des vorhandenen Fähigkeitspotentials empirisch begründet werden. Hierauf wird in der Kritik der LOTs noch zurückzukommen sein.

Unter dem Qualifikationsaspekt ist nach KLAUER "...der Praktiker...jedenfalls gut beraten, nicht zuviel Inhaltskategorien und Verhaltenskategorien in eine Aufgabenmenge zu packen, sondern besser zwei oder drei Lehrziele mit ganz speziellen Leistungen zu definieren" (1975, S.38). Damit werden nicht nur eine differenziertere Diagnose im allgemeinen und gezieltere Hinweise auf ggf. erforderliche Qualifikationsmaßnahmen im besonderen ermöglicht, sondern es wird auch die Gefahr eingeschränkt, die Vorteile dieses Testmodells durch eine Überfrachtung der Lehrziele aufzuheben oder gar in ihr Gegenteil zu verkehren: Eine Vernachlässigung des Validitätskriteriums könnte dazu führen, daß genau wie in der konventionellen Testdiagnostik den Probanden eine Unzahl von Testaufgaben vorgelegt wird, deren Relevanz für die konkrete Tätigkeit letztlich niemand mehr - und schon gar nicht die Betroffenen - beurteilen kann.

2.2.3 Kritik der Lehrzielorientierten Tests

Dem konservativen Charakter der KTT wurde als Antithese die dynamische Auffassung einer lebenslangen beruflichen Sozialisation im Sinne der Höherentwicklung individueller Fähigkeiten entgegengestellt. Erweitert man diese Antithese um die Forderung, daß eine solche Entwicklung nicht Zufälligkeiten überlassen, d.h. ungezielt ablaufen sollte, so scheint der Intention nach mit dem Modell der LOTs eine Alternative zur KTT vorgelegt worden zu sein, die für die betriebspsychologische Praxis der Personalbeurteilung im allgemeinen (darauf kann hier nicht näher eingegangen werden) und der Eignungsdiagnostik im besonderen als potentiell relevant angesehen werden muß. Daß auch die LOTs zumindest auf absehbare Zeit nicht die Alternative zur KTT darstellen können und ihre Relevanz insofern als "potentiell" relativiert werden muß, ist in den folgenden Problemen begründet.

Das vordergründigste, nichtsdestotrotz praktisch bedeutsamste Problem besteht darin, daß bislang noch keine, für eignungsdiagnostische Fragestellungen brauchbare LOTs entwickelt wurden. Deren Konstruktion wäre jedoch kurzfristig nur auf der Basis von Arbeitsanalysen möglich, die schon allein unter immanent-methodischen Gesichtspunkten höchst fragwürdig sind (vgl. HOYOS 1974; FRIELING 1975) - von der Berücksichtigung arbeitsorientierter Interessen erst gar nicht zu reden. Das hätte jedoch zur Konsequenz, daß der status quo der heutigen Arbeitsbedingungen bzw. -anforderungen, deren Festschreibung der konventionellen Eignungsdiagnostik als zentraler Kritikpunkt angelastet wurde (SCHARDT 1976), in der gleichen Weise wie bisher, nunmehr aber viel direkter - weil nicht über quantitative Gütekriterien vermittelt und verschleiert - in die Eignungsdiagnostik eingehen würde. Das sollte jedoch nicht zu dem - wie noch zu zeigen sein wird - voreiligen Schluß führen, eine solche pragmatische Lösung des Problems der Lehrzieldefinition als völlig unakzeptabel auszuschließen (schon allein deshalb nicht, weil dies de facto auf eine Prolongierung der konventionellen Eignungsdiagnostik hinausläufe).

Mit diesen Überlegungen ist im Grunde genommen schon ein weiteres Problem angesprochen, dem unter theoretischen Gesichtspunkten letztlich entscheidende Bedeutung zukommt und dem Validitätsproblem der KTT analog ist: Die Kontentvalidität bestimmt sich nämlich inhaltlich durch die Definition des Lehrziels in Form einer Idealnorm. Wie immer das Lehrziel auch bestimmt werden mag - ob intuitiv-pragmatisch oder arbeitsanalytisch -, entscheidend bleibt die Frage, was dieses Lehrziel beinhaltet. Da hinter jedem Lehrziel - auch und gerade im hierverwendeten eignungsdiagnostischen Sinne - ein bestimmtes

Menschenbild steht, nämlich Auffassungen darüber, was ein Mensch können und wie er sich verhalten sollte, steht hinter jeder Lehrzieldefinition letztlich eine politische Entscheidung. Genau dies ist der Grund dafür, daß oben nicht die Ableitung des Lehrziels aus einer Arbeitsanalyse schlechthin gefordert wurde, sondern deren Begründung durch eine Allgemeine Lohnarbeitspsychologie.

