

1. Grundlegende Annahmen und Definitionen

1. Welche Bedeutung hat die Diagnostik für pädagogisches Handeln, und wie läßt sich dies begründen?
2. Wodurch unterscheidet sich professionelle Diagnostik von Alltagsdiagnostik?
3. Auf welche grundlegenden Annahmen stützt sich Diagnostik, und wie gelangt man zu möglichst genauen und zutreffenden Aussagen?
4. Wie ist professionelle Pädagogisch-psychologische Diagnostik zweckmäßig zu definieren?

Vorstrukturierende Lesehilfe

Erziehung und Unterricht sind Lebensbereiche, in denen Diagnostik eine besonders große Rolle spielt. Ständig müssen pädagogische Entscheidungen unterschiedlicher Tragweite getroffen werden. Ihre Wirksamkeit hängt u.a. davon ab, wie zutreffend die individuellen Lernvoraussetzungen und Fähigkeiten, aber auch die emotionale Verfassung und die motivationale Bereitschaft eines Schülers erkannt und berücksichtigt werden. Obwohl Differenzierung und Individualisierung seit langem als Prinzipien der Unterrichtsorganisation anerkannt sind, ist der Status der professionellen Pädagogisch-psychologischen Diagnostik hierzulande unbefriedigend.

Die wissenschaftlich fundierte Diagnostik, wie sie sich seit grob einhundert Jahren entwickelt hat, fußt auf der Alltagsdiagnostik. Die Alltagsdiagnostik ist jedoch in vieler Hinsicht unzulänglich. Die Persönlichkeitstheoretischen und methodischen Konzepte, die tragfähige Lösungen für die meisten diagnostischen Fragestellungen möglich machen, werden erläutert. Eine wesentliche Rolle spielen dabei die Präzisierung der Merkmale und der Meßoperationen sowie die Verifizierung der diagnostischen Aussagen. Die instrumentelle Qualität der diagnostischen Hilfsmittel muß umso höheren Ansprüchen genügen, je mehr von den Entscheidungen abhängt, zu deren Begründung sie beitragen sollen. Das wichtigste Kriterium ist daher die **empirische Validität** der Methoden.

Abschließend wird Diagnostik als ein systematisches Vorgehen zur Gewinnung und Analyse von Merkmalsunterschieden an Personen definiert.

1.1 Bezeichnung des Gebiets

Die Bezeichnung **“Pädagogisch-psychologisch”** im Text dieses Buches bringt zum Ausdruck, daß die Diagnostik, um die es hier geht, beides zugleich ist, pädagogische und psychologische Diagnostik. Sie ist **pädagogisch**, weil ihre Fragestellungen aus der Erziehungspraxis stammen und weil unser Text sich auf diese Praxis bezieht. Überall da, wo es aus pädagogischen Gründen notwendig oder ratsam ist, die indivi-

duellen Bedingungen und die Ergebnisse menschlichen Lernens zu kennen, und immer wenn individuelles Verhalten mit pädagogischen Mitteln beeinflusst werden soll, ist Diagnostik unerlässlich. Diagnostische Tätigkeiten haben nicht nur einen beträchtlichen Anteil am Unterrichtsalltag, ihnen kommt auch ein hoher, kaum zu überschätzender Stellenwert zu: Ohne eine fundierte diagnostische Routine wäre ein professioneller, am Individuum orientierter Unterricht nicht möglich. Besonders sorgfältige Diagnosen sind vonnöten, wenn unerwartete Schwierigkeiten im Verhalten und in den Leistungen von Schülern auftreten und wenn wichtige, langfristig wirksame Entscheidungen über die pädagogische Behandlung von Schülern zu treffen sind, z.B. bei der Einschulung, bei Umschulungen, bei Kurszuweisungen oder auch beim "Sitzenbleiben".

Psychologisch ist diese Diagnostik deshalb, weil sie in der Regel Verhaltens- und Leistungsaspekte betrifft, die Gegenstand der theoretischen wie der empirischen Erkenntnisgewinnung in der Psychologie sind. Fachhistorisch kommt hinzu, daß die diagnostischen Methoden und Theorien, über die wir heute verfügen, vornehmlich von Psychologen entwickelt worden sind. Doch wäre es müßig, hier pädagogische und psychologische Anteile auseinanderzuidividieren. Die moderne Pädagogisch-psychologische Diagnostik hat ihre Wurzeln selbstverständlich in der überkommenen pädagogischen Praxis. Aus deren diagnostischen Bedürfnissen ist sie entstanden, und sie dient nichts anderem, als eben diese Praxis zu verbessern.

Es liegt deshalb nahe, statt von Pädagogisch-psychologischer einfach von Pädagogischer Diagnostik zu sprechen. Diese verkürzende Bezeichnung ist vor etwa 20 Jahren von Ingenkamp (vgl. 1985, S. 10) in Analogie zum angloamerikanischen educational measurement/assessment eingeführt und von anderen Autoren aufgegriffen worden (z.B. Klauer, 1978; Süllwold, 1983). Die Verbindungen "heilpädagogische" bzw. "sonderpädagogische" Diagnostik waren ohnehin schon früher geläufig. Gegen die praktische Kurzform "Pädagogische Diagnostik" ist nichts einzuwenden, solange man sich bewußt bleibt, daß Pädagogische Diagnostik in der Regel - mit jeweils unterschiedlicher Akzentuierung - auch psychologische Diagnostik ist.

Im Unterschied zur ärztlichen Diagnostik, die vorwiegend auf die Feststellung (oder den Ausschluß) krankhafter Abweichungen vom "Normal"-Zustand des Gesunden oder Unauffälligen gerichtet ist, umfaßt die Pädagogische Diagnostik die gesamte Spannbreite der vorkommenden Zustände und Zusammenhänge, die für die individuelle pädagogische Behandlung **aller** Schüler bedeutsam sind. Zwar werden die erhobenen Daten auch in der Pädagogischen Diagnostik auf Vergleichswerte bezogen, doch ziehen nicht nur die auffälligen, von einer Norm abweichenden, sondern grundsätzlich **alle** Werte je unterschiedliche pädagogische Konsequenzen nach sich. Von wenigen Ausnahmen abgesehen, ist in der Pädagogik - anders als in der Medizin - stets eine "Behandlung", nämlich eine bestimmte pädagogische Maßnahme oder "Intervention", angezeigt.

In den folgenden Abschnitten werden einige allgemeine pädagogische und psychologische Grundlagen der Diagnostik kurz dargestellt.

1.2 Allgemeine pädagogische Grundlagen

Ganz gleich, wer in dieser Praxis tätig ist, jede diagnostische Aktivität ist eingebunden in ein konkretes pädagogisches Handlungsfeld. Dem Pädagogen erscheint dies

selbstverständlich; der Psychologe, der nicht zugleich auch pädagogisch ausgebildet und erfahren ist, muß sich erst darauf einstellen und entsprechende Kenntnisse erwerben. In seiner allgemeinsten Bedeutung wird pädagogisches Handeln zumeist als "Erziehen" bezeichnet. In diesem weiteren Sinne ist Erziehung als das **intentionale Herbeiführen relativ dauerhafter Veränderungen von Personmerkmalen durch mentale Beeinflussung** zu verstehen. **Veränderung von Personmerkmalen** bedeutet, individuelle Ist-Zustände in neue, vorgegebenen Soll-Werten entsprechende oder angenäherte Ist-Zustände zu überführen. **Mentale Beeinflussung** heißt, daß die intendierten Veränderungen über kognitive Prozesse der Informationsaufnahme und -Verarbeitung herbeigeführt werden, die ihrerseits vom emotionalen und motivationalen Zustand des Lernenden abhängen. Ziel der Erziehung ist es, mengentheoretisch veranschaulicht, die Schnittmenge der Verhaltens- und Erlebenselemente von Lehrenden und Lernenden systematisch zu vergrößern. Pädagogische Effekte gehen aber nicht nur von der gezielten interpersonalen Beeinflussung aus. Andere Umweltbedingungen, namentlich soziakulturelle Faktoren, wie Vorbild- und Medienwirkungen, tragen ebenfalls, wenn auch eher beiläufig, zum Erwerb oder zur Veränderung von Merkmalen bei (sog. **funktionale** Erziehung).

Diese knappe Definition genügt für unsere Zwecke. Im Unterschied zu manchen pädagogischen Definitionen von Erziehung ist sie insofern wertungsfrei, als sie über Wert oder Unwert der Ziele, Intentionen und Ergebnisse des pädagogischen Handelns keine Vorentscheidung trifft. Sie ist zugleich offen in Richtung auf pädagogisch-psychologische **Verhaltensmodifikation** oder **Therapie**, von der wir sprechen, wenn es um die mentale Beeinflussung von Ausgangszuständen mit Krankheitswert geht. Im Regelfall stellt Erziehen eine Art **asymmetrischer sozialer Interaktion** dar, deren beabsichtigte (oder unbeabsichtigte) Wirkung sich stets auf Individuen richtet und nur an Individuen manifest werden kann.

Der pädagogische Grundsatz, dabei die individuelle Eigenart der Kinder zu beachten, ist alt. Ideengeschichtlich geht er in der Neuzeit auf den künstlerischen und intellektuellen Individualismus der Renaissance und des Humanismus zurück. Zahlreiche Denker haben seither die pädagogische Bedeutung des **Individualitätsprinzips** hervorgehoben und präzisiert. Schon der große Pädagoge Arnos Comenius (1592-1670), dem es eigentlich um eine breit angelegte Massenerziehung ging, hat gefordert, neben dem Alter auch die Unterschiede in der Begabung (Veranlagung) und im Lernfortschritt der Kinder systematisch zu berücksichtigen. Er formuliert bereits didaktisch **differenzierte** Vorschriften, die im wesentlichen bis heute Bestand haben. Dem Kulturkritiker und Aufklärer Jean Jacques Rousseau (1712-1778) wird die psychologisch bedachtsame Förderung der natürlichen Individualität des Kindes zum Angelpunkt für eine tiefgreifende Umgestaltung der Gesellschaft. Der bekannte Schweizer Erzieher Johann Heinrich Pestalozzi (1746-1827) hat dann den zentralen pädagogischen Begriff der "Individuallage" geprägt. Darin werden die zeitgeschichtlichen, die gesellschaftlichen und materiellen Umweltverhältnisse, in die ein Kind hineingeboren wird, mit den personalen Aspekten seiner Konstitution, seiner Urteilsfähigkeit, seiner Motivation und seiner Charaktereigentümlichkeiten zu einem dynamischen Gesamtkonzept vereint. Dessen psychologischer Kern besteht aus der Art und Weise, wie sich das Kind in seiner sozialen Umwelt erlebt. Pestalozzi fordert, daß alle Erziehung von der **Individuallage** des Kindes auszugehen hat. In bewußter Abkehr von jeglichem Schematismus und Formalismus in der Pädagogik wurde schließ-

lich der Ruf "Vom Kinde aus" zur sinnfälligsten Maxime der Schulreformbewegung des ausgehenden 19. und beginnenden 20. Jahrhunderts (s. Dietrich, 1982).

