



Universität Potsdam

Diether Hopf

Forschungsstand, Forschungsschwerpunkt und Institutionalisierung der pädagogischen Diagnostik

first published in:
Bildungsforschung : Probleme, Perspektiven, Prioritäten / im Auftr. der
Bildungskommission hrsg. von Heinrich Roth und Dagmar Friedrich. -
Stuttgart : Klett, 1975. - (Gutachten und Studien der Bildungskommission /
Deutscher Bildungsrat ; ...)

Postprint published at the Institutional Repository of Potsdam University:
In: Postprints der Universität Potsdam
Humanwissenschaftliche Reihe ; 112
<http://opus.kobv.de/ubp/volltexte/2009/3731/>
<http://nbn-resolving.de/urn:nbn:de:kobv:517-opus-37319>

Postprints der Universität Potsdam
Humanwissenschaftliche Reihe ; 112

13. Pädagogische Diagnostik

13.0. Einführung

Es gibt kein Lehren und Lernen, ohne daß sich nicht das Bedürfnis entwickelte, zuverlässige Daten über die Ausgangslage des Lernenden, den Fortschritt seiner Lernprozesse und das Lernergebnis, gemessen am Lernziel, zu gewinnen. So entstand die Testforschung. Sie ist für die einen zu einer Perfektion entwickelt worden, die kaum noch verbessert werden kann, für die anderen ist sie Symbol einer Technologie, die den Menschen gefährdet, weil der Test festlegt und abstempelt.

Wie sehr beide Meinungen Fehlurteile darstellen, beweist das Gutachten von Diether HOPF, einem der besten Kenner der Testforschung. Weder kann auf die Testforschung und ihre gezielte Weiterentwicklung verzichtet werden, noch kann die Pädagogik auf Tests verzichten. Die von den Pädagogen praktizierten Zensuren sind unzulänglich, aber auch die Tests sind es, weil die Erforschung ihrer pädagogischen Relevanz noch in den Anfängen steckt, aber beide können sinnvoll weiterentwickelt werden. Der Haupteinwand selbst gegen die besten Tests ist bis heute, daß sie nur ohne Berücksichtigung des „treatment“ messen können, d. h. die traditionellen Tests können die Lernumwelt und deren Wirkungsgrad nicht miterfassen, der Lernende ausgesetzt waren und sind.

HOPF benennt vier Funktionen der Pädagogischen Diagnostik: Beeinflussung des Lernens, Selektion und Klassifikation, Information und Hilfsinstrument für die Forschung. Die heutige Forschung zeigt eine deutliche Verlagerung der erkenntnisleitenden Interessen. Zwar bleibt die Diagnostik die Wissenschaft von den individuellen Unterschieden, aber sie zielt heute mehr auf die Erforschung der Ursachen von Erfolgen und Mißerfolgen ab, um die Befunde für didaktische Maßnahmen auswerten zu können, und sie tendiert heute mehr dazu, „Umweltmaße“ zu entwickeln, die fähig sind, die Güte von „Lernumwelten“ zu diagnostizieren. Sie geht insgesamt mehr auf die Förderung des Lernens und des Lernenden aus als auf Auslese.

Die Funktion der Tests bei der Beeinflussung des Lernens ist heute Lernförderung. Dazu ist eine Diagnose der Ausgangslage, der vorhandenen Vorkenntnisse und Fähigkeiten notwendig. Der Lehrer braucht Informationen über die Bereitschaft der Lernenden, sich zu engagieren und ausdauernd zu lernen. Es sind aber auch Informationen über die Variablen „Gelegenheit zum Lernen“ und die „Qualität des Unterrichts“ — Merkmale der „Lernumwelt“ — nötig. Die Verfahren der Pädagogischen Diagnostik müssen vor allem geeignet sein, eine sinnvolle, individuelle Förderung zu ermöglichen. Verbale Intelligenztestergebnisse dienen dann beispielsweise nicht der endgültigen Festlegung eines IQ, sondern der Beurteilung der Fähigkeit, Unterricht zu verstehen. Was die Motivationstests oder gar die Messung der „Qualität des Unterrichts“ betrifft, so steckt die Forschung noch völlig in den Anfängen. Ein Grund für diese Forschungslage besteht zweifellos in der

Jahrhunderte alten Tradition, Mißerfolge und Fehlleistungen einseitig auf die Schüler zurückzuführen, weniger auf die Effizienz des Lehrers und des Lehrsystems. Nötig wären vor allem diagnostische Instrumente, die die differentielle Ansprechbarkeit von Schülern auf verschiedene Unterrichtsmethoden zu untersuchen erlauben würden. Ebenso nötig wären Instrumente zur laufenden Beurteilung der Lernfortschritte und Lernrückstände in bezug auf die angestrebten Lernziele und Lerninhalte. Letzteres erfordert eine stärkere Integration von Pädagogischer Diagnostik und Curriculumentwicklung.

Am geringsten sind Meßinstrumente für das soziale und affektive Lernen entwickelt. Dazu gehört das Erkennen der Lernbedürfnisse der Kinder und Jugendlichen, der Interaktionsstrukturen in der Klasse, vor allem des „hidden curriculum“. Aber auch hier gibt es Ansätze und Möglichkeiten.

Was Selektion, Klassifikation und Prognose anbetrifft, so ist deutlich festzustellen, daß die Tendenz mit Recht von der Selektion weg zur Klassifikation geht, d. h. der Einsatz von Tests soll vor allem helfen, für jeden Schüler die für ihn wirksamste Lernumwelt zu finden — Klassifikation also als Voraussetzung eines differenzierenden und individualisierenden Unterrichts. Dazu bedarf es neuer Schwerpunkte der Testforschung, die umfangreiche Untersuchungen über „aptitude-treatment-interactions“ im Unterricht zur Voraussetzung haben. Für die Beratung des Lernenden sind diagnostische Verfahren nötig, die eine „positive Diskriminierung“ (CRONBACH und SUPPES 1969) ermöglichen.

Auch alle übrigen Informationen, die man sich von Tests verspricht, stehen unter einer neuen Anwendungs- und Forschungstendenz, es geht weniger um „Meßbänder“ als um „Schätzbänder“, es wird versucht, mehr inhaltliche Angaben über die ganz oder teilweise bewältigten Lernziele zu gewinnen, die an die Stelle von Zensuren gesetzt werden können.

Ein Kapitel widmet HOPF dem Einsatz von Tests bei der Erforschung und Reform des Bildungswesens. Hieraus sei ein entscheidender Satz zitiert: „Gleichgültig, ob es sich um Versuche mit individualisiertem Unterricht, um Erprobung von Formen des entdeckenden Lernens, um Lernsequenzen, bei welchen sich die Schüler ihre Ziele selbst setzen, oder um neue Inhalte handelt, stets ist die Pädagogische Diagnostik den curricularen Intentionen nachgeordnet, da sich auf diese Weise am ehesten sichern läßt, daß . . . die Pädagogische Diagnostik keine den Intentionen des Curriculum zuwiderlaufenden Rückwirkungen auf die Schüler ausübt.“

Die Konsequenzen für mittelfristige Forschungsschwerpunkte der Pädagogischen Diagnostik zieht HOPF im letzten Kapitel. Er macht präzise Aussagen über neue Möglichkeiten und Schwerpunkte bei der Erforschung

- kriteriumbezogener diagnostischer Verfahren,
- individueller Unterschiede,
- von Umweltmaßen,
- diagnostischer Verfahren im Bereich des sozialen und affektiven Verhaltens,
- psychometrischer Fragen,
- bei der Erstellung von Itembanks
- und der Mithilfe bei Schulversuchen.

Unerschöpfliche und nicht ausgeschöpfte Forschungsaufgaben!

Forschungsstand, Forschungsschwerpunkte und Institutionalisierung der pädagogischen Diagnostik

13.1. Vorbemerkung

Die bisher bekannt gewordenen Versuche, Bildungsforschung umfassend zu planen, sind unbefriedigend geblieben. Es ist nirgends gelungen, eine plausible Taxonomie oder akzeptable Kriterien für Prioritätsentscheidungen aufzustellen, die hierfür die notwendige Voraussetzung darstellen würden. Die Schwierigkeit besteht offenbar darin, daß zuvor ein wissenschaftlicher *und* politischer Konsens über die Bezeichnung der Forschungsfelder und ihre Rangordnung hergestellt werden muß.

Forschungsdefizite aufgrund praktischer, das heißt gesellschaftlich-politischer Relevanz der Forschung identifizieren heißt gesellschaftlich-politische Optionen den Entscheidungen über Forschungsprioritäten zugrunde legen. Geht man davon aus, daß in dieser Gesellschaft Dissens besteht über die Prinzipien der gesellschaftlichen Entwicklung, scheidet ein gesellschaftstheoretischer Konsens als Grundlage der Forschungsplanung aus. An seine Stelle können nur konsensfähige politisch-pragmatische Annahmen mittlerer Reichweite treten, die einerseits mit den unterschiedlichen Interessen noch in Übereinstimmung gebracht werden können, andererseits aber für die Forschungsplanung konkretisierbar bleiben.

Solche Annahmen lassen sich auf zwei Ebenen der Spezifizierung formulieren: erstens auf der Ebene der gesellschaftlichen Mängelanalyse und zweitens auf der Ebene eines — auf einer solchen Mängelanalyse beruhenden — bildungspolitischen Programms. Zwar liegen auch diesen Annahmen auf beiden Ebenen zuletzt noch gesellschaftliche Zielvorstellungen zugrunde, doch ihr pragmatischer Charakter läßt die Konsensbildung zu. Ob man eine Forschungsplanung, die gesellschaftlich-politische Relevanz beansprucht, auf der Ebene der Mängelanalyse oder der eines bildungspolitischen Programms ansiedeln will, ist eine Frage der Fruchtbarkeit dieser Ansätze für die Setzung von Prioritäten der Bildungsforschung.

Nimmt man die Sekundarstufe als Beispiel eines entscheidungsrelevanten und forschungsbedürftigen Problembereichs, so könnten Prioritäten offensichtlich auf der Grundlage sowohl einer Mängelanalyse als auch eines bildungspolitischen Programms bestimmt werden. Die Mängelanalyse würde rasch auf die Unzulänglichkeit der Selektionsmechanismen, auf die mangelnde Durchlässigkeit, auf curriculare Defizite, auf die Fragwürdigkeit der Legitimierung unterschiedlicher Lernziele, auf die Organisations- und Verwaltungsschwierigkeiten usw. hinweisen. Ein Überblick über

die Ergebnisse der Bildungsforschung würde dann ergeben, wo Forschungslücken liegen und welche Forschungsschwerpunkte unter den Gesichtspunkten dieser konkreten Mängelanalysen zu bilden wären.