Andererseits steht hinter jeder Lehrzieldefinition notwendigerweise die Absicht, einen Entwicklungsprozeß zu initiieren. Im Gegensatz zur KTT unterliegt damit den LOTs ein prinzipiell progressives Modell. Von dem Rationale der diagnostischen Fragestellung her stellt das Modell der LOTs somit eine echte Alternative zur KTT dar. Dabei sollte jedoch nicht übersehen werden, daß es nicht ausgeschlossen werden kann, in der Praxis auch objektiv reaktionäre Lehrziele zu definieren (z.B. Bereitschaft zur Akzeptanz fremdbestimmter Anordnungen). Aber auch diese, durchaus naheliegende Gefahr kann nicht - wie ebenfalls noch zu zeigen sein wird - als ein Argument für die völlige Ablehnung von LOTs herangezogen werden.

Im Zusammenhang mit dem Validitätskriterium entstehen jedoch noch eine Reihe weiterer methodischer Probleme, die hier nur angedeutet werden können. Abgesehen davon, daß der mit der Kontentvalidität verbundene theoretische Anspruch nur dann eingelöst werden kann, wenn entsprechende empirische Methoden zur ihrer Ermittlung zur Verfügung stehen (vgl. LIENERT 1967; KLAUER et al. 1974, Beitrag von SCHOTT), kann nicht der methodische Aspekt übersehen werden, daß Testresultate - unabhängig davon, nach welchem Testmodell sie zustande gekommen sind - in der Eignungsdiagnostik immer den Charakter von Prädiktoren haben. Aufgrund praktischer - wenn auch theoretisch nur schwer begründbarer - Erfahrungen ist allerdings bekannt, daß die Prognose eines Kriteriums qua Berufserfolg durch die Einbeziehung von Suppressorvariablen erheblich verbessert werden kann. Da solche Variablen gerade dadurch definiert sind, daß sie mit dem Kriterium niedrig, mit anderen Prädiktoren jedoch hoch korrelieren (negatives Vorzeichen des Beta-Gewichtes in der multiplen Regressionsgleichung), kommen diese unter dem Kriterium der Kontentvalidität für einen LOT nicht in Frage. Ein analoges Problem besteht hinsichtlich der Einbeziehung von Moderatorvariablen, deren Kriteriumskorrelationen definitionsgemäß in unterschiedlichen Gruppen unterschiedlich hoch sind (differentielle Vorhersagbarkeit). Sowohl unter dem Aspekt der Verwendung mehrerer LOTs als auch dem der Bestimmung der optimalen Aufgabenzahl für einen einzelnen LOT stellt sich das Problem der inkrementellen Validität, d.h. dem durch die zwischen den verschiedenen Prädiktoren vorhandenen Wechselbeziehungen vermittelten Beitrag, den ein spezifischer Test zur Erhöhung der Validität der gesamten Testbatterie bzw. aller Prädiktoren leisten kann. Da diese Probleme aber im vorliegenden Zusammenhang nicht zentral sind und ohnehin hierfür noch keine Lösungsvorschläge unterbreitet werden können, sei an dieser Stelle lediglich auf die einschlägige Literatur verwiesen (u.a. FISCHER 1974, S.75f.).

Abschließend soll nocheinmal das schon erwähnte Problem der Festlegung des Beherrschungsgrades kurz aufgegriffen werden. Es wäre denkbar, den Beherrschungsgrad oberhalb der geforderten Mindestgrenze im Hinblick auf arbeitsmarkt- bzw. beschäftigungspolitische Parameter variabel zu handhaben: Je

größer die Bewerberzahl im Verhältnis zu der Zahl der offenen Stellen ist, umso höher könnte der Beherrschungsgrad festgelegt werden! Damit würde jedoch der kriteriumsorientierte de facto durch einen an sachfremden Normen orientierten Ansatz ersetzt und die derzeit gängige eignungsdiagnostische Praxis durch die Hintertür wieder eingeführt werden: Der Vorteil eines LOT besteht eben genau darin, daß ausschließlich aufgrund immanenter Kriterien (Kontentvalidität) nur die dichotome Information "geeignet-ungeeignet", also ein absolutes, kriterienbezogenes Urteil geliefert wird. Das Binomialmodell, das nur durch einen solchen kriterienbezogenen Ansatz theoretisch begründet werden kann, würde somit ad absurdum geführt und der LOT zu einem "formellen Test mit lehrzielorientierten Aufgaben" (KLAUER et al. 1974, S.16).

Der entscheidende Vorteil der LOTs im Hinblick auf eine - ohnehin nur mittelfristig mögliche - Anwendung liegt in der relativen Einfachheit und damit der Transparenz des zugrundeliegenden Modells. Nur dieser Aspekt liefert die Legitimation dafür, LOTs trotz der hier skizzierten und weiterer Probleme und Gefahren als mittelfristige Alternative für die Konzipierung einer arbeitsorientierten Eignungsdiagnostik ins Auge zu fassen. Diese Transparenz der Konstruktionsprinzipien und der in sie eingehenden Voraussetzungen und Annahmen, insbesondere hinsichtlich der Definition des Lehrziels, läßt nicht nur eine Verbreiterung des Anwenderkreises auch auf Nicht-Psychologen zu. Vielmehr werden damit erstmals von testtheoretischer Seite aus die Voraussetzungen dafür geschaffen, den Interessen der Betroffenen auch im eignungsdiagnostischen Bereich im Rahmen der von Interessengegensätzen und Konflikten bestimmten industriellen Beziehungen wirksam zur Durchsetzung zu verhelfen: Aufgrund des immanenten Zwangs zur Zielzentriertheit (Kontentvalidität), unter der jede einzelne Testaufgabe gestellt wird, wird der gesamte diagnostische Prozeß in allen seinen Phasen (vgl. SCHARDT 1976) für die Betroffenen und ihre Interessenvertretungsorgane sehr viel eher durchschaubar, als dies derzeit noch bei den mit Hilfe der quantitativen Gütekriterien der KTT ihren konservativen Charakter verschleiern und Objektivität suggerierenden Testverfahren der Fall ist. Erst in dem Maße, in dem der eignungsdiagnostische Prozeß durchschaubar und damit auch kontrollierbar wird, wird er auch konfliktfähig.