Mit der zunehmenden Verwissenschaftlichung vieler Lebensbereiche werden in dieser Zeit auch die ersten diagnostischen Verfahren im heutigen Sinne entwickelt. Mit ihrer Hilfe sollte die alte Forderung nach der Differenzierung des Unterrichts besser in die Praxis umgesetzt werden, als dies aufgrund des üblichen Rückgriffs auf die Alltagserfahrung von Lehrern und Eltern möglich erschien. Bezeichnenderweise vollzog sich ein wesentlicher Teil dieser frühen Entwicklung innerhalb einer neuen Forschungsrichtung, die von ihren Verfechtern als "experimentelle Pädagogik" bezeichnet wurde (Ernst Meumann, 1862-1915; Wilhelm August Lay, 1862-1926). Sie zielte darauf ab, die Pädagogik insgesamt zu einer (möglichst) exakten Wissenschaft auszubauen und die Erziehungspraxis auf ein empirisch gesichertes Fundament zu stellen (vgl. z.B. Lay, 1903; Meumann, 1907. Beide gaben seit 1905 die ***Zeitschrift für Experimentelle Pädagogik*** heraus).

Differenzierung und Individualisierung sind heute unumstrittene Organisationsprinzipien für Schule und Unterricht. Die Pädagogisch-psychologische Diagnostik hat seit der Jahrhundertwende, vor allem in den USA, einen bemerkenswerten Aufschwung genommen. Trotz bedeutender Anteile deutscher Autoren an dieser Entwicklung und trotz erheblicher Fortschritte in den letzten Jahrzehnten ist die Akzeptanz der standardisierten diagnostischen Verfahren (Tests und Fragebogen) im deutschen Sprachraum hinter den Möglichkeiten, die sie bieten, aus verschiedenen Gründen zurückgeblieben (Tent, 1969, S. 28-33; Ingenkamp, 1985, S. 257-264). Zu diesen Gründen zählen gewisse Unzulänglichkeiten, die selbst bei sorgfältiger Konstruktion der Verfahren in Kauf genommen werden müssen, sowie ideologische Vorbehalte gegenüber der Funktion solcher Hilfsmittel und Probleme mit der sachgerechten Verwertung ihrer Ergebnisse (zur Kritik an der Pädagogischen Testdiagnostik vgl. Ingenkamp, 1989). Bis jetzt ist die wissenschaftlich begründete Diagnostik bei uns weder in der Unterrichtspraxis noch in der Lehrerbildung zur Selbstverständlichkeit geworden.

Der Verwissenschaftlichung von Praxis, allzumal der pädagogischen, sind sicher Grenzen gesetzt. Aber auch in der Pädagogik erfordert professionelles Handeln, praktische Entscheidungen wo immer möglich auf empirisch gesicherte Erkenntnisse zu stützen. Wenn man hier, wie andernorts, das Prinzip der Optimierung von Handlungsentscheidungen anerkennt, gibt es keine rationalen Gründe, auf dafür geeignete Erkenntnismittel zu verzichten. Die Unvollkommenheit der Instrumente ist kein Gegenargument, solange nachweislich bessere Alternativen nicht zur Verfügung stehen und die bekannten Eigenschaften und Schwächen der Verfahren angemessen beachtet werden. Der allgemeine triviale Grundsatz, das jeweils Bestmögliche zu tun, gilt hier ebenso uneingeschränkt wie der Grundsatz, empirisch und nicht ideologisch zu entscheiden, was unter gegebenen Bedingungen und im Hinblick auf gegebene Ziele das Bestmögliche ist. Dazu muß man die in Betracht kommenden Alternativen gründlich genug kennen.

1.3 Alltagsdiagnostik

Diagnostische Urteile zu bilden, ist ein alltäglicher Vorgang. Wir beobachten und registrieren die Erscheinung, das Auftreten und die sprachlichen Äußerungen anderer,

schätzen das Aufgenommene ein und schließen daraus auch auf “die Person”. Wir trachten danach, uns ein Bild vom anderen zu machen. Wir möchten wissen, mit wem wir es zu tun haben, “was für ein Mensch” der andere ist, was wir künftig von ihm zu erwarten haben und wie wir uns am besten darauf einstellen. Wir haben erkannt, daß unser eigenes Handeln auf ganz bestimmte Fähigkeiten, Zielvorstellungen, Bedürfnisse und Interessen zurückgeht, und zumeist können wir die (oft widerstreitenden) Beweggründe angeben, die uns veranlaßt haben, dieses oder jenes zu tun oder zu lassen. Entsprechende Gedanken machen wir uns darüber, weshalb sich andere so und nicht anders verhalten. Wir versuchen u.U., uns in den anderen “hineinzusetzen”. Wir wollen ihn so gut wie möglich verstehen. Denn Verständnis verschafft uns Sicherheit; es vermittelt den Eindruck, dem anderen gerecht zu werden, besser mit ihm auszukommen und gezielter auf ihn einwirken zu können. Wir vergleichen zu diesem Zweck die Menschen miteinander und orientieren uns dabei an Maßstäben, die unserer eigenen Erfahrung oder dem überkommenen “diagnostischen Regelwissen” entstammen (z.B. “Stille Wasser gründen tief”, “Der Apfel fällt nicht weit vom Stamm”, “An seinem Umgang erkennt man den Menschen” usw.).

Diese Vorgänge werden uns am ehesten dann bewußt, wenn es sich um jemanden handelt, mit dem wir in der Folgezeit häufig zusammentreffen wollen oder müssen, etwa wenn wir uns Freunde oder Partner aussuchen oder wenn z.B. ein Lehrer eine neue Klasse übernimmt, bzw. Schüler einen neuen Lehrer bekommen. Wir neigen dazu, wiederholte Beobachtungen zu verallgemeinern, und je besser wir einen Menschen zu kennen glauben, desto mehr gehen wir dazu über, ihm ganz bestimmte Fähigkeiten und Eigenschaften zuzuschreiben. Die Alltagssprache enthält eine nahezu unerschöpfliche Fülle von Ausdrücken und Wortkombinationen, mit deren Hilfe wir Menschen, uns selbst eingeschlossen, “charakterisieren” können. Jemand, bei dem uns dies nicht recht gelingt, bleibt uns fremd; wir bezeichnen ihn dann z.B. als “verschlossen”, “in sich gekehrt” oder “undurchsichtig”.

Selbstkritische Beobachter stellen in der Tat fest, daß es oft nicht einfach ist, zu schlüssigen Ergebnissen zu gelangen. Sie machen die Erfahrung, daß nicht nur einzelne Menschen, sondern auch einzelne Verhaltensbereiche unterschiedlich gut einzuschätzen sind. Sie wissen, daß sie sich täuschen können und sind deshalb auch nicht überrascht, wenn das Erwartete nicht eintrifft. Sie hüten sich, vorschnell zu urteilen und zu verallgemeinern. Sie bleiben bei vorsichtiger Vermutung, wo der Eilfertige sich schon eine feste Überzeugung bildet. Sie bedenken, daß Menschen, vor allem Kinder, sich mitunter stark und manchmal schnell verändern können. Sie differenzieren von Fall zu Fall zwischen Verhaltensweisen, bei denen sich Änderungen oder Schwankungen bemerkbar machen, und solchen, die vergleichsweise stabil erscheinen.

Dem aufmerksamen Beobachter entgeht auch nicht, daß sich Menschen vielfach unterschiedlich verhalten, je nachdem, wem sie gegenüberstehen, in welcher Situation und in welcher Verfassung sie sich befinden. Der einzelne Beurteiler ist kaum in der Lage, die ganze Spannweite des Verhaltens anderer zu überblicken. Es kommt hinzu, daß die Fähigkeit, das Verhalten unserer Mitmenschen “richtig” zu deuten, offenbar nicht jedem in gleicher Weise gegeben ist.

Zwar entwickeln wir schon als Kinder, im ständigen Austausch mit den Menschen in unserer Umgebung, ein diagnostisches “Gespür”, und mit wachsender Erfahrung verfeinern sich unsere diagnostischen Fertigkeiten und werden reichhaltiger. Und doch dauert es offenbar lange, bis wir jemanden “wirklich” kennen. So jedenfalls

belehrt uns das Sprichwort vom Scheffel Salz, den man erst mit ihm gegessen haben muß. Aber können wir uns überhaupt auf unsere Menschenkenntnis verlassen? Ist nicht jeder Mensch so unverwechselbar und einmalig mit seinem individuellen Wegdegang und seiner vielschichtigen Umwelt, daß jeder Versuch, ihm diagnostisch gerecht zu werden, grundsätzlich scheitern muß? Die Literatur ist voller Widersprüche. Sie spricht vom ewigen Rätsel, das der Mensch dem Menschen bleibe, aber auch davon, daß das Erfassen nur eines Menschen in seiner ganzen Tiefe den Weg dazu eröffne, alle zu verstehen (s. Kasten S. 21).

1.4 Professionelle psychologische Diagnostik

Der Rückgriff auf die Alltagspraxis soll verdeutlichen, daß die wissenschaftlich begründete Diagnostik von heute in der Tat aus dieser Praxis hervorgegangen ist und nach wie vor enge Beziehungen dazu hat. In den Anfängen ihrer gut hundertjährigen Geschichte stehen konkrete Fragestellungen aus dem klinischen und pädagogischen, aus dem betrieblichen, dem forensischen und dem militärischen Bereich im Vordergrund. Doch spielen auch mehr theoretische Interessen an interindividuellen Unterschieden schon frühzeitig eine Rolle. Ihre theoretischen Grundlagen werden teils in unmittelbarem Zusammenhang mit der diagnostischen Praxis, teils erst nachträglich entwickelt.

Alle wesentlichen Elemente, Annahmen und Probleme der professionellen Diagnostik sind in der diagnostischen Alltagspraxis vorgebildet oder lassen sich daraus herleiten. Beide stützen sich auf das in den Grundzügen gleiche Persönlichkeitsmodell. Danach kommt allen Menschen ungeachtet ihrer Individualität dasselbe Spektrum psychophysiologischer Funktionen, Erlebnisqualitäten und Verhaltenskategorien zu, und danach lassen sich im beobachtbaren Verhalten Zusammenhänge erkennen, die man in Regeln fassen kann. Beide gehen auf dasselbe menschliche Grundbedürfnis nach Überschaubarkeit und Sicherheit der Lebensvollzüge zurück, und beide dienen demselben allgemeinen Zweck: menschliches Verhalten "kalkulierbar" zu machen und durch möglichst zutreffende Vorhersagen die Ungewißheit im zwischenmenschlichen Umgang zu reduzieren.

Diagnostik ist selbstverständlich nur ein Teil unseres Bemühens, Ungewißheit zu verringern. In erster Linie wird dies durch überindividuell verbindliche, einheitliche Verhaltensrichtlinien wie Sitten und Gebräuche, Verträge, Regeln, Normen und Sanktionen bewirkt. Von daher wäre "individuelles" Verhalten eher als ein nicht weiter reduzierbarer Unsicherheitsfaktor zu verstehen, der sich der Prognostizierbarkeit grundsätzlich entzieht. Für die partielle Gültigkeit dieses Arguments scheint die Tatsache zu sprechen, daß die psychologische Diagnostik zwar die relativ besseren Informationen liefert, in weiten Bereichen aber nach wie vor keine befriedigende Vorhersage machen kann.