In dem Bereich der Sekundarstufe läßt sich die gesellschaftliche Mängelanalyse zum gegenwärtigen Zeitpunkt im wesentlichen als abgeschlossen betrachten. Darüber hinaus ist aber auch die Übersetzung der Ergebnisse der Mängelanalyse in bildungspolitische Programme bereits weitgehend geleistet worden und teilweise sogar die Entwicklung konstruktiver Lösungen in Form von Handlungsmodellen gelungen¹. Dies gilt bis zu einem gewissen Grade auch für eine Reihe weiterer entscheidungsrelevanter und forschungsbedürftiger Problembereiche: Insbesondere durch die Arbeiten des Deutschen Bildungsrates sind übergreifende Mängelanalysen vorgelegt sowie bildungspolitische Vorstellungen zu zentralen Fragen unseres Bildungswesens entwickelt und – seit Erscheinen des Strukturplans für das Bildungswesen – weiter vorangetrieben worden. Hierzu zählen beispielsweise die Problembereiche: Frühes Lernen; Integration allgemeiner und beruflicher Bildung; Abschlüsse.

Soweit bildungspolitische Programme als konstruktive Antwort auf die Mängelanalyse zu verstehen sind, dürften sie aufgrund der ihnen zugrunde liegenden eindeutigen bildungspolitischen Option besser als eine bloße Mängelanalyse geeignet sein, die Aufstellung eines Prioritätenkatalogs und die Planung der Bildungsforschung anzuleiten.

Im vorliegenden Gutachten soll daher der Versuch unternommen werden, den zuletzt genannten Ansatz, nämlich die Ermittlung von Forschungsdefiziten und die Erarbeitung eines Schwerpunktprogramms für das Forschungsfeld pädagogische Diagnostik auf der Grundlage der vorliegenden Mängelanalysen und der bezeichneten konstruktiven bildungspolitischen Reaktionen hierauf, weiter zu verfolgen. Die Überlegungen zu den Forschungsprioritäten der pädagogischen Diagnostik erfolgen dabei vorwiegend auf dem Hintergrund des Problembereichs Neugestaltung der Sekundarstufe, insbesondere der Gesamtschule. Wenngleich dies nicht bedeutet, daß nicht auch die Prämissen der Reformansätze in der Sekundarstufe hinterfragt werden können, ist auf diese Weise doch eine mittlere Konkretisierungsebene gewonnen, auf welcher Prioritäten diskutiert werden können.

Die notwendige Beschränkung auf einen solchen Problembereich, dessen Auswahl sein derzeitiges bildungspolitisches Gewicht nahelegt, wird insofern nicht zu einer übermäßigen Einengung des Forschungsfeldes führen, als der Versuch unternommen wird, unter Berücksichtigung der Forschungsliteratur die *Hauptfunktionen* der pädagogischen Diagnostik zu kennzeichnen und zu diskutieren. Zudem umgreift der Bereich Sekundarstufe eine Reihe von Problemen, die auch für die anderen Bereiche zentrale Bedeutung besitzen, so daß sich hierbei Fragen stellen, die mit hoher Wahrscheinlichkeit auch in anderen Problembereichen aufgeworfen würden, wenn diese in ähnlicher Weise den Untersuchungsgegenstand bildeten. Man wird daher davon

¹ Vgl. zum Beispiel die Gesamtschulempfehlung des Deutschen Bildungsrates (deren Zustandekommen durch das Paradigma Analyse — Programm — Konstruktion im übrigen nicht erschöpfend erklärt wird).

ausgehen können, daß die Mehrzahl der im Gutachten bezeichneten Forschungsdefizite zugleich solche des Bildungssystems insgesamt darstellen. Eher umgekehrt ist also sogar damit zu rechnen, daß die Darstellung der anstehenden Forschungsprobleme im Gutachten zunächst den Eindruck erwecken wird, als würden hier völlig unrealistische Forderungen an die Forschung gestellt oder als seien die Mängel der gegenwärtigen pädagogischen Diagnostik so umfassend, daß für absehbare Zeit keine Hoffnung auf Verbesserung bestünde. Dieser Eindruck muß zunächst auch entstehen, wenn z. B., wie in Abschnitt 13.2.1. aufgrund einer lernpsychologischen Analyse diejenigen Punkte identifiziert werden, an welchen diagnostische Verfahren verwendbar wären. Damit wird jedoch keineswegs unterstellt, daß es auch wünschenswert wäre, alle Möglichkeiten zum Einsatz von Tests oder sonstigen Verfahren wahrzunehmen. Im Gegenteil werden an mehreren Stellen Argumente dafür vorgebracht, durch Forschung das bei gegebenen Lernzielen unverzichtbare *Minimum* an Diagnostik zu ermitteln, das unter Umständen weit unter dem derzeit Üblichen liegt. Die Forschung zur pädagogischen Diagnostik sollte demnach in zwei quasi entgegengesetzten Richtungen verlaufen: Einerseits sollten existierende oder im Gutachten darüber hinaus erwähnte Ansätze vorangetrieben, und das heißt: die schon bestehende pädagogische Diagnostik sehr entschieden weiterentwickelt und ausdifferenziert werden, andererseits sollten aber auch durch Forschung diejenigen Bedingungen spezifiziert werden, unter denen der Bedarf an diagnostischen Verfahren sich drastisch reduzieren läßt, indem beispielsweise der Selbststeuerung des Lernens durch die Schüler in flexiblen Formen der Unterrichtsorganisation größerer Raum gewährt wird. Daß dabei die zuerst genannte Richtung sehr viel ausführlicher diskutiert wird als die zweite, ist vor allem durch die derzeitige Forschungslage bedingt; auf diesen Forschungsstrang beziehen sich dann auch die in Kapitel 13.4 vorgeschlagenen, mittelfristig erreichbaren Forschungsziele sowie die Hinweise auf bereits vorliegende Forschungsergebnisse, deren kurzfristige Implementierung möglich und sinnvoll erscheint.

Es muß betont werden, daß hier nicht daran gedacht ist, Forschungsplanung unter den Gesichtspunkt der Legitimation bildungspolitischer Entscheidungen zu stellen. Der Problembereich wird vielmehr als Paradigma für das heuristische Verfahren zur Ermittlung von Forschungsdefiziten und deren Beantwortung durch ein Forschungsprogramm gewählt. Die Wahl eines aktuellen Bereichs hat im übrigen zur Folge, daß die Notwendigkeit einer ständigen Revision von Forschungsplanung besonders deutlich wird. Denn der Versuch, eine endgültige Prioritätenliste zu erstellen, erweist sich in gleichem Sinne wie eine endgültige Reformplanung als illusorisch.

Die folgenden Ausführungen stehen zu einem Teil in der Tradition einer Diagnostik, die sich als Wissenschaft von den individuellen Unterschieden versteht; eine Alternative zu diesem Ansatz ist auch schwer denkbar, wenn Fragen der individuellen Lernförderung diskutiert werden. Andererseits führt jedoch die in diesem Gutachten erhobene Forderung, Schule und Unterricht müßten sich weitgehend den Möglichkeiten und Bedürfnissen der Schüler anpassen, zu einem zweiten Strang von Überlegungen und Argumenten: Wenn die „Lernumwelten“ so gestaltet werden

sollen, daß sie der Lage der Schüler entsprechen, muß man ihre hierfür relevanten Merkmale kennen und messen können. Dazu bedarf es der Entwicklung von „Umweltmaßen“ und, im Zusammenhang damit, der Erforschung der Ursachen von Erfolg und Mißerfolg in der Schule. Erst wenn Forschungs- und Entwicklungsarbeiten auch auf diesem Sektor vorangetrieben sind, erhält der Lehrer die Möglichkeit, den Gründen für Lernschwierigkeiten auf die Spur zu kommen und seine Förderungsmaßnahmen danach einzurichten. Der Gedanke, daß die pädagogische Diagnostik nicht nur individuelle, sondern auch situative Unterschiede zu erfassen erlauben sollte, und daß die diagnostischen Befunde erst aufgrund der Kenntnis der Ursachen von Erfolg und Mißerfolg sinnvoll interpretierbar sind, kehrt daher in diesem Gutachten mehrfach wieder.

13.2. Funktionen der pädagogischen Diagnostik

Unter pädagogischer Diagnostik werden alle Verfahren, Vorgänge und Maßnahmen verstanden, die der Messung und Beurteilung des Input, Output und des Verlaufs des pädagogischen Geschehens dienen². Die wichtigste und verbreitetste Form der gegenwärtigen pädagogischen Diagnostik besteht in den Zensuren und Zeugnissen. Hinzu kommen einige Leistungs- und Intelligenztests, soziometrische Verfahren sowie auf Beobachtungen beruhende Urteile von Lehrern über Schüler (zum Beispiel während des Unterrichts) oder von Schulaufsichtsbeamten und Schulleitern über Lehrer; ferner gelegentliche Versuche der Evaluation von Curricula oder von Unterrichtsergebnissen.

Die vielfältigen Funktionen, welche die pädagogische Diagnostik erfüllen kann, lassen sich für die Absichten dieses Gutachtens in vier Gruppen unterteilen: die Beeinflussung des Lernvorgangs; Selektion, Klassifikation, Prognose; Information sowie Forschung und Entwicklung. Dabei wird der von der Beeinflussung des Lernens handelnde Abschnitt (13.2.1.) aufgrund der darin diskutierten, allgemeinen lernpsychologischen Überlegungen weniger deutlich auf die spezifischen Probleme der Sekundarstufe oder der Gesamtschule zu beziehen sein als die übrigen Teile.

13.2.1. Beeinflussung des Lernens

Es dürfte keinem Zweifel unterliegen, daß Beurteilung, Bewertung und Diagnose zentrale Elemente des schulischen Lernens darstellen und auf das Resultat des Lernens sowie auf das weitere Lernen erheblichen Einfluß ausüben. So ist beispielsweise immer wieder gezeigt worden, daß sich wiederholende Mißerfolge, die in einer Folge negativer Urteile zum Ausdruck kommen, die Motivation zu weiterem

² Der Begriff wird hier also in weiterem Sinne gebraucht als beispielsweise bei Ulich und Mertens, 1973, 9, wo er, in Analogie zum Begriff der psychologischen Diagnostik, auf die „Erforschung und Erfassung individueller Unterschiede“ beschränkt bleibt; vgl. auch Hörmann, 1964.