2.3 Probabilistische Testmodelle

Die Probabilistischen Testmodelle (PTM), wie sie vor allem von RASCH (1960, 1961, 1966) konzipiert wurden, sollen in diesem Beitrag nur kurz abgehandelt werden, da dieser Ansatz - um das wichtigste Ergebnis der folgenden Überlegungen schon vorwegzunehmen - nicht dazu geeignet erscheint, in absehbarer Zeit zur Entwicklung einer arbeitsorientierten Eignungsdiagnostik beizutragen.

Vom theoretischen Ansatz her stimmen die PTM weitgehend mit den LOTs überein: Das explizite Ziel dieses von seinen Vertretern auch als individualistisch bezeichneten Ansatzes leitet sich aus der Kritik der Populationsabhängigkeit der KTT ab und beinhaltet den Versuch, Individuen als Individuen miteinander zu vergleichen, ohne dabei jedoch auf eine Referenzpopulation Bezug zu nehmen. STENE formuliert das folgendermaßen: "Die Frage etwa, ob sich ein Kind während eines oder mehrerer Jahre verbessert hat, kann unmöglich mit Hilfe verschiedener Tests beantwortet werden, die an Populationen unterschiedlich alter Kinder geeicht wurden" (1968, S.229). Aus diesem Grunde versuchte RASCH, "...to develop probabilistic models in the application of which the population can be ignored. It was a discovery of some mathematical significance that such models could be constructed, and it seemed remarkable that data collected in routine psychological testing could be fairly well represented by such models" (1966, S.89).

Indem der Kennwert eines bestimmten Individuums in einem bestimmten Test als Einheitsgröße in Form des Parameters Fähigkeit (ability) ausgedrückt wird, werden nicht mehr relative, auf Referenzpopulationen rekurrierende Urteile im Sinne der KTT, sondern auf die intraindividuelle Leistungsvariabilität bezogene Urteile ermöglicht. Solche intraindividuell relativierten Urteile werden dann zu absoluten und ein nach dem RASCH-Modell konstruierter Test zu einem LOT, "...wenn die vom Lehrziel geforderte Leistung als Bezugsgröße gewählt würde" (KLAUER et al. 1974, S.25; vgl. dazu auch den Beitrag von FRICKE).

Der methodologische Kern der PTM besteht darin, daß die beiden grundlegenden Parameter Fähigkeit - als Charakteristikum dafür, eine bestimmte Art von Aufgaben richtig zu lösen - und Leichtigkeit (easiness) - als Charakteristikum der Aufgabe analog dem Schwierigkeitsindex der KTT - unabhängig voneinander geschätzt werden. Dadurch wird es ermöglicht, statistische Inferenzschlüsse nur aufgrund dieser beiden Parameter zu treffen, d.h. ein Rekurs auf Populationsparameter ist nicht erforderlich. Dies bezeichnet RASCH mit spezifischer Objektivität (RASCH 1966, STENE 1968).

Die bedingte Wahrscheinlichkeit dafür, daß ein bestimmtes Item von einer bestimmten Person richtig beantwortet wird, läßt sich aufgrund dieser beiden Parameter berechnen. Anhand dieser bedingten Wahrscheinlichkeit (Meßmodell) können zwei Personen hinsichtlich ihrer Fähigkeit miteinander verglichen werden, Aufgaben einer bestimmten Art richtig zu lösen. Je "fähiger" eine Person also zur Lösung einer bestimmten Aufgabe ist, umso größer ist die bedingte Wahrscheinlichkeit für eine richtige Lösung; ist eine Aufgabe schwierig - die Leichtigkeit also klein - und verfügt die Person über eine mittlere Fähigkeit, nähert sich die Wahrscheinlichkeit für eine richtige Lösung dem Wert Null, die (Gegen-)Wahrscheinlichkeit für eine falsche Lösung

dem Wert Eins. Dabei werden die Antworten einer Person auf zwei Items als stochastisch unabhängig angesehen (lokale stochastische Unabhängigkeit). Das steht nicht im Widerspruch zu der Möglichkeit, daß zwei Items in einer Stichprobe von Item-Beantwortungen korreliert sein können (was meist auch der Fall ist, insbesondere, wenn es sich um Aufgaben ähnlicher Leichtigkeit handelt) (STENE 1968, S.231ff.). Zur Kennzeichnung einer Person wird es als ausreichend angesehen, die Anzahl der richtigen Lösungen heranzuziehen, d.h. gleiche Anzahl richtiger Lösungen impliziert auch die gleiche Fähigkeit (das Pattern der Aufgabenlösungen liefert keine zusätzlichen Informationen). Diese Annahme ist jedoch theoretisch unbeweisbar bzw. unabgeleitet.