Daraus zu folgern, man solle in der psychologischen und pädagogischen Praxis auf gründliche Diagnosen verzichten, wäre ein Fehlschluß. Die allgemeine Erfahrung und die vielfach belegte Gültigkeit diagnostischer Methoden sprechen gegen eine solche Auffassung. In wichtigen Teilbereichen menschlichen Verhaltens können wir offenbar mit hinreichender intraindividuelle Kontinuität und mit hinreichender Beständigkeit interindividuelle Differenzen rechnen. Hinreichend heißt, daß eine darauf ge-

Wie gut ist die Alltagsdiagnostik?

(1) Meili (in Meili & Steingrüber, 1978, S. 25-26) hat die *Intelligenz* dreier sechsjähriger Jungen mit deutlich verschiedenen IQ unter drei Bedingungen schätzen lassen. Die Beurteiler konnten sich auf Standfotos oder Filmaufnahmen stützen oder das reale Verhalten der Jungen beobachten. Sie sollten die Jungen hinsichtlich ihrer Intelligenz in eine Rangordnung bringen. Die Ergebnisse waren nicht besser als der Zufall. Die drei Darbietungsarten unterschieden sich darin nicht (s. Tabelle).

Beurteilung der Intelligenz nach Foto, Film und Natur; Einstufung in % (Richtige Beurteilungen kursiv).

Intelligenz	Darbietungsart									
	Foto			Film			Natur			
		gut	mittel	schwach	gut	mittel	schwach	gut	mittel	schwach
Beste(r)	(IQ 115)	28	62		17	50	33	12		41
Mittlere(r)	(IQ 100)	14	21	62	25	29	46	47	17	36
Schwache(r)	(IQ 77)	58	17	25	58	21	21	41	36	23

(2) Schwieriger ist es, die Brauchbarkeit der Beurteilung von *Temperamentsmerkmalen* (Persönlichkeitseigenschaften i.e.S.) zu überprüfen. Cohen (1969, Abschn. II) ließ mit Hilfe von 15 bipolaren Schätzskalen, die "Dominanz", "Beliebtheit" und "Gewissenhaftigkeit" erfaßten, sowohl persönlich bekannte als auch unbekannte Personen beurteilen. Von den unbekannt Personen lagen Fotos, Handschriften und Selbsteinschätzungen vor. Bei Einschätzung aufgrund persönlicher Bekanntschaft stimmten die Beurteiler zwar mit $r = 0.34$ im Mittel signifikant höher überein als bei Einschätzung anhand von Fotos, Handschrift und Selbsteinstufung; dies reicht jedoch als Grundlage für die Gültigkeit von Eindrucksurteilen nicht aus. Auch bei simultaner Beurteilung mehrerer Informationsquellen schwankten die Korrelationen im Mittel lediglich zwischen $r = 0.03$ und maximal 0.27.

(3) Die Ergebnisse der Forschung zur Zuverlässigkeit und Gültigkeit (Validität) diagnostischer Eindrucksurteile hat Merz (1963, S. 44) wie folgt zusammengefaßt:

1. Die *Zuverlässigkeit von Eindrucksurteilen* ist, von Sonderfällen abgesehen, gering. Es wurden knapp mittlere Koeffizienten gefunden. Jedoch ergeben sich fast unter allen Umständen gewisse Übereinstimmungen zwischen verschiedenen Beurteilern, gleichgültig, wie unzureichend die zur Verfügung stehenden Informationen sein mögen.
2. Die *Validität von Eindrucksurteilen* ist verständlicherweise noch geringer, es wurden Koeffizienten zwischen etwa 0,00 und 0.50 gefunden. Auch bei deutlichen Übereinstimmungen zwischen verschiedenen Beurteilern kann die Validität gleich Null sein.
3. Die *Validität von Beurteilungen* steht nur in recht lockerem Zusammenhang mit verschiedenen Persönlichkeitsmerkmalen der Beurteiler. Auch der Zusammenhang zu anderen unabhängigen Variablen ist gering.
4. Der *Inhalt der Beurteilungen* ist enger an andere Bedingungen gebunden als an die Individualität des Beurteilten. Solche Bedingungen sind u.a. die Eigenart des Beurteilers, die Eigenart der sozialen Beziehung zwischen Beurteiler und Beurteiltem und allgemeine Faktoren, wie etwa Stereotype.
5. Das auffälligste *Einzelergebnis* besteht wohl darin, daß die Validität von Beurteilungen weitgehend unabhängig ist von Art und Umfang der Informationen, welche dem Beurteiler über den Beurteilten zur Verfügung stehen."

gründete Diagnostik Entscheidungen und Behandlungszuweisungen ermöglicht, die insgesamt zu nachweislich besseren Ergebnissen führen, als das bei Vernachlässigung dieser Erkenntnisquelle und Anwendung anderer Strategien der Fall wäre.

Aus wissenschaftspragmatischer Sicht ist dabei unerheblich, ob das deterministische Menschenbild, das diesem Ansatz zu Grunde liegt, universell gültig ist oder nicht. Die theoretische Leitvorstellung, daß wenn alle Bedingungen bekannt wären, auch alles vorhergesagt werden könnte, ist, auf menschliches Verhalten angewendet, eine durchaus zweckmäßige Utopie, zumal sie das Auftreten unvorhergesehener Ereignisse keineswegs ausschließt. Ob die perfekte Vorhersage individuellen Verhaltens jemals Realität wird (und ob sie wünschenswert ist), kann vorläufig offen bleiben. Es ist die gemeinsame Aufgabe von Forschung und Praxis zu erkunden, wie weit eine solche heuristische Generaltheorie trägt. Trotz seiner theoretischen und empirischen Unzulänglichkeiten ist dieser Ansatz ein gangbarer und erfolversprechender Weg, den zu verlassen voreilig wäre, solange es an nachweislich besseren Alternativen fehlt. Für einen solchen Schritt ist die Geschichte der professionellen psychologischen Diagnostik zu kurz.

Diagnostische Feststellungen, z.B. "Thomas ist gar nicht so dumm, wie wir anfangs dachten" oder "Thomas hat einen IQ von 104", weisen Individuen einer Klasse von Personen zu, die sich in bestimmter Hinsicht untereinander gleichen oder ähneln. In diesem Beispiel wird behauptet, daß Thomas einer Klasse seiner Bevölkerung angehört, die man als durchschnittlich intelligent bezeichnet, unabhängig davon, wie stark sich die Mitglieder dieser Klasse im übrigen, z.B. nach Alter, Geschlecht, sozialer Herkunft oder Lebenswandel, unterscheiden.

Von dem Vorgehen im Alltag hebt sich die professionelle Diagnostik im wesentlichen durch folgende Kriterien ab:

- (a) Präzisierung der Begriffe, insbesondere der Merkmale, die erfaßt werden sollen
- (b) Präzisierung der Meßoperationen durch
 - Standardisierung der Verfahren
 - Ökonomisierung der Informationsaufnahme und -Verarbeitung
 - Bereitstellung von Vergleichsmaßstäben
- (c) Verifizierung der diagnostischen Aussagen und der darauf gestützten Entscheidungen.

Auf diese Kriterien wird in den folgenden Abschnitten näher eingegangen.

1.4.1 Präzisierung der Merkmale

1.4.1.1 Person und Merkmal

In der Persönlichkeitstheorie wird "Person" als eine je einzigartige, unteilbare Ganzheit mit vielschichtigen Bezügen zu sich und ihrer Umwelt verstanden. Psychologisch gesehen, sind Personen hochkomplexe, sich selbst bewußte Systeme mit dem "Ich" als Zentrum des Erlebens und der Verhaltenssteuerung. Aus den bereits dargelegten Gründen könnte der Anspruch, solche individuellen ganzheitlichen Gefüge erschöpfend abzubilden, weder von der professionellen Diagnostik noch von irgendeiner anderen existierenden Diagnostik eingelöst werden. Darauf bezogene Vorbehalte gegenüber der psychologischen Diagnostik sind auch deshalb unerheblich, weil wir uns in der Praxis je nach Fragestellung offenbar ohne zu großen Informationsverlust auf

bestimmte Ausschnitte beschränken können. Unsere Aussagen kennzeichnen nicht die Person schlechthin, sie beziehen sich auf die Ausprägung definierter **Merkmale**, die der betreffenden Person zusammen mit einer Vielzahl anderer **Personen als Merkmalsträger** zukommen. Dies gilt auch dann, wenn mehrere Einzelmerkmale zu einem Merkmals-Ensemble höherer Ordnung integriert werden (z.B. "Petra ist hochgradig introvertiert").

Mit "Merkmal" ist ein für die Diagnostik zentraler Begriff eingeführt, der uns bereits aus anderen Teilgebieten der Psychologie geläufig ist. Im psychologischen Sprachgebrauch bezieht sich der Begriff "Merkmal" (oder Variable) auf einen **stabil unterscheidbaren Aspekt mit mindestens zwei Ausprägungsvarianten, anhand derer Objekte gruppiert und Veränderungen an Objekten festgestellt werden können**. Unter "Objekt" sind hier Personen, Verhaltensprodukte (z.B. schriftliche Klassenarbeiten) und Sachverhalte (z.B. "Hans wächst in einem anregungsarmen häuslichen Milieu auf") zusammengefaßt. Die kaum übersehbare Fülle von Merkmalen ist nach verschiedenen, sich mehrfach überschneidenden Gesichtspunkten zu ordnen. Gerade in der psychologischen Diagnostik müssen wir - aus theoretischen wie aus praktischen Gründen - auf eine solche Ordnung Wert legen, weil diagnostische Aussagen nur dann richtig formuliert und verstanden werden können, wenn Art, Qualität und Funktion der Merkmale, für die Informationen vorliegen, sorgfältig beachtet worden sind. Tabelle 1 (S.24/25) enthält eine schematisierte Übersicht über die wichtigsten Aspekte zur Unterscheidung diagnostisch relevanter Merkmalsklassen.

Als Psychologen sind wir in erster Linie an Merkmalen interessiert, die den einzelnen Personen zukommen, insbesondere natürlich an psychologischen. Doch setzt die operationale Definition mancher psychologischer Merkmale, z.B. des IQ, die Kenntnis nicht-psychologischer Merkmale, vor allem von Zeit- und Altersvariablen, voraus, die dadurch die Funktion von unabhängigen Variablen erhalten. Ebenso kann die Bedeutung individueller diagnostischer Rohwerte durch andere biologische und soziographische Merkmale, wie Geschlecht, Krankheit, besuchter Schultyp, soziale oder ethnische Herkunft, und durch spezifische Umweltvariablen, z.B. Kriminalität des Vaters oder aktuelle Scheidungsauseinandersetzungen der Eltern, relativiert werden.

1.4.1.2 Anlage und Umwelt

Unter **Umwelt** verstehen wir die Gesamtheit aller Reize, die während seines Lebens auf ein Individuum wirken, wobei die Reizauswahl und die Reizwirkung durch die vom Individuum ausgehende Aktivität mehr oder weniger mitbestimmt werden. Wir unterscheiden unscharf, aber zweckmäßig, drei Klassen von Umweltvariablen: physikalisch-chemische (z.B. intrauterines Milieu; Klima, Wetter, Luftqualität), materielle (Versorgung mit Bedarfsgütern) und soziokulturelle (personale Beziehungen, mentale Anregungsbedingungen). Ferner unterteilen wir die Umweltvariablen nach der Dauer und Kontinuität ihrer Anwesenheit, wiederum unscharf, in langfristig wirksame und aktuelle (situative). Die Umweltfaktoren (U) sind offensichtlich am Zustandekommen des Verhaltens (V) von Personen beteiligt, doch wissen wir in den meisten Fällen nicht genau, wie das geschieht, welchen Anteil sie daran haben und wie sie mit den genetischen Faktoren (G) zusammenwirken.