Lernen erheblich schwächen, wie umgekehrt Erfolgserlebnisse zur Erhöhung der Ausdauer und Anstrengungsbereitschaft beitragen können. Auch weiß man, daß dem Schüler während des Lernverlaufs angebotene Orientierungspunkte sich auf den Lernerfolg im allgemeinen positiv auswirken.

Nun läßt sich freilich nahezu alles, was in Schule und Unterricht geschieht, als Lernvorgang begreifen, und es muß der Versuch unternommen werden, durch wie immer künstliche Unterteilungen diesen Bereich zu strukturieren. Trotz aller gegen eine solche Abtrennung bestehenden Bedenken wird daher im folgenden zwischen kognitivem Lernen sowie sozialem und affektivem Lernen unterschieden. Wegen der Breite jedes dieser Bereiche läßt sich freilich nur anhand weniger Ausschnitte der Versuch unternehmen, Schwerpunkte für Forschung und Entwicklung der pädagogischen Diagnostik zu bezeichnen. Daß die Bereiche des motorischen und ästhetischen Lernens im vorliegenden Gutachten nicht zur Sprache kommen, sollte nicht als Argument gegen ihre Relevanz mißverstanden werden.

13.2.1.1. Kognitives Lernen

Um bestimmen zu können, an welchen Stellen im Lernprozeß die pädagogische Diagnostik sinnvoll verwandt werden kann, bedarf es einer theoretisch angeleiteten Beschreibung des schulischen Lernens. Wenngleich eine solche Forderung für die Bestimmung von Forschungsprioritäten in allen Funktionsbereichen der pädagogischen Diagnostik gilt, soll nur an dieser Stelle der Versuch gemacht werden, anhand eines Modells schulischen Lernens diejenigen Nahtstellen zu bezeichnen, an denen die zur Förderung des Lernens dienlichen Informationen gewonnen werden können. Die Ergiebigkeit einer solchen Verfahrensweise soll am Beispiel des von CARROLL (1963, deutsch 1973) vorgelegten Modells skizziert werden³.

CARROLL unterscheidet fünf Variablen, deren Kenntnis es ermöglichen soll, kognitives Lernen sowie den Erwerb von Fertigkeiten zu erklären. Die Variablen sind so definiert, daß sie sich weitgehend in Zeiteinheiten ausdrücken lassen. Dies gilt insbesondere für die zunächst genannten drei Variablen: Fähigkeit, Ausdauer und Gelegenheit zum Lernen.

Die *Fähigkeit* eines Schülers wird ausgedrückt in der Zeit, die er benötigt, um ein festgelegtes Kriterium bei einer bestimmten Aufgabe zu erreichen, und zwar unter optimalen Lernbedingungen, das heißt, wenn sowohl die Qualität des Unterrichts als auch die Motivation des Schülers, das Kriterium zu erreichen, hoch sind. Unter diesen Voraussetzungen gibt die Lerngeschwindigkeit Auskunft über die Lernfähigkeit.

Die *Ausdauer* erweist sich am Zeitaufwand, den ein Schüler bereit ist, mit Lernen zu verbringen. Welches auch die Ursachen für die Motiviertheit und Ausdauer des

³ Es versteht sich, daß hier nur stichpunktartig auf das Modell eingegangen werden kann. Im übrigen muß man sich vor Augen halten, daß man möglicherweise zu etwas anderen Schlußfolgerungen käme, wenn man ein anderes Lernmodell als das CARROLLSche den Überlegungen zugrunde legen würde.

Schülers sein mögen, es wird angenommen, daß der Lernerfolg um so höher liegt, je mehr Zeit aufs Lernen verwendet wird.

Gelegenheit zum Lernen ist definiert als die Zeit, die zum Lernen einer bestimmten Aufgabe tatsächlich zur Verfügung steht. Ist diese zu kurz, so wird der Lernerfolg selbst bei hoher Motivation ausbleiben. Dies dürfte insbesondere beim Klassenunterricht unvermeidlich sein, bei dem der Lehrer sein Unterrichtstempo zum Beispiel auf eine angenommene mittlere Leistungsgruppe zuschneidet.

Der Grad, in welchem Inhalt und Methode des Unterrichts dergestalt strukturiert sind, daß es dem Schüler leicht fällt zu folgen, wird in der vierten Variable, *Qualität des Unterrichts*, gefaßt. Hierzu gehören unter anderem die adäquate Kommunikation zwischen Lehrer und Schüler, richtige und verständliche Erklärung der Inhalte, optimale Formen von Wiederholung, Übung und Feedback sowie die Identifikation und Berücksichtigung besonderer Schwierigkeiten, die Schüler mit bestimmten Inhalten haben.

Schließlich enthält das Modell die Variable *Fähigkeit, den Unterricht zu verstehen*, die mit verbaler Intelligenz vergleichbar sein mag. Bei nicht-optimaler Unterrichtsqualität werden von den Schülern zusätzliche Anstrengungen erfordert, dem Unterricht zu folgen; entsprechend verringert sich der Lernerfolg. Manche Befunde deuten darauf hin, daß Unterschiede in der Art und Qualität des Unterrichts zwar geringe Auswirkungen bei rasch lernenden Schülern zeigen, für die langsamen Lerner jedoch von erheblicher Bedeutung sind.

Es ist hier nicht der Ort, die Zusammenhänge zwischen den genannten Variablen und ihre Implikationen für die Erklärung und Modifikation des schulischen Lernens, das eigentliche Modell also, im Detail zu erörtern; seine theoretische und praktische Fruchtbarkeit hat sich vielfach bestätigt, beispielweise in dem von BLOOM (1971, deutsch 1973) aus bestimmten Teilaspekten des Modells abgeleiteten Ansatz des Mastery Learning. Der Grad des Lernerfolgs, das heißt der Grad, in dem ein Schüler bei einer bestimmten Aufgabe ein bestimmtes Kriterium erreicht, ist in dem Modell eine Funktion der Zeit, die der Schüler tatsächlich mit dem Lernen der Aufgabe verbringt, im Verhältnis zu der Zeit, die er angesichts seiner Fähigkeit, der Qualität des Unterrichts, und der Fähigkeit, den Unterricht zu verstehen, benötigen würde, um das Kriterium zu erreichen.

Die beschriebenen, für den Prozeß des schulischen Lernens zentralen Variablen und ihr Zusammenhang in dem Modell bieten die Möglichkeit, die Rolle der pädagogischen Diagnostik im Bereich des kognitiven Lernens näher zu charakterisieren. CARROLLS (1970) Diskussion einiger Meßprobleme, die beim sogenannten Mastery-Learning auftreten, hat in der folgenden Darstellung Berücksichtigung gefunden.

Eine notwendige Voraussetzung für die Anwendung des Modells auf konkrete Lernsituationen besteht zunächst darin, daß die jeweilige Aufgabe und das Kriteriumverhalten, worauf das Lernen gerichtet ist, spezifiziert werden. In welcher Form und auf welcher Abstraktionsebene dies geschieht, hängt zwar nicht nur von den Bedürfnissen der pädagogischen Diagnostik ab, sondern beispielsweise auch von der Auswirkung der Spezifikationsform auf die Flexibilität der Unterrichtsgestaltung durch den Lehrer. Für die Feststellung des Lernerfolgs bedarf es jedoch einer mög-

lichst präzisen Beschreibung der Lernaufgabe, ihrer Komponenten und Lernziele in einer überprüfbar Form, damit Schüler und Lehrer Aufschluß darüber erhalten, ob oder bis zu welchem Grad ein bestimmtes Kriterium erreicht worden ist. Dies ist auch notwendig, um zu bestimmen, welche Voraussetzungen erfüllt sein müssen, bevor ein Schüler mit dem Lernen der Aufgabe beginnen kann. Denn wenn man intendiert, die Schüler zu guten Lernerfolgen zu führen, wird man den Unterricht auf ihre jeweiligen Vorkenntnisse und Fähigkeiten abstimmen müssen. Die hierzu notwendigen Informationen müssen durch geeignete Formen der pädagogischen Diagnostik gewonnen werden.

Zur Vorbereitung und Durchführung seines Unterrichts sowie zur Planung individueller Lernhilfen benötigt der Lehrer Kenntnisse über die Fähigkeiten seiner Schüler. Diese schlagen sich, nach der Konzeption des hier zugrunde gelegten Modells, in der Zeit nieder, die ein Schüler braucht, um unter sonst optimalen Bedingungen eine Aufgabe zu bewältigen. Der Zeitbedarf der verschiedenen Schüler einer Klasse variiert, wie man weiß, außerordentlich stark, insbesondere in unausgelesenen Schülergruppen, wie sie beispielsweise in einer Gesamtschule anzutreffen sind. Versäumt es der Lehrer, diese Unterschiede zu berücksichtigen, so ist bei zahlreichen Schülern entweder Langeweile, Unlust und Desinteresse an der Aufgabe oder aber der vorzeitige Abbruch des Lernens auf ein bestimmtes Ziel hin die Folge. Insbesondere in dem zweiten Fall, in welchem nicht genügend Gelegenheit zum Lernen geboten wird, muß sich der Lernerfolg spürbar verringern.

Ähnliches gilt für die Ausdauer oder die Motivation des Lernalters: Wenn ein Schüler nicht in der Lage oder gewillt ist, so lange zu lernen, bis er das vom Lehrer oder von ihm selbst gesetzte Kriterium erreicht hat, wird er hinter den gestellten Ansprüchen zurückbleiben. Da es möglich ist, Schüler zu motivieren — sei es durch die Verwendung besonders ansprechender Materialien, sei es durch die Auswahl einer besonders attraktiven Aufgabe oder durch die methodische Gestaltung des Unterrichts —, braucht der Lehrer Informationen über die Bereitschaft der Schüler, sich zu engagieren und ausdauernd zu lernen. Es ist klar, daß auch hier erhebliche individuelle Unterschiede bestehen, sowohl in Hinsicht auf die besondere Ansprechbarkeit für bestimmte Lernziele und -inhalte als auch bezüglich der unterschiedlichen Sensibilität gegenüber den verschiedenen möglichen Maßnahmen zur Förderung der Lernwilligkeit durch den Lehrer.