Als Hauptvorteil des stochastischen Testmodells werden die Populationsunabhängigkeit, die Abbildbarkeit beider Parameter auf Verhältnisskalenniveau, die Eindimensionalität sowie das Charakteristikum, daß nur Wahrscheinlichkeits- und keine deterministischen Aussagen getroffen werden, angesehen. Hinzuzufügen wäre, daß der Anwendungsbereich der probabilistischen Modelle keineswegs auf die Testtheorie beschränkt ist. So unternahm FISCHER (1973) den Versuch, die Wirkungskontrolle von Massenkommunikation mit Hilfe des stochastischen Ansatzes zu verbessern (vgl. auch zu analogen Versuchen STENE 1968, S.267).

Ohne weiter auf die mathematisch-statistischen und methodologischen Grundlagen einzugehen, können schon an dieser Stelle folgende Kritikpunkte formuliert werden:

- (1) Die für die diagnostische Beurteilung zentrale Annahme, nach der die gleiche Anzahl richtiger Lösungen auch die gleiche Fähigkeit impliziert, ist theoretisch unabgeleitet bzw. unbeweisbar - ein Einwand, der auch den LOTs gegenüber zu erheben wäre.
- (2) PTMs versagen dann, wenn alle Items bzw. kein Item richtig gelöst wurde bzw. bei Items, die von allen oder von keinem Probanden richtig gelöst wurden, weil in diesen Fällen die bedingten Wahrscheinlichkeiten jeweils Null bzw. Eins sind.
- (3) PTMs sind deshalb kaum praktikabel - und darauf weist sogar STENE (1968) hin - weil die Schätzung der Modellparameter bzw. die Methoden zur Kontrolle der Voraussetzungen des Modells derart kompliziert und aufwendig sind, daß sie selbst mit Hilfe von EDV-Anlagen kaum gehandhabt werden können.
- (4) HUBER weist auf die aus der Populationsunabhängigkeit resultierende Schwierigkeit bei der Beurteilung der diagnostischen Valenz hin, da das Konzept der logischen Validität, auf das RASCH implizit rekurriert, nur für psychologisch triviale Tests akzeptiert werden könne (1973, S.248).
- (5) Der entscheidende Einwand gegen eine Verwendung von PTMs im Rahmen einer arbeitsorientierten Eignungsdiagnostik gründet sich auf den von GUTJAHN vorgelegten empirischen Modellvergleich zwischen der RASCH- und der GAUSS-Skalierung (1971, S.221 ff.). Der Autor konnte nachweisen, daß auch RASCH implizit auf Verteilungsannahmen rekurrieren muß, es sich also keineswegs um eine stichprobenfreie Testeichung und Personmessung handelt, wie WRIGHT (1967) seinen Artikel programmatisch überschrieben hat. Genauer gesagt: "Das RASCH-Modell setzt eine der Normalverteilung sehr ähnliche Verteilungsform voraus,

so daß in praxi die Ergebnisse beider Skalierungsmethoden schwer zu unterscheiden sind" (GUTJAHR 1971, S.255). Damit bleibt die Intention einer personfreien Testeichung und testfreien Personmessung und damit auch das Attribut moderne Testtheorie bloße Proklamation.

Zusammenfassend ist festzustellen, daß man aufgrund der genannten methodischen und praktischen Schwierigkeiten, der methodologischen Einwände sowie der Tatsache, daß nach rund 15jähriger Forschung n.m.W. noch kein einziges praktisch brauchbares Testverfahren auf dieser Grundlage entwickelt worden ist, nicht um den Schluß umhin kann, daß eine Neuorientierung der Eignungsdiagnostik mit Hilfe der PTMs auf absehbare Zeit kaum zu bewerkstelligen sein dürfte.

3. Forschungsaufgaben

Die Analyse der konventionellen Eignungsdiagnostik anhand ihrer methodologischen Grundlagen, des Modells der KTT, und der Nachweis ihres kapitalorientierten, konservativen Charakters (SCHARDT 1976) führte zu der eingangs formulierten Forderung nach der Konzipierung einer arbeitsorientierten Eignungsdiagnostik. Die damit aufgeworfene Frage nach der Verwertung arbeitspsychologischer Erkenntnisse im Interesse der Betroffenen, nämlich der abhängig Beschäftigten, setzt unter dem inhaltlichen Aspekt voraus, daß solche Methodologien (Testtheorien) und Methoden vorliegen, die von ihrem wissenschaftstheoretischen Charakter her eine solche interessengebundene Verwertung ermöglichen. Ziel der vorstehenden Ausführungen war der Versuch, eine erste Antwort auf diese Frage zu geben. Dessen zentrales Ergebnis läßt sich folgendermaßen zusammenfassen:

Die Eignungsbeurteilung, als psychometrische Ermittlung individueller Merkmalsausprägungen im differentialdiagnostischen Sinne, beruht derzeit noch ausschließlich auf dem methodischen Instrumentarium der KTT. Diese in ihren einzelnen Aspekten dargelegte und kritisierte Praxis schließt eine Psychodiagnostik im Interesse der abhängig Beschäftigten weitgehend aus (SCHARDT 1976). Damit wird die nur langfristig realisierbare Forderung nach einer materialistisch fundierten arbeitsorientierten Eignungsdiagnostik unabweisbar, die insbesondere den dynamischen Aspekt menschlicher Fähigkeiten und Interessen im Sinne einer lebenslangen beruflichen Sozialisation auf der Basis einer Wechselwirkung mit der objektiven Arbeitssituation zu reflektieren hätte. Mit den LOTs scheint ein Testmodell vorgelegt worden zu sein, in dem wesentliche Prämissen einer solchen Alternative gewährleistet werden könnten, so daß deren Weiterentwicklung und Übertragung auf die Eignungsdiagnostik sowie methodenkritische Erprobung lohnen dürfte. Ob die schon längere Zeit vorliegenden PTMs ebenfalls eine solche Alternative darstellen können, muß - nicht zuletzt aufgrund praktischer Erwägungen - sehr bezweifelt werden, bedarf aber ebenfalls der weiteren Überprüfung. Als unmittelbar realisierbare Minimalforderung mit dem Ziel der stärkeren Berücksichtigung arbeitsorientierter Interessen ist die zufallskritische Absicherung psychodiagnostischer Daten mit Hilfe der Methoden der PED zu fordern.

Da der gesamte eignungsdiagnostische Begutachtungsprozeß neben der Eignungsbeurteilung im engeren Sinne einerseits notwendigerweise auf einer Arbeitsanalyse basieren muß (SCHMIDTKE & SCHMALE 1961) - was insbesondere bei den LOTs deutlich wird - und andererseits erst die psychologische Grundlage für die zu treffende Personalentscheidung (Selektion und Plazierung) liefert (SCHARDT 1976), kann eine arbeitsorientierte Eignungsdiagnostik auf eine Einbeziehung auch dieser beiden Phasen nicht verzichten. Hinsichtlich der Arbeitsanalyse ist festzustellen, daß

- die Analyse der tätigkeitsspezifischen bzw. berufsrelevanten Arbeitsanforderungen sich derzeit noch in einem höchst unbefriedigenden Zustand befindet, da eine Verwertung der bislang vorliegenden Ansätze im Interesse der abhängig Beschäftigten nur partiell und unter großen Einschränkungen möglich ist. Hier sind langfristig angelegte Forschungsaufgaben anzugehen, die auf die Entwicklung einer Allgemeinen Lohnarbeitspsychologie hinauslaufen, zu der kürzlich ein erster, vielversprechender Ansatz von VOLPERT (in GROSKURTH & VOLPERT 1975) vorgelegt wurde (vgl. SCHARDT 1975).

Hinsichtlich der auf der Basis der psychologischen Eignungsbeurteilung zu treffenden klassifikatorischen Entscheidungen ist festzustellen, daß

- trotz einer Vielzahl von entscheidungstheoretisch begründeten und auch z.T. empirisch hinreichend überprüften Klassifikationsmethoden die unterschiedlichen Selektions- und Plazierungsentscheidungen fast ausnahmslos nach völlig subjektiven Kriterien, d.h. letztlich willkürlich getroffen werden. Die von den vorliegenden Theorien und Methoden her (vgl. die Sammelreferate von JANKE 1964, 1969; SCHMALE 1969; LIENERT & v.KEREKJARTO 1969; BAUMANN 1974) unmittelbar realisierbare Aufgabe besteht somit darin, die empirische Begründung klassifikatorischer Personalentscheidungen allgemein verbindlich zu machen. Darüber hinaus wären als längerfristige Forschungsaufgabe die Weiterentwicklung taxonometrischer Methoden sowie deren Überprüfung anhand von Methodenstudien anzugehen (vgl. SCHARDT 1975).

Aus diesen allgemeinen Feststellungen lassen sich eine Reihe von konkreten Aufgabenstellungen für eine kurz- und mittelfristig angelegte Forschungsarbeit ableiten, mit deren Hilfe die Voraussetzungen für die Konzipierung einer arbeitsorientierten Eignungsdiagnostik geschaffen werden könnten. Folgenden Aufgaben müßte dabei Priorität zukommen:

(1) Es muß ein Katalog von inhaltlichen und methodischen Mindestkriterien entwickelt werden, denen Test- und Beurteilungsverfahren genügen müssen. Das beinhaltet hinsichtlich des erstgenannten Aspekts:

- Kriterien, die sicherstellen, daß die Verfahren ausschließlich mit dem Ziel eingesetzt werden können, für die Betroffenen optimale Voraussetzungen zur Entfaltung ihrer Persönlichkeit in beruflich-fachlicher, personaler und sozialer Hinsicht zu schaffen; d.h. die Arbeitsplätze nach dem Fähigkeits- und Interessenpotential des Bewerbers auszuwählen und nicht - wie bisher - umgekehrt. Dies ist wiederum untrennbar verbunden mit dem Problem der Gestaltung der Arbeit - der Humanisierung der Arbeitswelt. Weiterhin sind Kriterien erforderlich, mit deren Hilfe alle Verfahren aus der betrieblichen Praxis ausgeschlossen werden können, die nicht dem genannten

Ziel dienen (z.B. Verfahren, die in die Privatsphäre des Bewerbers eindringen).