Tabelle 1.1: Vereinfachte Übersicht über diagnostisch bedeutsame Merkmalsklassen.

<i>Klassenbildender Gesichtspunkt</i>	<i>Merkmalsklassen</i>	<i>Beispiele</i>
<i>Merkmalsträger und Objektbereich</i>	Person-Merkmale - biologische - soziobiographische - psychologische Umwelt-Merkmale - physikalisch-chemische - materielle - soziokulturelle	Geschlecht, Körpergröße Sozialschichtzugehörigkeit kognitive Entwicklung Lufttemperatur, Lärmpegel Prokopf-Einkommen der Familie soziales Klima
<i>Datenherkunft</i>	direkt (am Pb) erhobene Merkmale indirekt (über den Pb) von Dritten erhobene Merkmale	Schulleistungsergebnisse, Erfolgs-/Miß- erfolgsattribuierung, Verhaltensstile Auskünfte von Lehrern, Eltern, Behörden (Akten)
<i>Anzahl der unterscheidbaren Ausprägungen</i>	alternativ (qualitativ) dichotomisiert mehrkategorial (qualitativ) kontinuierlich	Geschlecht, rechts/links oberhalb/unterhalb des Medians Berufe, Konfessionszugehörigkeit Alter, Intelligenz, Extraversion
<i>Skalenniveau</i>	nominalskaliert ordinalskaliert intervallskaliert Verhältnisskala	Beruf des Vaters/der Mutter soziale Hierarchien (Hack- ordnungen), Präferenzen IQ, emotionale Stabilität Alter, Körpergröße
<i>Verteilungsform</i>	normalverteilte Merkmale nicht-normal-, z.B. asymmetrisch verteilte Merkmale	IQ, Ängstlichkeit, Konzentrationsleistungen Schulnoten in der Grund- schule, Fehlerzahlen
<i>Meßgenauigkeit</i>	Merkmale mit hoher innerer Konsistenz Merkmale mit geringer innerer Konsistenz	kognitive Leistungsgeschwindigkeit, IQ, Extraversion Aggressivität im PFT, die meisten Merkmale in Formdeuteverfahren
<i>Dauerhaftigkeit der Merkmalsausprägung</i>	längerfristig stabile Merkmale (periodisch) schwankende Merkmal kurzfristige, aktuelle Merkmale	Intelligenz Jugendlicher, Introversion/Extraversion Verstimmtheit, Wohlbehagen Zustandsangst, Erregtheit, Arger, Müdigkeit, Freude
<i>Komplexitätsgrad</i>	hoch komplexe (mehrdimensionale) Merkmale typologisch gebündelte Merkmale "einfache" (eindimensionale) Merkmale	Intelligenz, Interessenspektrum, Schulerfolg Extraversion/Introversion, Neurotizismus, Maskulinität Sehscharfe, feinmotorische Steuerung, erlebte elterliche Strenge
<i>Generalisierungsbreite</i>	situationsübergreifende Merkmale situationsgebundene Merkmale	emotionale Labilität, Intelligenz, Rigidität spezielle berufliche Fertigkeiten, Phobien, Prüfungsangst, spezielle Einstellungen
<i>Instrumentelle Funktion</i>	abhängige Merkmale (Kriterien) unabhängige Merkmale (Prädiktoren) deskriptive Merkmale explikativ verwendete Merkmale Moderatorvariablen	Schulleistung von IQ Schulleistung für Lebenserfolg Schulversagen soziale Deprivation für Schulversagen Motivation im Zusammenhang von IQ und Schulleistung
<i>Diagnostische Relevanz (relative Validität)</i>	hoch valide Prädiktoren gering valide Prädiktoren	Intelligenz für Mathematikleistung, Vorkenntnisse Selbstbild und Verhaltensstile für Zeugnisnoten

Fortsetzung von Tabelle 1.1:

Klassenbildender Gesichtspunkt	Merkmalsklassen	Beispiele
Theoretischer Status (diagnostische Aussageebene)	beobachtete Merkmale (Verhalten und Verhaltensprodukte) Indikatorvariablen erschlossene Merkmale (latente Merkmale, Konstrukte, intervenierende Variablen, Dimensionen, Dispositionen)	Schreibgeschwindigkeit, Lesefehler, Gedächtnisleistungen, Testrohwerte, Aufsätze Zeugnisnoten, IQ kognitive Leistungsfähigkeit (g-Faktor), Sprachverständnis, Leistungsmotivation, Eigenschaften wie Zuverlässigkeit, Hilfsbereitschaft

Unsere allgemeine Vorstellung, derzufolge jedes menschliche Verhalten als eine Funktion von Anlage- und Umweltfaktoren zu verstehen ist,

$$[1.1] \quad V = f(G, U)$$

wird nach dem gegenwärtigen Erkenntnisstand gewöhnlich durch die Modellannahme spezifiziert, daß sich die empirische Varianz (Var) eines psychologischen Merkmals (X) additiv aus genetischen Anteilen (G), Umweltanteilen (U), den Kovarianzen (Cov G,U) zwischen diesen Komponenten sowie den Wechselwirkungen (G · U) und Meßfehleranteilen (E) zusammensetzt. Unter der Voraussetzung, daß außer dem Meßfehler auch die Wechselwirkungskomponente mit den anderen Anteilen nicht korreliert, ergibt sich folgende Varianzzerlegung:

$$[1.2] \quad \text{Var}(X) = \text{Var}(G) + \text{Var}(U) + 2\text{Cov}(G,U) + \text{Var}(G \cdot U) + \text{Var}(E).$$

Obwohl diese Gleichung für psychologische Merkmale bis jetzt nicht befriedigend ausgefüllt werden kann, hat sie neben ihrer theoretischen Bedeutung auch eine **unmittelbare praktische Konsequenz** für die Diagnostik: Sie verbietet in der großen Mehrzahl der Fälle die "ätiologisch" einseitige Interpretation psychologischer Befunde. Ob, bzw. wieweit z.B. ein festgestellter Rückstand in der Sprachentwicklung bei im übrigen unauffälliger Intelligenz auf das häusliche Milieu des Kindes, auf erworbene organische Mängel oder auf genetische Faktoren zurückgeht, ist den Testwerten nicht zu entnehmen und auch anhand von Anamnesedaten vielfach nicht zu entscheiden.

Doch selbst wenn eine Gleichung vom Typ [1.2] für ein Merkmal in einer Bevölkerung vorläge, wäre zu berücksichtigen, daß die Elemente der Gleichung von Individuum zu Individuum unterschiedliche Werte annehmen können und wir die Zusammensetzung im Einzelfall kennen müßten. Außerdem ist damit noch nicht gesagt, was die aufgeklärten Varianzanteile im Hinblick auf die Möglichkeit bedeuten, das registrierte Verhalten durch gezielte Beeinflussung zu verändern (Merz & Stelzl, 1977).

Selbst wenn wir sicher sein könnten, daß z.B. die kognitive Leistungsfähigkeit, wie sie in Intelligenztests gemessen und im IQ zusammengefaßt wird, im Mittel zu höchstens 20 % durch Umwelteinflüsse determiniert ist, bliebe offen, welche pädagogischen Handlungsspielräume damit eröffnet würden. Die Gleichsetzungen "genetisch bedingt = schwer beeinflussbar" und "umweltbedingt = leicht beeinflussbar" gelten nur eingeschränkt. Organische Mängel, die das Lernen erschweren, können auf genetischen Faktoren beruhen, aber auch die Folge von Umwelteinwirkungen sein. Es kommt hinzu, daß die Wirksamkeit pädagogischer Behandlung vom Lebensalter als

einer wichtigen Moderatorvariablen abhängen kann. Offenbar gibt es auch in der menschlichen Entwicklung so etwas wie "sensible Phasen", in denen stabile Verhaltensmuster leichter als nachher oder vorher erworben werden können (z.B. Urvertrauen, Spracherwerb, Sozialverhalten).

Mit Vorbehalt lassen sich die verschiedenen Verhaltensbereiche allenfalls in eine sehr grobe, komplementäre Rangfolge ihrer Determiniertheit durch Anlage und Reifung bzw. durch Umwelteinflüsse und Lernen bringen. An deren einem Ende mit den höchsten genetischen Anteilen und relativ geringer individueller Variabilität befinden sich die Reflexe und das Instinktverhalten, abnehmend über die Psychomotorik, die Intelligenz und Persönlichkeitsmerkmale i.e.S. bis hin zum anderen Ende mit Einstellungen, Gewohnheiten, Meinungen und z.B. der Konfessionszugehörigkeit, bei denen es unmittelbar einleuchtet, daß hier die Umwelteinflüsse eine deutlich größere Rolle spielen.

1.4.1.3 Kollektiv und Individuum

In diesem Zusammenhang ist an ein weiteres, analoges Problem zu erinnern, das die Diagnostik direkt betrifft. In der Praxis beruhen viele individualdiagnostische Urteile, vor allem Verhaltensvorhersagen, auf **ideographischen Rückschlüssen** aus Datenverhältnissen, die theoretisch wie empirisch für Kollektive gelten. Wird z.B. die Empfehlung, ein schulpflichtig gewordenes Kind besser noch nicht einzuschulen, u.a. auf das schwache Abschneiden in einem Schuleingangstest gestützt, ist unbekannt, ob dieses Kind zu denen gehört, für die die Prognose aufgrund der substantiellen Korrelation zwischen Testergebnis und Schulerfolg zutrifft, oder ob es der Gruppe von Kindern angehört, für die sich - aufgrund welcher Randbedingungen auch immer - die Vorhersage nachträglich als falsch erweisen würde. Da wir empirisch nie mit perfekter Abhängigkeit rechnen können, hat dies zur Folge, daß wir grundsätzlich nicht wissen, ob die Vorhersage im Einzelfall zutrifft oder nicht. Die Kennwerte der Kollektive haben lediglich die Funktion von Erwartungswerten für eine Vielzahl von Einzelfällen. Dieses Problem ist bislang nicht befriedigend zu lösen, auch nicht, oder nur sehr bedingt, über individuelle Meßwiederholungen im Sinne der psychometrischen Einzelfalldiagnostik (s. Abschnitt 9.1). Wir müssen uns mit dem Nachweis begnügen, daß wir je nach gegebener Datenlage, insbesondere nach Maßgabe der Korrelation zwischen den Variablen, bei einer größeren Zahl von Urteilen insgesamt weniger Fehler begehen, als dies bei Anwendung anderer verfügbarer Entscheidungsstrategien der Fall wäre. Dies ist eine der unvermeidlichen Konsequenzen aus der erwähnten Beschränkung unseres Erkenntnispielraums.