Von besonderer Bedeutung für die Diagnose der Ursachen unbefriedigender Lernerfolge sind Kenntnisse über die Qualität des Unterrichts. Diese wirkt sich einerseits auf die Motivation und Bereitwilligkeit der Schüler aus, die Anstrengungen des Lernens auf sich zu nehmen, andererseits steht sie in enger Beziehung — möglicherweise in einer Wechselwirkung — zu der Fähigkeit, Unterricht zu verstehen, insofern sich die unterschiedliche Qualität des Unterrichts stärker auf solche Schüler auswirkt, die eine geringere Fähigkeit besitzen, dem Unterricht zu folgen; diese steht, wie CARROLL (1970) vermutet, mit dem, was als verbale Intelligenz bezeichnet wird, und der Fähigkeit zu schlußfolgerndem Denken in Zusammenhang. Mängel in der Qualität des Unterrichts wirken sich negativ auf den Lernerfolg aus, und zwar auch dann, wenn in bezug auf die übrigen Variablen günstige Voraussetzungen bestehen.

Die Variablen „Qualität des Unterrichts“ und „Gelegenheit zum Lernen“ unterscheiden sich von den übrigen Variablen insofern, als die einen Merkmale des Schülers betreffen, die anderen jedoch Merkmale der „Lernumwelt“ bezeichnen und insofern als eher manipulierbar angesehen werden können. Dieser für die Bestimmung der Forschungsschwerpunkte der pädagogischen Diagnostik wichtige Aspekt wird uns unten ausführlicher beschäftigen.

Anforderungen an die pädagogische Diagnostik

Aus dem bisher Gesagten geht hervor, daß für Lernergebnisse, die unter dem gesetzten Kriterium liegen, mehrere Ursachen in Frage kommen. Der Vorzug des hier zugrunde gelegten Modells schulischen Lernens besteht unter anderem darin, daß man eine begrenzte Anzahl möglicher Ursachen für ausbleibende Lernerfolge benennen kann, die mit Hilfe geeigneter Verfahren⁴ der pädagogischen Diagnostik analysiert werden können, so daß dann eine sinnvolle, individuelle Förderung möglich wird. Gegenstände der pädagogischen Diagnostik sind danach erstens die Merkmale, mit denen die Schüler in den Unterricht kommen (Fähigkeiten, Ausdauer, Vorkenntnisse), zweitens die Lernbedingungen, die der Unterricht erzeugt, und drittens die Kriteriumsleistung. Dabei muß man davon ausgehen, daß die Lernvoraussetzungen und Fähigkeiten bei jedem Schüler anders ausgeprägt sind und in anderer Weise durch die verschiedenen Aufgaben und Lernziele abgerufen und stimuliert werden; ferner, daß sie mit der Qualität des Unterrichts, aber auch untereinander in Beziehung stehen können. Daraus folgt, daß Lernen in hohem Maße individualisiert erfolgen mußte, zumindest aber, daß die Verfahren der pädagogischen Diagnostik geeignet sein müssen, die individuellen Verschiedenheiten abzubilden⁵

CARROLL hat darauf hingewiesen, daß man bei der Messung der Lernfähigkeit davon ausgehen muß, daß diese stärker aufgabenspezifisch ist, als man gemeinhin angenommen hat. Darauf deuten insbesondere die Untersuchungen im Kontext eines strikt individualisierenden Unterrichts hin⁶. Dort ist zugleich deutlich geworden, daß die Annahme, die Lernfähigkeit anhand der pro Zeiteinheit erworbenen Kenntnisse diagnostizieren zu können, offenbar zu einfach ist: Unterschiedliche Lerninhalte bedingen unterschiedliche Lernverläufe, so daß die Feststellung einer bestimmten Veränderung pro Zeiteinheit bei einer linear ansteigenden Lernkurve eine andere Bedeutung besitzt als bei einer negativ akzelerierten Kurve. Für die Auswahl einer bestimmten Aufgabe, die ein Schüler lernen soll, bedarf es daher so-

⁴ Wenn hier und im folgenden von „Verfahren“, „Instrumenten“ usw. der pädagogischen Diagnostik die Rede ist, sind darunter nicht unbedingt Tests oder sonstige formalisierte oder mit großem Aufwand entwickelte Hilfsmittel zu verstehen. Vielmehr werden diese Begriffe in einem weiten Sinne gebraucht; beispielsweise ist die systematische Beobachtung auch ein „Instrument“, mit dem man Verhalten „messen“ kann.

⁵ Vgl. hierzu auch Abschnitt 13.2.2.

⁶ Beispielsweise im Zusammenhang der Individually Prescribed Instruction, vgl. LINDVALL und BOLVIN, 1967.

wohl der Kenntnis des zu erwartenden aufgabenspezifischen Lernverlaufs als auch der Beurteilung der Vorkenntnisse des Schülers zur Bestimmung des Punktes auf der Lernkurve, an dem er sich gerade befindet. Während die erste Forderung sich erst nach Abschluß ausgedehnter Lernforschungen erfüllen lassen dürfte, wäre es eine mittelfristig lösbare Aufgabe der Forschung im Bereich der pädagogischen Diagnostik, die Voraussetzungen für die Diagnose der spezifischen Ausgangslage der Schüler zu schaffen sowie Maße zu entwickeln, die eine Schätzung der individuellen Lernfähigkeit erlauben, nach welcher sich die vom Lehrer zu schaffende Gelegenheit zum Lernen richten kann. Intelligenztests dürften — unter anderem aufgrund der möglichen Aufgabenspezifität der Lernfähigkeit — ungeeignete Verfahren für die genannten Zwecke darstellen. Skepsis gegenüber ihrer Brauchbarkeit für die Förderung des Lernens ist angebracht auch angesichts des Befundes, daß bei optimaler, individualisierter Instruktion die Korrelation zwischen der Geschwindigkeit des Lernfortschritts und der Intelligenz gegen Null geht⁷.

Dagegen läßt sich vermuten, daß Messungen der *verbalen* Intelligenz Auskunft geben über die Fähigkeit, den Unterricht zu verstehen. Nicht minder wichtig dürfte hier allerdings auch die Messung des Hörverständnisses sein, da eine Reihe von Untersuchungen Zusammenhänge zwischen Schulerfolg und Leistungen in Hörverständnis tests zum Vorschein gebracht haben.

Über die Ausdauer, mit der ein Schüler sich aktiv einer Aufgabe widmet, dürfte der Lehrer in der Klassensituation nur mit Mühe verlässliche Auskünfte erhalten. Hier wäre es nicht nur wichtig, die angesichts einer bestimmten Aufgabe vorfindbare Ausdauer zu beobachten, sondern auch die potentielle Ausdauer zu kennen, das heißt die maximalen Werte, die ein Schüler erreichen kann; erst dann ließe sich mit größerer Sicherheit auf Defizite in anderen Variablenbereichen schließen, wenn ein Schüler aufgrund eines für ihn relativ hohen Engagements das gesetzte Erfolgskriterium nicht erreicht.

Diagnostische Verfahren, die diesen Zwecken dienen könnten und ohne allzu großen Aufwand im Unterricht verwendbar wären, stehen nicht zur Verfügung. Beispielsweise hat VEROFF (1973) dargelegt, daß zur Messung der in diesem Zusammenhang relevanten Leistungsmotivation derzeit noch keine Instrumente greifbar sind, die verlässlich sind und zugleich nicht allzu viel Zeit in Anspruch nehmen.

Betrachtet man die Verfahren, die derzeit zur Messung und Beobachtung der Qualität des Unterrichts zur Verfügung stehen, und vergleicht sie mit den der Messung individueller Unterschiede dienenden, ausgefeilten Instrumenten, so wird rasch deutlich, daß es sich hier um einen Bereich handelt, in welchem die Forschung kaum

⁷ Vgl. McCLELLAN, 1972, S. 170 (Hinweis auf GLASER). Vgl. auch MICHELS, W. C.: "Some lessons from high school physics." In: Proceedings, 1959 Invitational Conference on Testing Problems, ETS, Princeton 1960, S. 17—26. MICHELS fand, daß im Laufe eines Jahres die Korrelationen zwischen IQ und den Leistungen in einem neu konstruierten Physik-Curriculum (PSSC) stetig abnahmen, der sorgfältig vorbereitete Unterricht also den „klassischen“ Zusammenhang zwischen Schulleistung und Intelligenz reduzierte. Der zu Beginn der Lernperioden beobachtbare Zusammenhang ist andererseits für die Unterrichtsplanung ohne Bedeutung. (Weitere Literatur hierzu u. a. bei SCHWARZER, 1973, und LOHNES, 1973.) Vgl. auch die kürzlich von McCLELLAND, 1973, vorgebrachte, umfassende Kritik an der Verwendung von Intelligenztests im Bildungswesen.

begonnen hat und nur ganz vereinzelte brauchbare Ergebnisse vorliegen⁸. ROSENSHINE (1970) hat in seinem Überblick über die bisherigen Untersuchungen über das Lehrerverhalten im Unterricht gezeigt, daß man bislang nicht über Meßinstrumente verfügt, deren Ergebnisse mit dem Lernerfolg der Schüler in klarem Zusammenhang stehen. Ein Grund für diese Forschungslage besteht zweifellos in der Jahrhundertalten Tradition, Mißerfolge und Fehlleistungen auf die Schüler zurückzuführen, nicht aber die Tätigkeit des Lehrers in Frage zu stellen. Darüber hinaus stellen die bislang entwickelten Verfahren außerordentlich grobe Instrumente dar, so daß bei der Erforschung von Zusammenhängen zwischen Unterrichtsqualitäten und Merkmalen der Lerner (zum Beispiel aptitude-treatment-interactions) sehr undifferenzierte und sehr fein skalierte Maße in Beziehung zueinander gesetzt werden müssen. CARROLLS Modell macht jedenfalls deutlich, daß die Qualität des Unterrichts nicht weniger als Ursache für Leistungsausfälle in Betracht kommt als etwa die Lernfähigkeiten oder die Ausdauer der Schüler.

Angesichts der überaus plausiblen Hypothesen über das Bestehen von Zusammenhängen zwischen Unterrichts- und Schülermerkmalen sowie der differentiellen Ansprechbarkeit von Schülern durch verschiedene Unterrichtsmethoden bedarf es dringend der Erforschung dieser vom Lehrer gesetzten Lernbedingungen, insbesondere der bislang meist vernachlässigten kognitiven Transaktionen und Interaktionen im Unterricht, sowie der Entwicklung entsprechender diagnostischer Instrumente.