Hinsichtlich des methodischen Aspekts sind erforderlich:

- Quantitative und verfahrenstechnische Kriterien für die Ermittlung der Reliabilität und Validität der Testverfahren, wobei die Validitätskriterien aus den o.g. inhaltlichen Kriterien abzuleiten sind. Dies hat gleichzeitig zur Konsequenz, daß die Testverfahren nach dem neuesten Stand wissenschaftlicher Erkenntnisse konstruiert werden müssen.
- Exakte Kriterien zur Anwendung der Verfahren, um eine maximale Durchführungs- und Auswertungsobjektivität sicherzustellen. Diese Forderung schließt z.B. aus, daß Verfahren von fachlich nicht hinreichend geschultem Personal - und das sind derzeit nur Diplom-Psychologen und u.E. Psychologisch-Technische Assistent(inn)en - angewendet werden. Der weit verbreiteten Beurteilung durch Vorgesetzte, die fast jeder Willkür offen steht, würde damit ein Riegel vorgeschoben.
- Eine exakt definierte Festlegung des Indikationsbereichs aller praktizierten Verfahren, um eine mißbräuchliche Anwendung auszuschließen. Diese muß mit den Validitätskriterien begründet sein. In positiver Hinsicht ist darin die Festlegung der Fragestellungen beinhaltet, zu denen die Testverfahren im Sinne einer Entscheidungsfindung herangezogen werden können, so daß die fast beliebige Verwendung irgendwelcher Beurteilungen ausgeschlossen wird.
- Kriterien, die die Verwertung von Testergebnissen für die Beratung des Bewerbers objektivieren und damit optimieren. Dies beinhaltet z.B., daß alle Daten nach den Kriterien der PED zufallskritisch abgesichert sein müssen wie auch darüber hinaus die interpretative Kombination von mehreren Testergebnissen (z.B. Profilvergleiche) methodisch-statistisch fundiert sein muß.

(2) Überprüfung sämtlicher in der betrieblichen Praxis verwendeten Verfahren anhand dieser Kriterien mit dem Ziel des Verbots der Anwendung aller Verfahren, die nicht diesen Mindestanforderungen genügen.

(3) Erarbeitung der psychologisch-inhaltlichen und methodisch-statistischen Grundlagen für die in einem zweiten Schritt zu entwickelnden psychodiagnostischen Verfahren, die den genannten grundlegenden Zielsetzungen entsprechen. Diese sehr aufwendige Arbeit beinhaltet unter dem inhaltlichen Aspekt:

- Die Entwicklung von Kriterien bzw. Kategorien für die Beurteilung von Arbeitsanforderungen unter der Prämisse der Entfaltung der Persönlichkeit (daß hierzu die REFA-Prüfliste nicht ausreicht, liegt auf der Hand).
- Die Operationalisierung dieser Kategorien und anschließende Gruppierung auf empirischer Grundlage, nach dem
- Kriterium des - notwendigerweise empirisch ermittelten - Zusammenhangs zwischen individuellen Merkmalen, die im Sinne der Persönlichkeitsentfaltung wünschenswert sind, und den vorher spezifizierten Arbeitsanforderungen.
- Die auf dieser Grundlage vorgenommene Konstruktion von Testverfahren, nach Möglichkeit mit Hilfe alternativer Methoden (z.B. LOTs).

Unter dem methodischen Aspekt ist u.a. die kritische Rezeption und Überprüfung der Brauchbarkeit neuerer Entwicklungen auf dem Gebiet der Meß- und Testtheorie (PED, LOT, PTM) der Erkenntnisse bezüglich impliziter Prozesse bei der diagnostischen Urteilsbildung (z.B. Halo-Effekt), der Weiterentwicklung und

empirischen Überprüfung insbesondere (multivariater) taxonometrischer Verfahren zu fordern.

(4) Entwicklung von Kriterien und Richtlinien für die Ausbildung des Personenkreises, der mit Personalbeurteilungen befaßt ist.

Abschließend soll nachdrücklich darauf hingewiesen werden, daß mit den genannten, sicherlich noch vorläufigen Aufgabenstellungen nur eine, wenn auch notwendige Voraussetzung für die Anwendung einer arbeitsorientierten Eignungsdiagnostik in der betrieblichen Praxis geschaffen werden kann. Der Glaube, daß allein mit der Bereitstellung wissenschaftlicher Erkenntnisse auch schon deren praktische Verwertung - und zwar im intendierten Sinne - gesichert sei, erweist sich in der betrieblichen Praxis immer wieder als illusionär. Auch was die praktische Umsetzung einer arbeitsorientierten Eignungsdiagnostik anbetrifft, hat die historische Erkenntnis der Arbeiterbewegung Gültigkeit: Die Durchsetzung der eigenen Interessen ist und bleibt eine Frage betrieblicher und gesellschaftlicher Kräfteverhältnisse. Den Interessen der abhängig Beschäftigten auch im eignungsdiagnostischen Bereich Rechnung zu tragen, wird nur in dem Maße möglich sein, in dem die Betroffenen selbst, in und mit ihren Organisationen - in erster Linie also den Gewerkschaften - bereit und in der Lage sein werden, eine arbeitsorientierte Eignungsdiagnostik zu verwirklichen.