Auch diese allgemeinen Charakteristika psychologischer Merkmalszusammenhänge sind bei der Formulierung diagnostischer Befunde zu beachten. Unabhängig von ihrem Verwertungszweck sind diagnostische Aussagen in der Regel zunächst deskriptive Feststellungen, die sich auf den Ist-Zustand von Merkmalsausprägungen und deren Verknüpfung an Individuen beziehen. Wie präzise unsere Aussagen sein können und welche Schlußfolgerungen sich daraus ziehen lassen, hängt von der Qualität der Daten, d.h. von der Beschaffenheit der benutzten Verfahren und von den Untersuchungsbedingungen ab.

1.4.1.4 Diagnostische Konstrukte

In diesem Zusammenhang wird deutlich, daß psychologische Merkmale nichts unmittelbar Gegebenes sind. Ein Merkmal ist in jedem Fall die **begriffliche Fassung** eines Aspekts, in dem sich Individuen voneinander und zu verschiedenen Zeiten unterscheiden können. Das begriffliche Abstraktionsniveau kann dabei zwischen der Wiedergabe eines unmittelbar beobachteten Verhaltens, z.B. sich-Melden im Deutschunterricht am Freitag letzter Woche, und erschlossenen theoretischen Konstrukten, wie "Lerneifer", "Selbstbewußtsein", und "Geltungssucht", variieren. Den unterschiedlichen Aussageebenen entsprechen Unterschiede im theoretischen Status der Aussagen. Je höher das Abstraktionsniveau, desto mehr Implikationen sind darin enthalten.

Konstrukte sind hier als eine Art sprachlicher Kürzel zu verstehen, die inhaltlich definierte Bereiche (grundsätzlich) beobachtbaren Verhaltens von Individuen zusammenfassen. Diagnostische Aussagen auf Konstruktebene stellen individualisierte Dispositionsprädikate dar (Herrmann, 1973). Unter **Disposition** wird die Bereitschaft zu bestimmten Handlungen bzw. eine (genügend große) Wahrscheinlichkeit für das Auftreten einer bestimmten Klasse von Verhaltensweisen verstanden. So impliziert z.B. die Feststellung, "Steffi ist hoch leistungsmotiviert", daß Steffi - unabhängig von der Qualität ihrer Leistungen - seit einiger Zeit bemüht ist und voraussichtlich auch weiterhin bemüht sein wird, bei einer Vielzahl verschiedener Leistungsanforderungen jeweils "ihr Bestes zu geben". Solche Aussagen sind nur dann gerechtfertigt, wenn das betreffende Konstrukt, hier die überdauernde Leistungsmotivation, als hinreichend gesichert gelten kann, und wenn sie sich auf Meßoperationen stützen, von denen vorgängig gezeigt worden ist, daß sie **konstitutiv** für das Konstrukt sind. Konstrukte gelten in dem Maß als empirisch gesichert, in dem unterscheidbare Ansätze zu ihrer operationalen Realisierung für dieselben Meßwertträger-Kollektive zu konkordanten Ergebnissen führen, im Idealfall, wenn sie sich als "methodeninvariant" erweisen. In der psychologischen Diagnostik ist dies bisher für das Konstrukt "Intelligenz" am vergleichsweise besten gelungen.

Wir unterscheiden demnach die hypothesengeleitete **induktive Gewinnung** der Konstrukte in der differentiell-psychologischen und diagnostischen Grundlagenforschung von ihrer **deduktiven Verwendung** in der diagnostischen Praxis. Es liegt auf der Hand, daß deren Ergebnisse insgesamt nicht besser ausfallen können, als es dem Konsolidierungs-Status der Konstrukte entspricht. **Die Sicherung und die sachgerechte Handhabung von Konstrukten ist ein Kernstück professioneller Diagnostik.** Für die Praxis ist es dabei nicht entscheidend, ob ein Konstrukt restlos auf empirisch beobachtetes Verhalten reduziert werden kann. Eine solche Forderung strikt einzulösen, ist offenbar nicht möglich, aber auch nicht nötig. Konstrukte können durchaus einen undeckbaren Bedeutungsüberschuß enthalten, jedenfalls solange und soweit sich mit ihren diagnostischen Ableitungen befriedigende Resultate erzielen lassen, d.h. solange und soweit sich damit Verhaltensvorhersagen nachweislich verbessern lassen.

Es wäre allerdings ein Mißverständnis, wenn man Konstrukte dieser Art als verursachende Instanzen im Sinne latenter Verhaltensdeterminanten interpretieren wollte. Die Aussage "Thomas kann gut denken, weil er überdurchschnittlich intelligent ist", wäre eine bloß schein-kausale Verknüpfung tautologischer Argumente, denn Intelligenz ist durch Denken-Können definiert. Doch ist es vertretbar, Konstrukte in quasi-explikativer Weise zu benutzen. "Da wir bei Thomas (u.a.) einen hohen IQ ermittelt haben, erwarten wir, daß er im Mathematikunterricht gut zurechtkommt". Eine sol-

che Aussage stützt sich auf die Tatsache, daß die Wahrscheinlichkeit des Erfolges im Mathematikunterricht (unter sonst gleichen Bedingungen) deutlich mit dem IQ zunimmt. Auch hierbei greifen wir lediglich auf eine "Wenn-Dann"-Beziehung (\rightarrow) im Sinne eines deskriptiv-empirischen Zusammenhangs zwischen operationalen Repräsentanten der Konstrukte Intelligenz und Schulerfolg zurück; seine Enge bemißt sich an der Höhe eines Korrelationskoeffizienten.

Ungeachtet seiner wissenschaftstheoretischen Belastung stellt das hier skizzierte Verständnis von Konstrukten einen (vorläufig) gangbaren Weg dar, in der Diagnostik Merkmale zu definieren und zu präzisieren, ohne daß dabei das dialektische Verhältnis von theoretischem und operationalem Zugang aus dem Blickfeld gerät. Damit steht uns - unabhängig von ihrer theoretischen und methodischen Ausgestaltung im einzelnen - eine tragfähige allgemeine Basis für eine theoriegeleitete und zugleich handlungsorientierte Diagnostik zur Verfügung.

Ein Beispiel für die Beziehung zwischen Verhaltensmerkmalen und Konstrukten ist in der Abbildung 1 (S. 29) wiedergegeben. Es verdeutlicht die verschiedenen Aussageebenen und den damit zunehmenden Abstraktionsgrad der Merkmale. Zugleich wird die Komplexität von Stufe zu Stufe größer. Dies beruht in diesem Ordnungsansatz auf der empirischen Korrelation zwischen den Merkmalen. Zusammenfassungen dieser Art sind möglich und diagnostisch zweckmäßig, soweit gezeigt werden kann, daß die einbezogenen Merkmale miteinander systematisch höher als mit anderen Merkmalen korrelieren. Das Beispiel orientiert sich an dem deskriptiv-hierarchischen Persönlichkeitsmodell von Eysenck (z.B. 1975) und stützt sich auf Konstruktionsdaten des Persönlichkeitsfragebogens für Kinder (PFK 9-14; Seitz & Rausche, 1976, 1992).

Weitere Möglichkeiten zur Bildung komplexer diagnostischer Klassen bestehen in der Gruppierung von Personen nach der Ähnlichkeit ihrer Merkmalsprofile. Die Gruppen werden dabei so zusammengefaßt, daß die Mitglieder einer Gruppe sich möglichst ähnlich, die Gruppen untereinander möglichst unähnlich sind (Clusteranalyse; s. Abschnitt 4.3.). Dies spielt z.B. bei der Interaktion von Unterrichtsmethode und Schülertyp eine Rolle ("Aptitude-Treatment-Interaction", ATI; vgl. Cronbach & Snow, 1977).

1.4.1.5 Person, Situation und aktuelle Befindlichkeit

Da Gleichungen vom Typ der Formel [1.2] (s.S. 25) vorläufig nicht zu realisieren sind und dies für eine brauchbare Diagnostik auch nicht nötig ist, betrachten wir Verhalten unscharf, aber heuristisch vertretbar, als eine Resultante aus Personmerkmalen (P) und situativen Umweltbedingungen (S):

$$[1.3] \quad v = f(P, S).$$

Die längerfristigen Umwelteinwirkungen sind hier in den Ist-Zustand der Personmerkmale eingegangen.

Analog zu der in [1.2] angegebenen Varianzzerlegung setzt sich die theoretische Varianz des Verhaltens aus Anteilen zusammen, die auf Unterschiede zwischen den Personen (P) und zwischen den Situationen (S) sowie auf die Kovarianz (Cov P, S) und die Wechselwirkung (P · S) dieser Komponenten zurückgehen.

Bei diesem gelegentlich auch "interaktionistisch" genannten Ansatz ist zu beachten, daß die Merkmalsklassen (P) und (S) - abgesehen von ihrer Kovarianz und Wech-

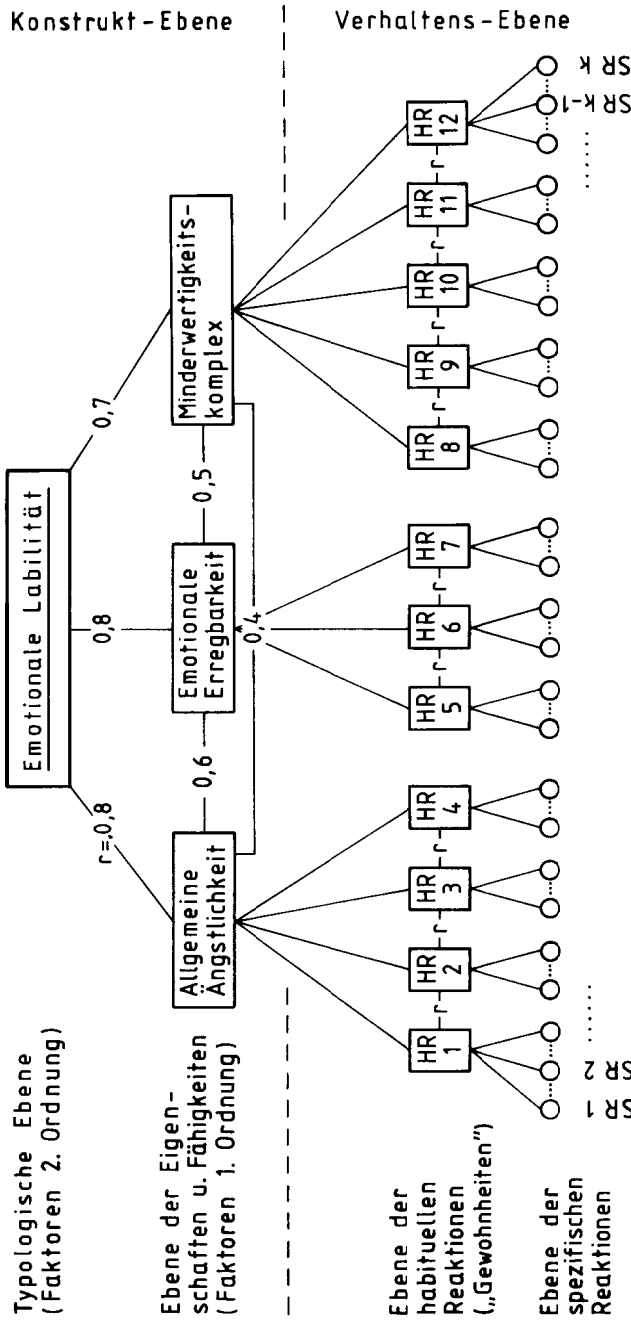


Abbildung 1.1: Konstrukt "Emotionalität" bei Kindern (nach dem deskriptiv-hierarchischen Persönlichkeitsmodell von Eysenck und Konstruktionsdaten des PFK 9-14 von Seitz & Rausche). Bedeutung der habituellen Reaktionen (HR):

- HR 1: sich vor körperlicher Verletzung fürchten
- HR 2: sich vor Dunkelheit und Alleinsein fürchten
- HR 3: sich vor Bedrohung fürchten
- HR 4: Angstträume haben
- HR 5: leicht irritiert und frustriert sein
- HR 6: unruhig und ungeduldig sein
- HR 7: nervös sein und zu versagen fürchten
- HR 8: andere für hübscher halten
- HR 9: andere für fähiger halten
- HR 10: sich nach anderen richten
- HR 11: sich verspottet fühlen
- HR 12: darunter leiden, manches nicht zu können

selwirkung - schon von vornherein nicht unabhängig voneinander variieren. Zwar können wir im allgemeinen Ereignisse und Sachverhalte, die außerhalb einer Person vorkommen, eindeutig von Merkmalen unterscheiden, die unmittelbar oder mittelbar einer Person zuzuschreiben sind. Doch enthält "Situation" nach unserer Definition von Umwelt Anteile von Personvarianz. Die Unterscheidung von Person und Umwelt ist theoretisch notwendig; für das praktische Vorgehen bleibt sie fiktiv, weil Person und Umwelt psychologisch nur quasi-unabhängige Merkmalsklassen darstellen.