Schließlich seien noch die Anforderungen knapp umrissen, die sich in bezug auf die Messung des Kriteriums stellen. Hier geht es um den traditionellen Gegenstand der schulischen Leistungsbeurteilung, die Frage nämlich, ob und in welchem Grad ein Schüler die gesetzte Aufgabe bewältigt und das Lernziel erreicht hat. Klassenarbeiten, mündliche Prüfungen, Zensuren und Zeugnisse sind die üblichen Verfahren, mit deren Hilfe hier geurteilt wird; wenige standardisierte Tests sowie, bei einer Minderzahl von Lehrern, selbst konstruierte informelle Tests kommen hinzu.

Entscheidend dürfte bei der Beurteilung des Kriteriumsverhaltens die genaue Passung des Instruments auf die Ziele und Inhalte des Unterrichts — die zwischen den Lehrern und bei demselben Lehrer von Mal zu Mal variieren —, seine Ziel- und Aufgabenspezifität sowie seine Präzision sein. Darüber hinaus bedarf es auch der Beurteilung der Lernfortschritte der Schüler im Laufe des Unterrichts, des rechtzeitigen Erkennens von Lücken und Unklarheiten, kurz: eines Instrumentariums der Verlaufsbeurteilung des Schülerverhaltens. Auch hier wiederum sind die entscheidenden Eigenschaften der erforderlichen Beurteilungsverfahren Flexibilität in bezug auf Inhalte und Ziele (auch bei Teilaufgaben!) und Kriterienbezogenheit.

An dem von CARROLL entwickelten Modell schulischen Lernens sollte dargestellt werden, an welchen Gelenkstellen im Prozeß des kognitiven Lernens durch pädagogische Diagnostik Informationen gewonnen werden können, mit deren Hilfe eine auf die individuelle Lage des Schülers abgestellte Förderung möglich wird. Es genügt nicht, Mißerfolge beim Lernen oder partiell defizitäre Lernergebnisse als solche zu erkennen und zu einer an anderen Schülern gewonnenen Norm in Beziehung zu set-

⁸ Vgl. GAGE, 1972.

zen, sondern es kommt darauf an, Lernerfolge und -mißerfolge in eine präzise Beziehung zu den Lernzielen und -inhalten des Unterrichts zu setzen sowie die Ursachen von Lerndefiziten zu erkennen. Dabei kann es nicht darum gehen, in jedem einzelnen Falle Messungen aller aufgeführten Variablen vorzunehmen (beispielsweise dürfte es in der schulischen Praxis nicht sinnvoll sein, die bei der Bearbeitung jeder einzelnen Aufgabe zu Tage tretende Ausdauer eines Schülers festzustellen; eher wäre es angezeigt, die auf die Ausdauer und Motivation negativ einwirkenden Einflüsse wie zum Beispiel kontinuierliche Mißerfolgserlebnisse des Schülers aufgrund einer Serie schlechter Beurteilungen zu beseitigen); es sollte jedoch zumindest die Möglichkeit zu ihrer Messung geschaffen werden. Dazu bedarf es umfangreicher Forschungen und der Entwicklung von Untersuchungsverfahren, mit deren Hilfe jeder der das Modell ausmachenden Parameter genau und verläßlich gemessen werden kann.

Die sich aus dem CARROLLSchen Lernmodell ergebenden Anforderungen an die pädagogische Diagnostik lassen sich vorläufig in drei Schwerpunktbereiche zusammenfassen, in welchen Forschung und Entwicklung betrieben werden müssen:

- a) der Bereich der Messung des Kriteriumverhaltens. Hier geht es vor allem um die Entwicklung von auf Lernziele und Leistungsdimensionen⁹ bezogenen Verfahren, die eine hohe Flexibilität aufweisen, also nicht nur am Ende, sondern auch während des Lernvorgangs verwendet und präzise auf den Unterricht zugeschnitten werden können. Darüber hinaus gilt es, die Voraussetzungen dafür zu schaffen, daß das Kriteriumverhalten möglichst breit erfaßt werden kann und nicht nur in seinen leicht meßbaren Ausschnitten; hierzu werden diagnostische Verfahren unterschiedlicher Art benötigt;
- b) der Bereich der Messung individueller Unterschiede, etwa in bezug auf Lernfähigkeit, Lernvoraussetzungen, Ausdauer, Fähigkeit, dem Unterricht zu folgen;
- c) der Bereich der Messung von Qualitäten der „Lernumwelt“, hier: der Unterrichtsmethoden und -gestaltung durch den Lehrer, darüber hinaus jedoch auch der Lernhemmnisse, die außerhalb des Unterrichts liegen und Ursache für Mißerfolge beim Lernen in der Schule sein können (und demzufolge bei der Förderung wie der Beurteilung Berücksichtigung finden müssen). Durch die Gewinnung von Informationen über die Determinanten des Lernerfolgs und -mißerfolgs wird es möglich, über die Deskription individueller Unterschiede hinaus zu deren Ursachen vorzustoßen und entsprechende Maßnahmen zu treffen.

13.2.1.2. Soziales und affektives Lernen

Auch bei diesem Gegenstandsbereich wäre es – ähnlich wie bei der Betrachtung des kognitiven Lernens – sinnvoll, von einem Modell oder einer Theorie ausgehend die sich an die pädagogische Diagnostik richtenden Anforderungen zu bestimmen und daraus mögliche Forschungsschwerpunkte abzuleiten. Die Forschungslage in diesem

⁹ ROEDER, 1974.

Gebiet erlaubt es allerdings nicht, diesen Weg einzuschlagen, vor allem nicht in einer so stark verkürzten Form wie im vorhergehenden Kapitel. Es soll deshalb der Versuch unternommen werden, relativ unsystematisch einige Probleme des sozialen¹⁰ und affektiven Lernens zu benennen, die wichtige Implikationen für die pädagogische Diagnostik besitzen.

Um kurz zu charakterisieren, worauf sich die pädagogische Diagnostik in diesem Bereich richtet, seien einige gerade im Zusammenhang der Gesamtschuldiskussion häufig vertretene Ziele stichwortartig wiedergegeben: Hier geht es beispielsweise um die Erziehung zur Kooperationsfähigkeit, zur Kommunikationsbereitschaft, politischen Verantwortung, Fairneß, Achtung vor den Menschenrechten, oder auch um die Förderung der Lernmotivation, um die Entwicklung angemessener Einstellungen zu den Mitmenschen, insbesondere zu Minoritäten, oder um die Ausbildung der Phantasie¹¹.

Die wesentliche Funktion der pädagogischen Diagnostik besteht auch in diesem Lernbereich darin, dem Schüler diejenigen Informationen zur Verfügung zu stellen, durch die er für die Entwicklung seiner Persönlichkeit den notwendigen feedback erhält, sowie dem Lehrer die Möglichkeit zu eröffnen, die Auswirkungen seiner Tätigkeit zu beurteilen. Diese Funktion wird auch in der Öffentlichkeit ein um so größeres Gewicht erhalten, je stärker soziales und affektives Lernen in einem traditionell „akademischen“ Curriculum akzentuiert werden; auch werden sich die herkömmlichen Formen und Inhalte der pädagogischen Diagnostik als um so inadäquater erweisen, je deutlicher im Zuge einer stärker auf die besondere Förderung von „Benachteiligten“ gerichteten Schule die herkömmlichen Lernziele zurücktreten. Aber auch in der derzeitigen Schule spielt soziales und affektives Lernen eine erhebliche Rolle¹², wenngleich es nur selten angeleitet und beobachtet wird. Einer hierauf gerichteten pädagogischen Diagnostik würde die Aufgabe zufallen, die Auswirkungen von Schule und Unterricht deutlich zu machen und dadurch überhaupt erst eine Diskussion über die Wünschbarkeit bestimmter Formen des sozialen und affektiven Lernens zu ermöglichen.

Aufgaben der pädagogischen Diagnostik

Wenn man soziales und affektives Lernen unterstützen und steuern sowie Informationen darüber erhalten will, welche Auswirkungen Schule und Unterricht auf die Schüler ausüben, benötigt man Verfahren, die es ermöglichen,

¹⁰ Auf die Künstlichkeit der Trennung zwischen sozialem, affektivem und kognitivem Lernen ist wiederholt hingewiesen worden, vgl. z. B. KRAPPMANN, 1973. Die Unterscheidung erfolgt hier aus Gründen der einfacheren Darstellung bestimmter diagnostischer Probleme.

¹¹ Ziele dieser Art werden auch in der angelsächsischen Literatur gerade in letzter Zeit oft diskutiert, vgl. z. B. Proceedings of the 1971 Invitational Conference, 1972, p. 130 und 137; HALSEY, 1972, 195; HOLTZMAN, 1971, 547. Einzelne dieser Dimensionen sind unter dem Aspekt der Beurteilung bereits bearbeitet worden, vgl. z. B. ÜLICH und MERTENS, 1973, 128 ff.

¹² Vgl. z. B. die Schulanalysen von JACKSON, 1968 und 1973.

- Bedürfnisse, Defizite, Probleme der einzelnen Kinder zu erkennen, um sie in der Planung und Durchführung des Unterrichts berücksichtigen zu können;
- Interaktionen zwischen Lehrer und Schüler und Schülern untereinander zu diagnostizieren, so daß der Lehrer Rückmeldung über die Auswirkung seines Verhaltens und damit die notwendige Information zur Steuerung seines weiteren Verhaltens erhält;
- die Effekte längerer Unterrichtsperioden auf das Verhalten der Kinder zu erkennen;
- die Auswirkungen nicht-intendierter Lernprozesse kennenzulernen, die in der Schule ablaufen.