Literatur

- BAUMANN, U. 1974. Gruppierung und Klassifikation: Statistische Probleme in der klinischen Psychologie. In: W.J.SCHRAML & U.BAUMANN (Hrsg.) Klinische Psychologie II. Methoden, Ergebnisse und Probleme der Forschung. Bern: Huber, S.77-112.
- DIEHR, D., FLAKE, R., HÜTTEMEISTER, L. & STÖMPFIG, G. 1970. Untersuchung der wirtschaftlichen Auswirkungen werksärztlicher Tätigkeit - Möglichkeiten der Produktivitätssteigerung. Bad Godesberg (zit.n. REGUS 1973).
- FISCHER, G.H. 1968. Abriß der klassischen Testtheorie. In: G.H.FISCHER (Hrsg.) Psychologische Testtheorie. Bern: Huber, S.15-53.
- FISCHER, G.H. 1973. Lineare logistische Modelle zur Beschreibung von Einstellungs- und Verhaltensänderungen unter dem Einfluß von Massenkommunikationen. Psychologisches Institut der Universität Wien, Research Bulletin, No.14, Nov.1973.
- FISCHER, G.H. 1974. Einführung in die Theorie psychologischer Tests. Grundlagen und Anwendungen. Bern: Huber. (Vollst.Neufassung von FISCHER 1968).
- FRIELING, E. 1975. Psychologische Arbeitsanalyse. Stuttgart: Kohlhammer.
- GROSKURTH, P. & VOLPERT, W. 1975. Lohnarbeitspsychologie. Berufliche Sozialisation: Emanzipation zur Anpassung. Frankfurt am Main: Fischer.
- GUTJAHR, W. 1971. Die Messung psychischer Eigenschaften. Berlin (DDR): Deutscher Verlag der Wissenschaften.
- HACKER, W. 1973. Allgemeine Arbeits- und Ingenieurpsychologie. Berlin (DDR): Deutscher Verlag der Wissenschaften.
- HOYOS, C.Graf 1974. Arbeitspsychologie. Stuttgart: Kohlhammer.
- HUBER, H.P. 1973. Psychometrische Einzelfalldiagnostik. Weinheim: Beltz.
- JANKE, W. 1964. Klassifikation. In: R.HEISS (Hrsg.) Handbuch der Psychologie, Bd.6: Psychologische Diagnostik. Göttingen: Hogrefe, S.3-16. (3.Aufl.1971).
- JANKE, W. 1969. Grundlagen der Klassifikation. In: M.IRLE (Hrsg.) Bericht über den 26.Kongreß der Deutschen Gesellschaft für Psychologie in Tübingen 1968, Göttingen: Hogrefe, S.135-154.
- KLAUER, K.J. 1975. Wie man einen lehrzielorientierten Test entwickelt. Psychologie heute, 2, (H.2, Febr.), S.37-39.
- KLAUER, K.J., FRICKE, R., HERBIG, M., RUPPRECHT, H. & SCHOTT, F. 1974. Lehrzielorientierte Tests. Düsseldorf: Schwann.
- LIENERT, G.A. 1967. Testaufbau und Testanalyse. Weinheim: Beltz (2.Aufl.).
- LIENERT, G.A. & KEREKJARTO, M.v. 1969. Möglichkeiten der Ex-Post-Klassifizierung depressiver Symptome und Patienten mittels Faktoren- und Konfigurationsanalyse. In: H.HIPPIUS & H.SELBACH (Hrsg.) Das depressive Syndrom. München: Urban & Schwarzenberg, S.218-256