So wie sich Umwelteinflüsse nach kürzer- oder längerfristigen, situationsspezifischen oder situationsübergreifenden unterscheiden lassen, unterscheiden wir relativ stabile "habituelle" Personenmerkmale (P_{stab} ; Eigenschaften, Fähigkeiten, "Persönlichkeitsmerkmale" i.e.S.) von Merkmalen der aktuellen Befindlichkeit (P_{akt}), die die stabileren Merkmale überlagern. Unter der Voraussetzung, daß die Wechselwirkungskomponente mit den übrigen Anteilen nicht korreliert, ist daher die theoretische Varianz des Ausdrucks P in Formel [1.3] definiert als

$$[1.4] \quad \text{Var}(P) = \text{Var}(P_{\text{stab}}) + \text{Var}(P_{\text{akt}}) + 2\text{Cov}(P_{\text{stab}}, P_{\text{akt}}) + \text{Var}(P_{\text{stab}} \cdot P_{\text{akt}}).$$

Für ein einzelnes Merkmal X gilt, daß es sowohl von konstruktverwandten Merkmalen (im Sinne von Abb. 1; s.S. 29) als auch von konstruktfernden überlagert sein kann, z.B. die kognitive Leistungsfähigkeit von der aktuellen Konfliktbelastung oder vom augenblicklichen Gesundheitszustand. Eine Kovarianz z.B. von Leistungsfähigkeit (stabile Komponente) und Konfliktbelastung (aktuelle Komponente) läge vor, wenn sich die Leistungsschwächeren häufiger in leistungsmindernden Konfliktsituationen befänden als die Leistungsstärkeren. Von Wechselwirkung zwischen Leistungsfähigkeit und Konfliktbelastung wäre zu sprechen, wenn die tatsächlich gezeigte Leistung nur bei den habituell Leistungsschwachen oder deutlich stärker bei ihnen als bei den habituell Leistungsstarken beeinträchtigt wäre.

Damit ist angedeutet, daß wir menschliches Verhalten als ein hochkomplexes Geflecht von Merkmalen zu verstehen haben, dessen systematische Untersuchung Aufgabe der Persönlichkeitsforschung und der Differentiellen Psychologie ist (Amelang & Bartussek, 1990). Weitere Aspekte, durch die sich Merkmale funktional voneinander abheben, sind der Tabelle 1 (s.S. 24/25) zu entnehmen. Darauf wird später noch eingegangen.

1.4.2 Präzisierung der Meßoperationen

1.4.2.1 Standardisierung, Ökonomisierung und Meßgenauigkeit

Die **grundlegenden diagnostischen Prinzipien** bestehen darin, durch Standardisierung der Untersuchungsbedingungen **erstens** die Situationseinflüsse (S) möglichst konstant zu halten, so daß das registrierte Verhalten theoretisch allein als Funktion der Personenmerkmale (P) verstanden werden kann. Für S = konstant gilt:

$$[1.5] \quad v = f(P).$$

Zweitens wird durch eine repräsentative Auswahl der Aufgaben, die den Pbn gestellt werden, der Reaktionsbereich inhaltlich so spezifiziert, daß sich die registrierte

Verhaltensstichprobe einem bestimmten Merkmal(skomplex) X_j zuordnen läßt. Für $S = \text{konstant}$ und inhaltlich spezifiziert gilt:

$$[1.6] \quad V = f(X_j),$$

wobei (X_j) Element der Menge (P) aller Merkmale ist, die wir an Personen unterscheiden können.

Das erste Standardisierungsprinzip entspricht der Forderung nach **Objektivität** des Vorgehens. Die Verfahren sollen vor allem durch eindeutige, verbindliche Vorschriften ("Instruktionen") für die Durchführung und die Auswertung sowie für die Interpretation der Ergebnisse gewährleisten, daß der diagnostische Befund so **wenig wie möglich** von den **äußeren** Umständen abhängt, unter denen die Untersuchung stattgefunden hat. Deren Anteil an der empirischen Varianz der Meßwerte soll gegen null gehen. Zu den äußeren Umständen zählen auch Untersucher und Auswerter (vgl. Abschnitt 2.2).

Auf dem zweiten Standardisierungsprinzip beruht die **Validität** (diagnostische Gültigkeit) der Methoden. Sie liegt empirisch zunächst in dem Maße vor, wie es bei der Konstruktion eines Verfahrens gelingt, den theoretisch bestimmten Merkmalsbereich, auf den es sich richten soll, tatsächlich abzudecken. Dies gilt für Klassen- oder Prüfungsarbeiten grundsätzlich ebenso wie für formelle Testverfahren. Unsere Erhebungsinstrumente sind jeweils **operationale Definitionen** der Merkmale, die uns diagnostisch interessieren. So kann z.B. die Intelligenz nach dieser oder jener **Intelligenztheorie** als der Grad der kognitiven Leistungsfähigkeit eines Individuums definiert werden; **diagnostisch ist sie durch das definiert, was der Test mißt** (Boring, 1923, zit. nach Conrad, 1983, S. 107; vgl. Abschnitte 1.4.3 und 2.2). Mit der Standardisierung dienen die beiden genannten Prinzipien zugleich der **Ökonomisierung**; d.h. in der professionellen Diagnostik werden die Daten nicht gelegentlich oder zufällig, sondern unter möglichst **einheitlichen Bedingungen** möglichst **systematisch** und **treffsicher** erhoben. Wie beim Experiment geht es um künstlich herbeigeführte, kontrollierte und grundsätzlich wiederholbare Verhaltensbeobachtung; anders als beim Experiment mit seinem Prinzip der Bedingungsvariation gilt hier der Grundsatz der **Bedingungskonstanz**.

Je ähnlicher ("homogener") die Anforderungen der Aufgaben oder Fragen und je geringer der Einfluß der äußeren Umstände, desto genauer ("reliabler") wird das betreffende Personmerkmal gemessen. Die Varianz der Meßergebnisse soll möglichst vollständig auf systematische Unterschiede in der Merkmalsausprägung bei den untersuchten Personen ("Meßwertträgern") zurückgehen.

In dem Maße, in dem die so entstandene Reihung der Personen bei einer späteren Meßwiederholung erhalten bleibt, kann man von einem "stabilen" Merkmal sprechen. Ein Merkmal ist theoretisch umso stabiler, je näher die Korrelation zwischen erster und zweiter Messung an die Meßgenauigkeit des Verfahrens herankommt und je länger die Messungen auseinanderliegen. Empirisch bleibt die feststellbare Stabilität eines Merkmals auf die Meßgenauigkeit des diagnostischen Verfahrens beschränkt. "Stabil" bezieht sich hier auf die psychologische oder pädagogische Bedeutung des Merkmals, unabhängig davon, wie stark sich inzwischen die Meßwerte **absolut** verändert haben. Dies spielt z.B. eine Rolle, wenn es um die Messung von Lernfortschritten geht und daraus auf unterschiedliche "Fähigkeiten" der Schüler geschlossen werden soll. Schlüsse dieser Art werden in der Regel gezogen, wenn **längerfristige** Vorhersagen ("Prognosen") gefordert sind, z.B. bei der Einschulungs- oder Eignungs-

diagnostik. Dabei geht man auf die Konstruktebene über, weil Vorherzusagendes Merkmal (“Kriterium”) und Ausgangsmerkmal (“Prädiktor”) phänotypisch verschieden sein können, wie das z.B. bei der Verwendung von Intelligenztests für die Diagnose der “Sonderschulbedürftigkeit” oder die Prognose des Berufserfolgs der Fall ist. Die Stabilität der Merkmale und der Randbedingungen ist Voraussetzung für Vorhersagen. Die erreichbare Güte einer Vorhersage wird begrenzt durch das empirisch ermittelte Maß an Stabilität der beteiligten Merkmale.

1.4.2.2 Vergleichsmaßstäbe

Diagnostische Aussagen kennzeichnen die individuelle Ausprägung eines Merkmals an einer Person. Angenommen, die Schülerin Gabi habe bei einer Reihe von Rechtschreibaufgaben 16 Punkte erhalten. Wie ist diese Leistung zu beurteilen? Die bloße Mitteilung eines solchen Wertes läßt nicht erkennen, was er bedeutet. Dazu sind offenbar weitere Informationen nötig, z.B. welches ist die höchste erreichbare oder die höchste erreichte Punktzahl? Wie ist die Skala definiert? Darf man annehmen, daß jemand mit 8 Punkten nur “halb so gut” in Rechtschreiben ist? (Oder “doppelt so gut”, falls nämlich Fehler gezählt wurden?) Von der Definition und der Qualität der Skala abgesehen (s. Tabelle 1), brauchen wir Bezugsgrößen, die es gestatten, den individuellen Wert auf dem angenommenen Merkmalskontinuum zu lokalisieren. Dies kann auf verschiedene Weise geschehen. Wir können den Wert z.B. auf die Verteilung aller Werte beziehen, die von Schülern desselben Alters oder derselben Schulstufe erreicht werden, und angeben, ob er über oder unter einem ausgezeichneten Kennwert, etwa dem Median oder einer anderen Marke, liegt, und wie weit er davon entfernt ist. Unabhängig davon können wir uns u.U. mit der Feststellung begnügen, ob z.B. vor Beginn einer neuen Unterrichtseinheit die nötigen Mindestanforderungen bei den Schülern erfüllt sind, oder ob ein bestimmter Sollwert erreicht ist, der uns erlaubt anzunehmen, daß die Schüler die betreffende Fertigkeit inzwischen hinreichend sicher beherrschen.