Schon bei der Aufzählung dieser Aufgaben werden wesentliche Unterschiede zur pädagogischen Diagnostik im Bereich des kognitiven Lernens deutlich: Zum einen ist hier das „*hidden curriculum*“ der Schule von besonderem Interesse; dabei handelt es sich um ein breites Feld wichtiger und mutmaßlich massiver Lernvorgänge in der Schule, über deren Natur so gut wie nichts bekannt ist – ein erstaunlicher Tatbestand angesichts der Überlegung, daß es im Interesse jedes Lehrers liegen müßte zu wissen, welche – vielleicht gänzlich unerwünschten und unververtretbaren – Prozesse sich unbemerkt in Schule und Unterricht abspielen; dieser Forschungsgegenstand dürfte insbesondere für die als Ganztagschulen angelegten Gesamtschulen von Bedeutung sein. Zum anderen fällt sogleich die Empfindlichkeit dieses Gegenstandes ins Auge: Nicht nur, daß grobe, für jedermann sichtbare Beobachtungs- und Beurteilungsverfahren die Prozesse sozialen und affektiven Lernens empfindlich stören, wenn nicht unmöglich machen würden; schon die explizite Planung und Programmierung des Unterrichts auf ein soziales Lernziel hin, beispielsweise auf das Ziel der Verbesserung der Kooperationsfähigkeit, kann dazu führen, daß der gewünschte Lernprozeß gar nicht erst in Gang kommt¹³. Das Problem besteht daher nicht nur darin, unaufdringliche Verfahren¹⁴ anzuwenden, sondern die Planung, die Durchführung und die Bewertung des pädagogischen Geschehens eng aufeinander zu beziehen. Dies stellt an den Lehrer noch höhere Anforderungen als die erforderliche Integration von pädagogischer Diagnostik und Curriculum im Bereich des kognitiven Lernens. Daraus folgt auch, daß hier der Lehrerbildung ein besonderes Gewicht zukommt, insofern kaum Hilfen „von außen“ angeboten werden können, beispielsweise in Form fertiger Testverfahren, die der Lehrer nur noch anzuwenden bräuchte. Die Beurteilung und Diagnose des sozialen und affektiven Lernens verlangt vielmehr vor allem erst einmal ein sorgfältiges Studium der hier involvierten komplexen Phänomene und der Möglichkeiten und Folgen pädagogischer Interventionen, bevor an eine Diagnose und Evaluation gedacht werden kann.

¹³ Vgl. hierzu KRAPPMANN, 1973.

¹⁴ In diesem Bereich wird man sich häufig auf Beobachtungsverfahren beschränken müssen.

a) Versucht man auch hier vorläufige Forschungsschwerpunkte zu benennen, so zeigt sich besonders klar, daß es notwendig wäre, zunächst den Bereich des sozialen und affektiven Lernens in der Schule selbst genauer zu untersuchen, beispielsweise also in Erfahrung zu bringen, welches die psychischen „Kosten“ von Schule und Unterricht sind oder welche Merkmale und Inhalte das „*hidden curriculum*“ für verschiedene Teilnehmergruppen an unseren Schulen besitzt. Im Zuge der Entwicklung eines Modells z. B. des sozialen Lernens in der Schule würden dabei auch diagnostische Verfahren entworfen oder adaptiert werden, deren Brauchbarkeit für die Praxisbedürfnisse dann sinnvoll diskutiert werden könnte.

Im Bereich des sozialen und affektiven Lernens ist im übrigen die Forschung noch weit von dem Punkt entfernt, der bei der pädagogischen Diagnostik im Bereich des kognitiven Lernens schon erreicht ist. Dies zeigt sich sogar in dem wahrscheinlich noch am besten erforschten Gebiet, der Leistungsmotivation. Hier gibt es eine Fülle pädagogisch relevanter Befunde, beispielsweise den der Differenziertheit des Leistungsmotivs für verschiedene Personengruppen, nicht jedoch praktikable diagnostische Verfahren, mit deren Hilfe ein Lehrer feststellen könnte, bei welchem Schüler welche Ausprägung des Leistungsmotivs vorliegt, um sich in seinem Unterricht danach zu richten (vgl. VEROFF, 1973).

Sofern die theoretischen Arbeiten auf einem Gebiet schon vorangetrieben sind (wie beispielsweise in der Motivationspsychologie), bedarf es daher dringend der Entwicklung von Verfahren, die den Bedürfnissen der Schule entsprechen. Bereits entwickelte neue Instrumente müssen zudem auf ihre theoretische Fundierung überprüft werden. So sind beispielsweise die Versuche, dem Ungenügen an den sogenannten „Kopfnoten“ der Zeugnisse dadurch abzuhelfen, daß statt der bisher üblichen Ziffernoten nun präziser anmutende Ratingskalen entworfen werden, mit Skepsis zu betrachten. Solange über die damit zu messenden Eigenschaften reine common-sense Vorstellungen herrschen, stellt die Entwicklung von Instrumenten mit größerer *face-validity* eher eine Gefahr als eine Hilfe dar.

b) Wichtiger vielleicht noch als die Bearbeitung des genannten Forschungs- und Entwicklungsfeldes dürfte es sein, die Auswirkungen der bislang und künftig verwendeten Formen der pädagogischen Diagnostik im kognitiven Bereich auf das affektive und soziale Lernen zu untersuchen. Es ist ein erstaunlicher Tatbestand, daß man seit Jahrzehnten weiß, daß ständige Mißerfolge zu einer drastischen Reduzierung der Lernbereitschaft und zu einem negativen Selbstbild¹⁵ führen können, dennoch aber unverändert ca. 75 Prozent aller schlechten Zensuren an 25 Prozent der Schüler gegeben werden und nach wie vor die Vorstellung von der zu erstrebenden Normalverteilung der Zensuren weit verbreitet ist, bei der es notgedrungen einen bestimmten Anteil von Minderleistungen gibt. Hier dürfte es weniger darum gehen, neue Meßinstrumente zu entwickeln, mit denen die Fortschritte der Schüler beurteilt werden können, als vielmehr radikal neue Modelle des Lehrens, Lernens und

¹⁵ Vgl. u. a. WEIS, 1971, 547 f.

Beurteilen zu entwickeln und zu erproben, bei denen die genannten Nachteile entfallen. Versuche mit dem sogenannten Mastery Learning¹⁶ wären eine Möglichkeit; konsequent individualisierter Unterricht¹⁷ mit den entsprechenden Verfahren der Information der Schüler über ihren Leistungsstand eine andere. Auch andere Differenzierungsformen der Schule sind in diesem Zusammenhang von Belang: Wenn bei einem System von Fachleistungskursen immer nur die in der (normalen) Zensurenverteilung besten Schüler in den nächst höheren Kurs aufgestuft werden, müssen gerade die guten Schüler daran Interesse haben, nicht miteinander oder mit anderen zusammenzuarbeiten, weil sie sonst ihre eigene Position gefährden würden. Ein solches Differenzierungs- und Zensurensystem muß daher jede Kooperationsbereitschaft im Keime ersticken¹⁸.

Aus den wenigen, hier angedeuteten Argumenten wird bereits deutlich, daß die Formen und Verfahren der pädagogischen Diagnostik nicht isoliert von sonstigen Faktoren der Schul- und Unterrichtsgestaltung betrachtet werden können. Die Erprobung neuer Modelle der pädagogischen Diagnostik wird daher nur in sorgfältiger Abstimmung auf und im Zusammenhang mit anderen schulischen Faktoren, das heißt also in Form von vergleichsweise radikalen, vieldimensionalen Schulversuchen erfolgen können.

Unverkennbar handelt es sich hier insgesamt um einen ebenso wichtigen wie diffizilen Forschungsbereich, in welchem mit besonderer Sorgfalt auf die Entwicklung solcher Instrumente hingearbeitet werden muß, bei denen die Gefahr der Fehlinterpretation der gewonnenen Daten möglichst weitgehend reduziert ist. Auch wird man oft darauf verzichten müssen, für die Praxis brauchbare Verfahren zu konstruieren, weil ihre Verwendung sich entweder negativ auf den Lernprozeß auswirken würde oder weil ethische Bedenken ihren Gebrauch verbieten¹⁹.

13.2.2. Selektion, Klassifikation, Prognose

Man wird kaum fehlgehen in der Annahme, daß die gegenwärtig gebräuchliche pädagogische Diagnostik zu der oben (Abschn. 13.2.1.) beschriebenen Funktion, der Förderung des Lernens, nur selten und unsystematisch herangezogen wird, ja sogar, daß sie oft in einer Weise verwendet wird, die eher das Gegenteil, nämlich die Entmutigung des Lernenden, bewirkt. Um so verbreiteter ist die Verwendung von Zensuren und sonstigen Urteilen zum Zwecke der Selektion, Klassifikation und Prognose im Bildungswesen. Diese – zweite – Funktion der pädagogischen Diagnostik gilt es, im folgenden näher zu betrachten.

Die folgenreichste Verwendungsart von Urteilen über Schüler dürfte derzeit darin bestehen, daß mit ihrer Hilfe Ausleseentscheidungen getroffen werden. Die Auslese-

¹⁶ Vgl. BLOOM, 1971 (deutsch 1973).

¹⁷ Z. B. IPI, vgl. LINDVALL und BOLVIN, 1967.

¹⁸ Vgl. hierzu z. B. Gesamtschulen-Informationsdienst 3, 1972, 37 ff. (Berlin, Pädagogisches Zentrum).

¹⁹ Dies gilt beispielsweise für ein offensichtlich recht effektives Verfahren der Einstellungsmessung, die „Bogus Pipeline“, vgl. JONES und SIGALL, 1971.

situation ist dadurch charakterisiert, daß diejenigen Personen, die ein bestimmtes Merkmal in einer bestimmten Ausprägung besitzen, ausfindig gemacht und einem darauf bezogenen, speziellen treatment unterzogen werden sollen. Durch die Auslese werden zwei Gruppen geschaffen: die Ausgewählten und die Zurückgewiesenen. Die Güte der Entscheidung wird in der Regel nur an den Ausgewählten überprüft; die Fehlentscheidungen in bezug auf die Zurückgewiesenen bleiben unbeachtet. Selektionsvorgänge im Bildungswesen sind beispielsweise die Schulreifeprüfung, die Übergangsauslese für das Gymnasium oder die Selektion der Abiturienten bei Numerus-clausus-Fächern an der Universität. In allen diesen Fällen sind für die Aufgenommenen die Ausleseprüfung und das nachfolgende treatment eng aufeinander bezogen, während für die Abgewiesenen lediglich die negative Bestimmung getroffen wird, daß das gewünschte treatment nicht in Frage kommt.