- LORD, F.M. & NOVICK, M.R. 1968. Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley.
- Projektgruppe im WSI, 1974. Grundelemente einer Arbeitsorientierten Einzelwirtschaftslehre. Ein Beitrag zur politischen Ökonomie der Unternehmung. (WSI-Studien zur Wirtschafts- und Sozialforschung, Nr.23) Köln: Bund-Verlag.
- RASCH, G. 1960. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danmarks Paedagogiske Institut.
- RASCH, G. 1961. On general laws and the meaning of measurement in psychology. In: J.NEYMAN (Ed.) Proceedings of the forth Berkeley symposium on mathematical statistics and probability, Vol.IV. Berkeley: Univ.of California Press, S.321-334.
- RASCH, G. 1966. An individualistic approach to item analysis. In: P.F. LAZARFELD & N.W.HENRY (Eds.) Readings in Mathematical Social Science. Chicago: Science Research Associates, S.89-107.
- REGUS, M. 1973. Das Gesundheitswesen in der BRD. Situation und Perspektiven. In: H.U.DEPPE et al. Medizin und gesellschaftlicher Fortschritt. Köln: Pahl-Rugenstein, S.29-76.
- SCHARDT, L.P. 1975. Betriebliche Personalbeurteilung und Humanisierung der Arbeit. (Hektographiertes Manuskript). Düsseldorf (Dez.1975).
- SCHARDT, L.P. 1976. Wem nützt die Eignungsdiagnostik? Zum konservativen Charakter der klassischen Testtheorie und ihrer Anwendung in der betrieblichen Praxis. Bielefelder Arbeiten zur Sozialpsychologie, Nr.7 (April 1976).
- SCHMALE, H. 1969. Praktische Probleme der Klassifikation. In: M.IRLE (Hrsg.) Bericht über den 26.Kongreß der Deutschen Gesellschaft für Psychologie in Tübingen 1968. Göttingen: Hogrefe, S.154-159.
- SCHMIDTKE, H. & SCHMALE, H. 1961. Arbeitsanforderungen und Berufseignung. Methodik der eignungsdiagnostischen Normfindung. Bern: Huber.
- STENE, J. 1968. Einführung in Raschs Theorie der psychologischen Messung. In: G.H.FISCHER 1968, S.229-268.
- STRITTMATTER, P. (Hrsg.) Lernzielorientierte Leistungsmessung. Weinheim: Beltz.
- SEVE, L. 1973. Marxismus und Theorie der Persönlichkeit. Frankfurt am Main: Verlag Marxistische Blätter.
- VOLPERT, W. 1974. Die "Humanisierung der Arbeit" und die Arbeitswissenschaft. (Hefte zu politischen Gegenwartsfragen 16). Köln: Pahl-Rugenstein.
- WRIGHT, B.D. 1967. Sample-free test calibration and person measurement. In: Invitational conference on testing problems. Princeton. (Zit.n.GUTJAHR 1971).
- WSI-Forum, 1973. Arbeitsorientierte Einzelwirtschaftslehre contra Kapitalorientierte Betriebswirtschaftslehre. (WSI-Studien zur Wirtschafts- und Sozialforschung Nr.24) Köln: Bund-Verlag.

BIELEFELDER ARBEITEN ZUR SOZIALPSYCHOLOGIE

- Nr.1 (2/76) Hans Dieter Schmidt: Die Förderung nationalistischer Einstellungen durch internationale Sportwettkämpfe in sozialpsychologischer Sicht
- Nr.2 (2/76) Hans Dieter Schmidt, Amélie Schmidt-Mummendey, Friedrich-Wilhelm Schröder und Rüdiger Stallberg: Der Einfluß aggressiven Modellverhaltens, Status des Modells und Selbsteinschätzung auf offen aggressives Verhalten
- Nr.3 (3/76) Hans Dieter Schmidt, Amélie Schmidt-Mummendey, Ulrich Goertz, Günter-Maus, Friedrich-Wilhelm Schröder und Uwe Troske: Der Einfluß von Modell-Aggression und Modell-Status auf das Erteilen von Geldstrafen als instrumentell-aggressives Verhalten
- Nr.4 (4/76) Manfred Bornewasser: Naturwissenschaftliche und verhaltenstheoretische Orientierungen in der Sozialpsychologie (I)
- Nr.5 (4/76) Rosemarie Mielke und Hans Dieter Schmidt: Zur Psychologie politischen Verhaltens - eine empirische Analyse von Verhaltensmustern beim Kampf um Gleichberechtigung der Frau
- Nr.6 (4/76) Hans Dieter Schmidt: Selbstwahrnehmung und Beeinflussung
- Nr.7 (4/76) Lothar P. Schardt: Wem nützt die Eignungsdiagnostik? Zum konservativen Charakter der klassischen Testtheorie und ihrer Anwendung in der betrieblichen Praxis
- Nr.8 (5/76) Lothar P. Schardt: Ansätze zu einer arbeitsorientierten Eignungsdiagnostik
- Nr.9 (7/76) Manfred Bornewasser, Fritz Köster, Rosemarie Mielke und Cornelia Pawelcik: Erstellung und erste Überprüfung eines Fragebogens zur instrumentellen Arbeitsorientierung bei Jugendlichen
- Nr.10 (8/76) Hans Dieter Schmidt, Bernd Schiebel und Uwe Troske: Untersuchung der Stabilität von Modell-Effekten aggressiven Verhaltens gegenüber Populations- und Situationsunterschieden
- Nr.11 (9/76) Hans Dieter Schmidt: Selbstwahrnehmung als interpersonelle Wahrnehmung: Hypothesen
- Nr.12 (10/76) Hans Dieter Schmidt, Lothar P. Schardt und Rosemarie Mielke: Ein- und mehrdimensionale Skalierung der gewerkschaftlichen Kampfbereitschaft von Belegschaften
- Nr.13 (vergriffen)
- Nr.14 (1/77) Hans Dieter Mummendey, Rosemarie Mielke, Günter Maus und Bernhard Hesener: Untersuchungen mit einem mehrdimensionalen Selbsteinschätzungsverfahren

(Forschungsberichte in unregelmäßiger Folge. Redaktion: H.D.Mummendey, Universität Bielefeld, Fakultät für Soziologie, Postfach 8640, 4800 Bielefeld; dort für DM 1,50 pro Heft zu beziehen)