Ganz gleich, ob ein erhobener Istwert für pädagogisch befriedigend gehalten oder als veränderungsbedürftig betrachtet wird, in jedem Fall bedarf es dazu verlässlicher Orientierungsgrößen (“Normen”). Anders wäre die Bedeutung einer diagnostischen Information nicht einzuschätzen; sie bliebe wertlos. Dies gilt erst recht, wenn z.B. ermittelt werden soll, wie “beträchtlich” die Ausfälle im Leistungsspektrum eines Schülers sind, und zu klären ist, ob sie auf mangelhafte “Beschulung”, auf starke psychische Belastung (z.B. durch Ängste) oder auf Motivationsstörungen zurückgehen, bzw. ob der Schüler als “lernbehindert” im Sinne von “sonderschulbedürftig” gelten muß. Ebenso benötigen wir Normen, wenn es um das Erkennen besonderer “Begabungen” oder um die Feststellung geht, die Leistungen eines Schülers seien “durchschnittlich” und sein Verhalten “unauffällig”: ganz allgemein, wenn Leistungen und Verhalten intra- oder interindividuell verglichen werden sollen, sei es zu einem bestimmten Zeitpunkt, sei es, daß uns Veränderungen von einem zum anderen Zeitpunkt interessieren. Als Orientierungsgrößen können gesetzte **Sollvorgaben** (“Gabi hat in Mathematik das Klassenziel nicht erreicht”), individuelle **Bezugsnormen** (“Gabi hat in Deutsch-Schriftlich erhebliche Fortschritte gemacht”) oder **Gruppennormen** (“Im Englischen gehört Gabi zum besten Viertel ihres Jahrgangs”). Häufig ist es zweckmäßig, für Vergleiche zwischen Individuen und Merkmalen einheitliche Skalen zu benutzen.

Normen und die Handhabung von Normen, zumal in der Pädagogisch-psychologischen Diagnostik, sind nichts von Natur aus Gegebenes. Sie hängen von kulturellen und gesellschaftlichen Bedingungen ab und beruhen größtenteils auf Vereinbarungen, z.B. darüber, was, wann und wie in Schulen unterrichtet werden soll, oder welche Zulassungsbedingungen für den Besuch von Sonderschulen oder Universitäten gelten. Dementsprechend unterliegen sie dem Wandel, und sie sind grundsätzlich revidierbar. Dies trifft selbst für das zu, was man in einer Gesellschaft ungeachtet interkultureller Gemeinsamkeiten - unter Intelligenz versteht. Normen können mehr oder weniger engmaschig sein. Die bei uns übliche Skala für Schulnoten läßt fünf oder sechs, mit den manchmal vergegebenen Zwischennoten zehn bis zwölf Abstufungen zu. Intelligenzquotienten (IQ) sind zwei- bzw. dreistellig definiert und erwecken den Eindruck, man könne in der Gesamtbevölkerung mindestens 90 Ausprägungsgrade der kognitiven Leistungsfähigkeit unterscheiden (IQ zwischen 55 und 145). Wie eng das Raster von Normen sein darf, hängt in erster Linie von der Genauigkeit ab, mit der das abgebildete Merkmal gemessen wird, also von der Reliabilität des diagnostischen Verfahrens, bzw. der Stabilität des Merkmals. Sind diese gering, kann selbst eine einstellige Normenskala eine Differenzierung vortäuschen, die wegen der mangelnden Qualität der Meßoperation nicht gerechtfertigt ist.

In der Praxis kommt es vielfach nicht auf maximale Differenzierung, bzw. die maximal mögliche Meßgenauigkeit an. Häufig genügen Unterscheidungen wie "versetzt"/"nicht versetzt", "durchschnittlich", bzw. "unter-" oder "überdurchschnittlich" oder die Feststellung, daß die große Mehrheit der Schüler das Unterrichtsziel erreicht hat, ohne daß für jeden einzelnen nachgeprüft werden müßte, in welchem Maße sich seine Leistungen z.B. von denen der Mitschüler unterscheiden. Andererseits wird - im Zusammenhang mit der Handhabung von Grundgesetzartikeln, die die Freiheit der Berufswahl garantieren-für die Abschlußzeugnisse von Gymnasien eine ausgefeilte Arithmetik vorgeschrieben. Sie soll die schulische Gesamtleistung jedes Schülers auf einem Raster von 31 zulässigen Skalenwerten lokalisieren und damit eine feine Abstufung der kritischen Mindestwerte für die Zulassung zu bestimmten Studiengängen ermöglichen. Auch hier ist u.a. zu fragen, ob die Meßgenauigkeit der Diagnostik ausreicht, um die "Befunde" so stark zu differenzieren. Ist die Leistungsfähigkeit von Bewerbern, die z.B. mit der Note 1,9 den kritischen Wert nicht erreichen, tatsächlich geringer als die der anderen, die mit 1,8 zugelassen werden?

Allgemein gilt jedoch der Grundsatz, daß Normen pädagogisch umso ergiebiger genutzt werden können, je stärker sie zwischen den Ausprägungsgraden eines Merkmals zu differenzieren gestatten, vorausgesetzt, die diagnostischen Verfahren, auf denen sie beruhen, sind entsprechend meßgenau. Dies interessiert uns natürlich nicht bei beliebigen Merkmalen sondern nur bei solchen, von denen gezeigt werden kann, daß sie für Erziehung und Unterricht bedeutungsvoll sind, und worin diese Bedeutung besteht.

1.4.3 Verifizierung diagnostischer Aussagen

Diagnostische Aussagen beschreiben die individuelle Ausprägung von Merkmalen, auf denen sich Personen unterscheiden können. Wie alle wissenschaftlichen Aussagen über empirische Sachverhalte müssen die Aussagen der professionellen Diagnostik überprüfbar sein. Sie sollen nicht nur objektiv und hinreichend präzise sein, sie

müssen sich auch bewähren, d.h. sie müssen nachweislich und möglichst vollständig zutreffen. Erst damit wird die Diagnostik ihrer Funktion gerecht, zur Optimierung pädagogischer Entscheidungen beizutragen. Die diagnostischen Verfahren sind also darauf zu untersuchen, wieweit sie diesem Anspruch genügen. Auf die Verfahren bezogen, sprechen wir - wie bereits erwähnt - von deren **Validität** oder **Gültigkeit**. Damit ist das Ausmaß gemeint, in dem etwa ein Test für den Zweck, zu dem er verwendet werden soll, tatsächlich brauchbar ist; z.B. wie gut ein Schuleingangstest als Prädiktor das Kriterium Schulerfolg vorherzusagen gestattet, wenn die Kinder ihren Lernvoraussetzungen entsprechend gefördert werden.

An die Validität der Verfahren sind umso höhere Ansprüche zu stellen, je gewichtiger die zu treffende Entscheidung ist. Wo es entsprechend gute Verfahren (noch) nicht gibt, muß die daraus resultierende Unsicherheit berücksichtigt werden; d.h. die Randbedingungen und die wahrscheinlichen Konsequenzen alternativer Entscheidungen sind so sorgfältig wie möglich abzuwägen. Bleibt die empirische Fehlerquote auch bei Nutzung aller verfügbarer Prädiktoren hoch, sind u.U. die systembedingten Entscheidungszwänge zu revidieren. Dies betrifft z.B. die vom traditionellen westdeutschen Schulsystem geforderten Übergangsentscheidungen nach dem vierten Grundschuljahr (vgl. Tent, 1969). Der im vorigen Abschnitt angeführte Zugang zum Studium von Numerus-Clausus-Fächern ist ein anderes Beispiel für die Frage nach der empirischen Legitimation staatlicher Regelungsbefugnisse. Hierbei geht es hauptsächlich um einen Aspekt der Validität von Lehrerurteilen. Den Inhabern von Reifezeugnissen wird die unbefristete Eignung und Berechtigung bescheinigt, beliebige Fächer an wissenschaftlichen Hochschulen studieren zu können. Wieweit ist die scheinbar plausible Annahme gerechtfertigt, daß sich Abiturienten für bestimmte, zulassungsbeschränkte Studiengänge umso eher eignen, **je** besser die **Durchschnittsnote** ihres Zeugnisses ist? Und haben Absolventen mit besseren Prüfungsergebnissen auch mehr Erfolg im Beruf?

In diesen Beispielen ergibt sich die Validität der diagnostischen Verfahren aus dem Verwertungszusammenhang. Wir sprechen dann von **Kriteriumsvalidität** und von **prognostischer Validität**. Unter den Rahmenbedingungen unseres Schulsystems spielt dieser Validitätsaspekt eine unverhältnismäßig große, wenn auch inzwischen abnehmende Rolle.

Die Verfahren können aber unabhängig von ihrer aktuellen Verwertung auf ihre pädagogische oder psychologische Bedeutung überprüft werden. Welches Merkmal, oder welche Merkmalskombination, wird erfaßt? Geben z.B. die Deutschnoten tatsächlich nur die Leistung der Schüler im Deutschunterricht wieder oder gehen vielleicht das "Betragen" oder die Sympathie/Antipathie auf seiten des Lehrers mit ein? Welche Komponenten der kognitiven Leistungsfähigkeit sind in einem Intelligenztest berücksichtigt, und wie groß sind eventuell die Anteile von Motivation und Konzentration? Wieweit beeinflußt die Neigung, sozial erwünscht zu reagieren, die Ergebnisse eines Angstinventars oder Persönlichkeitsfragebogens? Dabei interessiert in erster Linie, wieweit die empirischen Daten mit theoretisch vorgegebenen Merkmalskonzepten (wie "Intelligenz" oder "Labilität") in Einklang stehen. Man spricht in diesen Fällen (unscharf) von **Konstruktvalidität**. Was ein konstruktvalider Test mißt, kann je nach diagnostischer Fragestellung und Vergleichsgröße (Kriterium) verschieden belangvoll sein. Ein Intelligenztest, der z.B. hochsich mit der objektiven Schulleistung in Mathematik korreliert, kann u.U. Zeugnisnoten nur mäßig genau und die Ergebnisse mündlicher Prüfungen noch weniger genau vorhersagen. Obwohl alle drei

Variablen mit Intelligenz zu tun haben, ist das Ausmaß unterschiedlich; dieses hängt u.a. von der instrumentellen Qualität des Kriteriums ab. Es ist also nicht sinnvoll, von **der** Validität eines Verfahrens zu sprechen; vielmehr gibt es je nach Verwendungszweck eine Mehrzahl unterscheidungsbedürftiger **Validitätsaspekte**.

In der Pädagogisch-psychologischen Diagnostik spielt darüber hinaus die **Lehrplangültigkeit** oder **curriculare Validität** eine besondere Rolle. Hier geht es um den (meist über Expertenurteile erbrachten) Nachweis, daß die Aufgaben in Schulleistungstests für die Lehrplananforderungen eines zeitlichen Ausschnitts aus einem Unterrichtsfach repräsentativ sind. So muß z.B. ein curricular valider Rechentest für das vierte Schuljahr genau die Typen von Aufgaben enthalten, die vom Lehrplan für den Mathematikunterricht auf dieser Schulstufe vorgesehen sind, also schriftliches Multiplizieren und Dividieren, Kopfrechnen und Textaufgaben ("Rechnerisches Denken" in Sachzusammenhängen). Aber auch jede Klassenarbeit muß selbstverständlich, meist für einen kleineren Ausschnitt, "lehrplangültig" sein. Zu beachten ist, daß die Leistungen der Schüler immer auch von der Güte des erteilten Unterrichts mitbestimmt werden. Die **Rückmeldungsfunktion** solcher diagnostischer Erhebungen gilt gleichermaßen der Schule wie den Schülern.