Im Unterschied zur Auslese zeichnet sich die Klassifikation dadurch aus, daß kein Schüler aufgrund der Diagnose zurückgewiesen, sondern *jeder* dem ihm entsprechenden treatment zugeführt wird. Beispiele für Klassifikation sind alle Formen von Beratung, etwa Berufsberatung oder Schullaufbahnberatung, für welche dem Ratsuchenden Vorschläge unterbreitet werden, die der bei ihm festgestellten Ausprägung der relevanten Merkmale angemessen sind und aufgrund derer er dann eine Entscheidung mit kalkulierbarem Risiko treffen kann; ein Klassifikationsvorgang ist — zumindest der Idee nach — auch die Einteilung der Schüler in Fachleistungskurse in der Gesamtschule oder die Option für Wahlfächer. Auch hier geht es darum, für *sämtliche* betroffenen Schüler das ihnen am besten entsprechende treatment ausfindig zu machen.

Soweit die Urteile über die Schüler nicht zu einem auf die Bedürfnisse aller Betroffenen speziell zugeschnittenen treatment führen, ist es angebracht, eher von Selektion als von Klassifikation zu sprechen. Dies gilt beispielsweise für die Entscheidung über das Sitzenbleiben eines Schülers aufgrund seiner Zensuren in zwei oder drei Fächern; die Wiederholung der Klasse bedeutet für ihn nicht, daß er einem treatment zugeführt wird, welches auf seine Leistungsausfälle auch nur versuchsweise abgestimmt wäre. Sitzenbleiben läßt sich insofern als ein Fall von Selektion betrachten.

Aufgrund der unvollkommenen Reliabilität und Validität aller in der pädagogischen Diagnostik gebräuchlichen und verwendbaren Beurteilungsverfahren sind Fehler bei der Gruppierung von Schülern, sei es zu Selektions-, sei es zu Klassifikationszwecken, unvermeidlich. Die bei der Selektion entstehenden Fehler lassen sich in zwei Typen unterteilen: A) unrichtige positive Auslese von fälschlicherweise als Merkmalsträger diagnostizierten Schülern, B) unrichtige negative Auslese von Schülern, die zurückgewiesen werden, obwohl sie Merkmalsträger sind. Am Beispiel des Numerus clausus für die Hochschulen heißt dies, daß einerseits eine Reihe von Abiturienten zugelassen werden, die nach den Normen der Hochschule nicht zugelassen werden dürften, und daß andererseits zahlreiche potentiell erfolgreiche Studenten abgewiesen werden.

Die Größe der bei der Selektion entstehenden Fehlerquote hängt von der Sicherheit ab, mit der das Kriteriumverhalten prognostiziert werden kann. In der um-

fangreichen Forschung, die zu dieser Frage vorliegt (insbesondere zur Übergangsauslese für die Sekundarschule), sind durchweg niedrige Validitätskoeffizienten ermittelt worden, so daß man in der Regel mit hohen Fehlerquoten rechnen muß. Bei derartigen Untersuchungen hat sich insbesondere die Definition und Messung des Kriteriumverhaltens als problematisch erwiesen. So ist beispielsweise die prognostische Validität von Abschlusssensuren der Schule meist an der Studienleistung nach ein oder zwei Jahren, allenfalls an den Examensensuren bei Abschluß des Studiums bestimmt worden, obwohl es keinesfalls ausgemacht ist, daß diese wiederum mit der Qualität der späteren Berufsleistung in Beziehung stehen. Gerade darauf aber müßte sich die Prognose beziehen.

Fehler bei der Auslese lassen sich nur bei den Aufgenommenen aufgrund ihrer Bewährung korrigieren. Dies geschieht gewöhnlich dadurch, daß weitere Schüler von dem Unterricht, für den sie nicht geeignet zu sein scheinen, ausgeschlossen werden, ohne daß die für sie passenden treatments im einzelnen bedacht und angewendet werden. Im Unterschied dazu besteht bei Klassifikationsentscheidungen die Möglichkeit, Fehlzuweisungen bei allen Schülern zu korrigieren und jedem einzelnen das ihm gemäße treatment zukommen zu lassen. Die hierzu geeignete Organisationsform ist, zumindest der Idee nach, die integrierte Gesamtschule, da dort entweder in Form breiter und variabler Differenzierung differentielle treatments angeboten werden, die durch eine geeignete pädagogische Diagnostik genutzt und zudem auf die Bedürfnisse der Schüler abgestimmt werden können, oder individualisierende Unterrichtsformen in heterogenen Schülergruppen entwickelt und praktiziert werden.

Forschungsschwerpunkte

a) Schon aus den wenigen, oben skizzierten Argumenten sowie in Anbetracht der Forschungslage läßt sich der Schluß ziehen, daß es nicht sinnvoll wäre, die Forschung und Entwicklung über Probleme der Selektion in der Schule zu intensivieren. Selbst in einem stabilen Schulsystem und unter Ausschöpfung aller Möglichkeiten²⁰ ergibt sich noch immer eine so hohe Quote von Fehlentscheidungen, daß nur durch Veränderung der Schulorganisation Abhilfe geschaffen werden kann, sei es mittels geplanter Strukturänderungen, sei es als Folge des Versuchs, jeglichen Fehler vom Typ B durch Herabsetzung der Mindestleistungen bei der Aufnahme in einen selektiven Schultyp zu vermeiden²¹. Eine perfekte Prognose der Entwicklung

²⁰ Vgl. z. B. VERNONS (1957) Ausführungen über die Übergangsauslese zur Sekundarschule im englischen Schulsystem seit dem Education Act von 1944.

²¹ Ein solcher Versuch führt dazu, daß die Aufnahmequote drastisch erhöht werden muß (bei der Übergangsauslese zum Gymnasium beispielsweise von ca. 25 Prozent auf ca. 70 Prozent) und sich als Konsequenz alle Probleme des Unterrichts in einer nahezu unausgelesenen Population stellen (vgl. HOPF, 1970). — In den USA ist zu beobachten, daß Tests durch den Übergang der Mehrheit der Schüler aufs College ihre ursprüngliche Selektionsfunktion allmählich verlieren und statt dessen zunehmend der Klassifikation dienen. Aufgrund von Beratung getroffene Entscheidungen beziehen sich nun auf die Auswahl unter verschiedenen Programmen. Dies hat beträchtliche Folgen für die Verwendungsweise der Instrumente. Vgl. z. B. NOVICK, 1971.

von Kindern und Jugendlichen wird es aus vielerlei Gründen auch in Zukunft nicht geben; von den Bedürfnissen und den Zielen des Schulsystems und der Schüler her gesehen bedarf es daher nicht der verstärkten Forschung und Entwicklung von Verfahren und Modellen der pädagogischen Diagnostik zum Zwecke der Selektion, sondern vielmehr zum Zwecke der Klassifikation im Zusammenhang eines differenzierenden und individualisierenden Unterrichts. Differenzierung und Individualisierung dienen dabei nicht nur der Absicht, das kognitive Lernen zu fördern beziehungsweise dem Lehrer die Erleichterungen zu verschaffen, die mit dem Unterricht in relativ homogenen Gruppierungen verbunden sind, sondern beispielsweise auch der Verbesserung des Gruppenklimas oder der Interaktion und Kommunikation zwischen Lehrer und Schüler²².

Die hier aufgeführten Argumente, welche gegen die derzeitige Form der Selektion sprechen, unterstellen freilich nicht, daß das Bildungssystem insgesamt völlig ohne Ausleseentscheidungen auskommen könnte: Spätestens nach dem Verlassen der Schule stehen Selektionsentscheidungen an, und sie werden wohl für absehbare Zeit auch nicht durch ausgedehnte Beratung und Betreuung in Klassifikationsentscheidungen umzuwandeln sein. Dagegen müßte es während der Schulzeit durchaus möglich sein, auf Selektion zu verzichten, indem man beispielsweise davon abgeht, heterogene Lerngruppen im Frontalunterricht lernen zu lassen und diejenigen auszusondern, die dabei aufgrund der unzureichend individualisierten Angebote in Lernschwierigkeiten geraten, sondern indem man – etwa mit Hilfe des „mastery learning“ – jedem die Möglichkeit gibt, die ihm gemäßen Ziele auch zu erreichen, und dann erst auf der Grundlage der Kenntnisse und Fähigkeiten, die er unter *diesen* Bedingungen erworben hat, über den weiteren Bildungsweg eine Entscheidung herbeiführt.

Zur Klassifikation benötigt man eine pädagogische Diagnostik, mit deren Hilfe es möglich ist, den sachstrukturellen Entwicklungsstand und sonstige, für die Ziele des Unterrichts und der Lernprozesse wichtige Voraussetzungen beim Schüler zu erfassen sowie die wesentlichen Merkmale der treatments zu bestimmen, die zu den jeweiligen Schülermerkmalen passen – eine naheliegende Forderung, die allerdings umfangreiche Untersuchungen über Zusammenhänge zwischen Lernvoraussetzungen und treatments impliziert²³.

Darüber hinaus benötigt man Verfahren, welche geeignet sind, die an verschiedenen Punkten der Entwicklung eines Schülers gewonnenen Diagnosen aufeinander zu beziehen. Es gibt bislang in Deutschland kein Instrument, das dieser Forderung genügt²⁴.

Die Anforderungen an die Instrumente der pädagogischen Diagnostik und an die Modelle der Interpretation und Gewichtung von Befunden stellen sich je nach

²² Vgl. z. B. THELENS, 1967, Vorschlag der Gruppierung nach „teachability“.

²³ Insofern können auch die bisher verwendeten Differenzierungsformen nicht als befriedigend gelten, da dort lediglich nach Inhalten oder nach „Niveau“ unterschieden wird, auf die Passung beispielsweise zwischen Unterrichtsstrategien, Medien usw. einerseits und Lernformen, die der einzelne Schüler bevorzugt, andererseits, keine Aufmerksamkeit verwendet wird.

²⁴ Als ein mögliches Modell solcher Verfahren könnten die Sequential Tests of Educational Progress (STEP) dienen. Vgl. zu diesen und anderen Problemen BOLDT, 1972.