Ganz gleich wie die Validität eines diagnostischen Verfahrens bestimmt wird, es geht jeweils um die Aufklärung der Varianz uns interessierender Personmerkmale, und zwar unabhängig davon, ob wir es mit relativ stabilen oder weniger stabilen Merkmalen zu tun haben. Die Validität ist das wichtigste Gütekriterium aller Diagnostik. Unsere Aussagen sollen so valide sein wie möglich, d.h. die Unterschiede, die wir feststellen, sollen so genau wie möglich zutreffen. Doch können auch weniger valide Verfahren nützlich sein. Ihre Anwendung ist gerechtfertigt, solange keine nachweislich besseren zur Verfügung stehen. Wie schon angeführt, müssen wir stets bedenken, **wie valide** ein Verfahren für den Zweck ist, zu dem wir es benutzen.

1.5 Zusammenfassung und Definition von Diagnostik

Verhalten, Leistungen, Eigenschaften und Fähigkeiten von Personen zu beurteilen, ist uns aus dem alltäglichen zwischenmenschlichen Umgang von früh an vertraut. Unsere Urteile zielen darauf ab, den anderen möglichst gut zu verstehen und einzuschätzen, was wir von ihm zu erwarten haben. Dies ist Teil unserer Bemühungen, mit Hilfe bestandsfester Erkenntnisse die Lebenswelt, in der wir agieren, überschaubar zu machen und künftige Ereignisse weniger ungewiß erscheinen zu lassen. Aufgabe der Pädagogisch-psychologischen Diagnostik ist es, dies für den Lebensbereich zu leisten, den wir Erziehung nennen. Erziehen heißt, Merkmale von Personen über mentale Beeinflussung möglichst dauerhaft zu verändern. Differenzierung und Individualisierung sind anerkannte Grundsätze pädagogischen Vorgehens.

Die professionelle Diagnostik dient der Verwirklichung dieser Grundsätze. Sie folgt damit dem allgemeinen Optimierungsgebot, das auch für pädagogisches Handeln gilt. Dabei knüpft sie an die Alltagsdiagnostik an. Ihr wissenschaftliches Fundament erhält sie durch die Klärung ihrer persönlichkeits- und meßtheoretischen Annahmen, durch die Präzisierung der Merkmale, auf die sie sich richtet, durch die genaue Analyse der Randbedingungen, unter denen sie abläuft, durch die Standardisierung und Präzisierung der Meßoperationen (der diagnostischen Erhebungsmethoden),

durch die Bereitstellung von Maßstäben zur Beurteilung der individuellen Meßergebnisse sowie durch die empirische Verifizierung ihrer diagnostischen Aussagen.

Als **Praxis** ist Diagnostik in der Regel "problemlösendes Handeln" im Sinne der Anwendung einer grundsätzlich "nutzenmaximierenden Technologie auf wissenschaftlicher Grundlage" (Wottawa & Hossiep, 1987). Im Umfeld der Erziehung geht es primär um einen jeweils pädagogisch definierten Nutzen, worin dieser auch immer bestehen mag. Je nach Fragestellung sind demzufolge Methoden zu verwenden, die den Ansprüchen einer wissenschaftlich fundierten Diagnostik genügen und deren Güte der Tragweite der pädagogischen Schlußfolgerungen entspricht, die man darauf stützen will.

Damit ist der Sache nach deutlich, was unter Pädagogisch-psychologischer Diagnostik zu verstehen ist, wozu sie dient, wie sie vorgeht, und was wir von ihr erwarten. In der Literatur finden sich zahlreiche, unterschiedlich genaue und umfassende Begriffsbestimmungen (z.B. Klauer, 1978; Michel & Conrad, 1982; Ingenkamp, 1985; Jäger & Petermann, 1992). Es erscheint uns zweckmäßig, Diagnostik, in Anlehnung an Tent & Waldow (1984, S. 5) zusammenfassend wie folgt zu definieren:

Definition für Diagnostik

"Diagnostik ist ein theoretisch begründetes System von Regeln und Methoden zur Gewinnung und Analyse von Kennwerten für inter- und intraindividuelle Merkmalsunterschiede an Personen."

Dazu gehören

- (a) die Formulierung diagnostischer Probleme und Fragestellungen
- (b) die Erhebung diagnostischer Daten und deren Integration **zu Diagnosen** sowie
- (c) die damit verknüpften Folgerwartungen (**Prognosen**) im Hinblick auf verfügbare oder wünschbare Behandlungsalternativen.

Bei den Erhebungsmethoden unterscheidet man die informelle, instrumentell meist schwächere Urteilsbildung durch Experten (z.B. Lehrer, Psychologen, Ärzte) aufgrund Verhaltensbeobachtung, Leistungseinschätzung und Gesprächsführung von der formalisierten Urteilsbildung mit Hilfe standardisierter Untersuchungsverfahren (Inventarien und **Tests**).

Mit dieser Definition sind die in Praxis und Forschung möglichen Fälle der Anwendung diagnostischer Prozeduren und der Verwertung diagnostischer Informationen erschöpfend abgedeckt. Wie alle empirisch-psychologischen Untersuchungen werden diagnostische Erhebungen stets an Individuen vorgenommen; diagnostische Aussagen beziehen sich daher primär auf Einzelpersonen, denen damit bestimmte Attribute zugeschrieben werden. Aus den individuellen Ergebnissen lassen sich je nach der Skalenqualität Kennwerte für Gruppen errechnen, so daß man - vor allem zu Forschungszwecken - z.B. Schulklassen, Schultypen und Schulstufen, Schülerkohorten oder Statusgruppen hinsichtlich bestimmter Merkmale insgesamt kennzeichnen und miteinander vergleichen kann.

Von **Diagnose** sprechen wir in diesem Zusammenhang, wenn Personen anhand relevanter und valider Einzelinformationen innerhalb eines pädagogisch bedeutsamen Klassifikationssystems einer bestimmten Klasse von Merkmalsträgern zugeordnet werden. So kann z.B. die Kombination des Rohwerts auf einem kognitiven Leistungstest mit dem Lebensalter bei einem jüngeren Kind bedeuten, daß es "überdurchschnittlich", derselbe Rohwert bei einem älteren, daß es "unterdurchschnittlich intelligent" ist. Ähnlich fassen wir verschiedene Informationen über die Sinnestüchtigkeit, die Schulleistung, die Intelligenz und die Vorgeschichte eines Schülers zu Diagnosen wie "lese-rechtschreibschwach" oder "lernbehindert" zusammen.

Unter **Prognose** versteht man die Erwartung ("Vorhersage") künftigen Verhaltens oder künftiger Leistungen aufgrund diagnostischer Erkenntnisse. Bezieht sich die Erwartung auf dasselbe Merkmal wie das zuvor diagnostisch erfaßte, ist die Treffsicherheit der Vorhersage eine Funktion der Stabilität des Merkmals; bezieht sich die Erwartung auf ein anderes Merkmal ("Kriterium"), hängt die Treffsicherheit neben der Stabilität des Prädiktors und des Kriteriums von der Enge des empirischen Zusammenhangs zwischen beiden ab. Im pädagogischen Alltag spielen Erwartungen dieser Art eine große Rolle; formalisierte Vorhersagen werden allerdings nur selten genutzt.

Als **diagnostischen Test** bezeichnen wir jedes systematisch konstruierte, routinemäßig anwendbare, standardisierte und normierte **Verfahren zur Erhebung individueller Reaktionsstichproben**, sofern dessen Meßgüte bekannt ist und für den Verwendungszweck ausreicht.- Diese strenge Bestimmung soll im Sinne unserer Diagnostik-Definition die Unterscheidung "weicher" von methodisch anspruchsvollen Verfahren gewährleisten und dem Verschleiß des Test-Begriffs entgegenwirken. Auch wenn sie methodisch hohen Ansprüchen genügen, sind diagnostische Ergebnisse stets deskriptive Aussagen über Ist-Zustände. Für sich genommen, besagen sie in der Regel noch nichts über die zugrundeliegenden "Ursachen". Dazu bedarf es zusätzlicher Analysen. Ebenso wenig ist diagnostischen Aussagen zu entnehmen, weshalb und wie der festgestellte Ist-Zustand geändert werden soll. Für unseren Anwendungsbereich wird dies von den pädagogischen Zielvorgaben und den Möglichkeiten zu ihrer Realisierung bestimmt ("Primat der Didaktik", Tent & Waldow, 1984; Schlee, 1985). Da Erziehung Veränderungen an Personenmerkmalen bewirken soll, kommt der **Veränderungsmessung** ("Verlaufsdagnostik") in der Pädagogisch-psychologischen wie in der klinischen Diagnostik eine besondere Bedeutung zu (s. Abschnitt 9).

Grundlegende Literatur:

Erziehungswissenschaftliche Grundlagen:

- Brezinka, W. (1978). *Metatheorie der Erziehung* (4. Aufl.). München: Reinhardt.
 Sauer, K. (1981). *Einführung in die Theorie der Schule*. Darmstadt: Wiss. Buchgesellschaft.
 Wilhelm, Th. (1977). *Pädagogik der Gegenwart* (5. Aufl.). Stuttgart: Kröner.

Zur Differentiellen Psychologie und Persönlichkeitsforschung:

- Amelang, M. & Bartussek, D. (1990). *Differentielle Psychologie und Persönlichkeitsforschung* (3. Aufl.). Stuttgart: Kohlhammer.

- Herrmann, Th. (1991). *Lehrbuch der empirischen Persönlichkeitsforschung* (6. Aufl.). Göttingen: Hogrefe.
- Hofstätter, PR. (1977). *Persönlichkeitsforschung* (2. Aufl.). Stuttgart: Kröner.
- Mogel, H. (1990). *Umwelt und Persönlichkeit. Bausteine einer psychologischen Umwelttheorie*. Göttingen: Hogrefe.

Zur Psychologischen und Pädagogischen Diagnostik:

- Ingenkamp, K. (1985). *Lehrbuch der Pädagogischen Diagnostik* (Studienausgabe 1988). Weinheim: Beltz.
- Jäger, R.S. & Petermann, F. (Hrsg.) (1992). *Psychologische Diagnostik. Ein Lehrbuch* (2., veränd. Aufl.). Weinheim: Psychologie Verlags Union.
- Klauer, K.J. (Hrsg.) (1978). *Handbuch der Pädagogischen Diagnostik*, Band 1. Düsseldorf: Schwann.
- Kleber, E.W. (1992). *Diagnostik in pädagogischen Handlungsfeldern*. Weinheim: Juventa.
- Stüllwold, F. (1983). Pädagogische Diagnostik. In K.J. Groffmann & L. Michel (Hrsg.), *Intelligenz- und Leistungsdiagnostik* (S. 307-386). Göttingen: Hogrefe.
- Wottawa, H. & Hossiep, R. (1987). *Grundlagen psychologischer Diagnostik*. Göttingen: Hogrefe.

Weiterführende Literatur zur Pädagogisch-psychologischen Diagnostik:

- Ingenkamp, K. (1990). *Pädagogische Diagnostik in Deutschland 1885-1932*. Weinheim: Deutscher Studien Verlag.
- Laux, H. (1990). *Pädagogische Diagnostik im Nationalsozialismus 1933-1945*. Weinheim: Deutscher Studien Verlag.