Typ der Differenzierung und Individualisierung verschieden. Individualisierung, wie sie zum Beispiel beim programmierten Unterricht erfolgt, ist auf eine andere pädagogische Diagnostik angewiesen²⁵ als eine auf die Freisetzung des vollen menschlichen Potentials jedes einzelnen²⁶ gerichtete Individualisierung, bei welcher es kein vorgeplantes „terminal behavior“ und keine externen Standards geben kann, auf welche die pädagogische Diagnostik bezogen ist. Ähnliches gilt für die Differenzierung: Eine Gruppierung der Schüler auf der Grundlage spezieller Leistungsbereiche und Leistungsdimensionen innerhalb eines Unterrichtsfaches²⁷ ist auf andere diagnostische Hilfen angewiesen als eine flexible Unterrichtsorganisation, wie sie sich beispielsweise an den „integrated days“ englischer Primary Schools beobachten läßt²⁸, die sich unter anderem durch ein hohes Maß an Selbstbestimmung und an Selbstbeurteilung der Schüler auszeichnet und nur auf ein Minimum an diagnostischen Verfahren angewiesen zu sein scheint. Vor allem durch die Analysen ROEDERS ist im übrigen deutlich geworden, daß die herkömmlichen Zensuren die differenzierungs- (und damit lern-)relevanten Strukturen der Lerninhalte überdecken und somit keine geeigneten Verfahren darstellen, mit deren Hilfe man Lernen fördern und Ausfälle gezielt angehen könnte. Für die Bedürfnisse der Gesamtschulen bedarf es hier zunächst der Entwicklung einer pädagogischen Diagnostik, die in den sich verändernden Differenzierungsformen Verwendung finden kann sowie auf Leistungsbereiche und Leistungsdimensionen beziehbar ist.

b) Neben der Differenzierung und Individualisierung stellt die Schülerberatung – hier vor allem die Schullaufbahnberatung und die Berufsberatung²⁹ – einen weiteren Fall von Klassifikation dar, für den eine brauchbare pädagogische Diagnostik erst noch geschaffen werden muß.

Unter dem in diesem Gutachten zur Diskussion stehenden Gesichtspunkt läßt sich die Beratung als ein Vorgang beschreiben, bei welchem die Ergebnisse einer Diagnose mit den verfügbaren Informationen über die Erfordernisse des Zielverhaltens in Form einer Prognose zueinander in Beziehung gesetzt werden.

Hierzu gilt es zunächst, diejenigen Schülermerkmale zu bestimmen, die für die Prognose des Kriteriumverhaltens relevant sind, beispielsweise Interessen, Fähigkeiten, Kenntnisse und Fertigkeiten. Besonders im Bereich der Fähigkeiten und Schulleistungen wäre es wichtig, die nach wie vor offene Frage der Gewinnung und Interpretation von Profilen einer Lösung näher zu bringen. Dringlich wäre ebenfalls die Erforschung der im Beratungszusammenhang auftretenden Probleme bei der Verwendung lernzielspezifischer Verfahren.

Die Diagnose der Schülermerkmale läßt sich für die Beratung erst dann fruchtbar machen, wenn man genaue Kenntnisse über die Charakteristika derjenigen Situationen besitzt, auf die hin die Beratung erfolgen soll, sowie über die Passung dieser beiden Datensätzen. Für die Schullaufbahnberatung bedeutet dies beispielsweise,

²⁵ Vgl. z. B. LINDVALL und BOLVIN, 1967.

²⁶ MCCLELLAN, 1972, 171, mit weiterer Erörterung dieser Unterscheidung und ihrer Implikationen.

²⁷ Vgl. ROEDER, 1974.

²⁸ Vgl. z. B. COOK und MACK, 1971.

²⁹ Zur Beratung bei psychosozialen Konflikten vgl. das Gutachten FRIEDRICH.

daß Informationen über die verfügbaren treatments existieren müssen, so daß der Berater weiß, was in unterschiedlichen Kursen oder Institutionen gefordert und angeboten wird, welche Lehr- und Lernmethoden dort üblich sind etc.³⁰. Für die Berufsberatung bedarf es genauer Kenntnisse der erforderlichen (und sich wandelnden) Qualifikationen beziehungsweise der zentralen Merkmale des Kriteriumverhaltens. Die sich hieraus herleitenden Anforderungen an eine brauchbare pädagogische Diagnostik sind nicht gering: Unter der plausiblen Hypothese einer *aptitude-treatment-interaction* bedarf es, wie schon in anderem Zusammenhang beschrieben, der Erforschung der treatment-Merkmale und der Entwicklung sonstiger „Umweltmaße“, die valide, verlässlich und zugleich praktikabel sind, sowie der Bereitstellung von Informationen über Berufsqualifikationen und Merkmale des Kriteriumverhaltens, auf deren Grundlage eine Beratung erfolgen kann. Die Misere der derzeit gebräuchlichen Berufseignungs- und -interessentests dürfte vor allem auf den Mangel an Kenntnissen der genannten Art zurückzuführen sein³¹; die Untauglichkeit der herkömmlichen Zensuren für über die Schule hinausgehende Prognosen hat sich ebenfalls allenthalben gezeigt: „Grades seem only to predict grades, nothing else.“³² Wichtige Forschungsaufgaben für die pädagogische Diagnostik stellen sich schließlich hinsichtlich der Probleme, die bei der Zusammenführung von Schülerdaten und Informationen über den Zielbereich der Prognose entstehen. In der Testforschung hat sich beispielsweise gezeigt³³, daß diese Instrumente den Erfolg eines Schülers in verschiedenen Colleges gerade nicht generell voraussagen, sondern eine nach treatments differenzierende Prognose gestellt werden muß. Für die Zwecke der Beratung genügt daher keinesfalls die Ermittlung eines einfachen Koeffizienten der prognostischen Gültigkeit eines Instruments, sondern es bedarf zumindest der Konstruktion von Validitäts-Matrizen, aus denen die Sicherheit von Prognosen verschiedener Subpopulationen für verschiedene „Umwelten“ beziehungsweise unterschiedliches Kriteriumverhalten abgeschätzt werden kann³⁴. Auch dann weiß man freilich noch nichts über diejenigen Merkmale solcher „Umwelten“, auf welche die Unterschiede zwischen den Validitätsindizes zurückzuführen sind, sondern findet zunächst nur „blinde“, korrelative Zusammenhänge von Prädiktor und Kriterium. Immerhin wäre dies ein erster Schritt auf dem Wege, Passung zwischen Lernvoraussetzungen und treatments herzustellen.

³⁰ In den USA nimmt man bei der Beratung längst auf die Tatsache Rücksicht, daß die verschiedenen Colleges ganz unterschiedliche Lernumwelten darstellen (vgl. hierzu z. B. ETS, Annual Report 1969—1971). Dies gilt in Deutschland vermutlich nicht weniger: So stellt es sozusagen das Geheimwissen eines Erziehungsberaters dar, die teilweise äußerst unterschiedlichen Anforderungen von Schulen innerhalb derselben Stadt, ja desselben Bezirks zu kennen und bei seiner Beratung zu berücksichtigen. Auch ist es eine Fiktion zu glauben, die verschiedenen Universitäten erforderten ein gleiches Fähigkeits- und Interessenprofil für ein erfolgreiches Studium; beispielsweise dürften die Erfolgchancen eines Psychologiestudenten mit einer bestimmten individuellen Ausprägung von Fähigkeiten, Interessen etc. sehr unterschiedlich aussehen, je nachdem, ob er in Bodum oder Berlin, in Hamburg oder Freiburg studiert.

³¹ Vgl. auch Proceedings of the 1971 Invitational Conference on Testing Problems, Princeton, 1972, S. 119 ff.

³² Vgl. GOODLAD, 1972, S. 10.

³³ Vgl. z. B. CRONBACH und SUPPES, 1969, 85 f.

³⁴ Vgl. hierzu auch SHULMAN, 1970.

c) Die zahlreichen Untersuchungen über die Konsequenzen der Selektion im Bildungswesen haben überzeugend deutlich gemacht, daß jede Selektion, aber auch manche Formen der Klassifikation (zum Beispiel setting in der Gesamtschule) Ungleichheiten in der Gruppenzusammensetzung mit sich bringen, die meist in der Unterrepräsentation von Schülern aus niedrigen Sozialschichten oder Angehörigen sonstiger Minoritäten in den lernschnelleren Gruppen bestehen.

Versuche, solche Ungleichheiten durch spezielle Verfahren der pädagogischen Diagnostik zu mildern, gibt es seit langem; aber weder der Versuch, sogenannte „*culture-free tests*“ zu konstruieren, noch die Verwendung vergleichsweise objektiver und auf die Messung allgemeiner kognitiver Fähigkeiten gerichteter Instrumente anstelle subjektiver Beurteilungsverfahren³⁵ hat zu befriedigenden Resultaten geführt. Neue Lösungsansätze müssen daher in andere Richtungen gehen: Einerseits bedarf es erheblicher Forschungsanstrengungen, um die Determinanten der im Durchschnitt geringeren Lernleistungen zum Beispiel von Schülern aus niedrigeren Sozialschichten zu bestimmen; allein die Tatsache, daß auch Kinder anderer Schichten, wenngleich seltener, ähnliche Ausfälle zeigen, läßt darauf schließen, daß die verfügbaren Schichtindizes äußerst grobe und für kompensatorische Bemühungen völlig unzureichende Ausgangsdaten darstellen. Die Entwicklung aussagekräftiger, das heißt differenzierter „Umweltmaße“ zur Bestimmung derjenigen Faktoren, die ein Kind zu einem „benachteiligten“ Schüler werden lassen³⁶, ist daher ein vordringliches Anliegen, wenn die Diagnose gezielte Förderungsmaßnahmen zur Folge haben soll.

Der Absicht, benachteiligten Schülern besondere, auf sie zugeschnittene Hilfen anzubieten, dürfte es andererseits eher dienlich sein, diagnostische Verfahren zu entwickeln, die bestehende Lerndefizite nicht überdecken (wie es die kulturunabhängigen Tests tendenziell tun), sondern diese im Gegenteil klar herauszuarbeiten, so daß Maßnahmen einer „positiven Diskriminierung“³⁷ möglich werden. Die Weiterentwicklung einer auf die Identifizierung individueller Unterschiede gerichteten pädagogischen Diagnostik sowie ihrer psychologischen Grundlagen ist daher auch aus diesem Grunde dringlich³⁸.

Für die vorerst unvermeidbaren Fälle von Auslese innerhalb und außerhalb des Schulsystems ist der Versuch, Ungleichheiten abzubauen, aber nicht nur auf die Erforschung der Ursachen von Lerndefiziten und auf die Konstruktion geeigneter Verfahren, sondern auch auf die Entwicklung und Erprobung neuer Modelle der Datengewichtung und Dateninterpretation angewiesen. Abgesehen von Modellen, die von a priori getroffenen Entscheidungen ausgehen — beispielsweise der Aufnahme gleicher Proportionen aus allen Bewerbergruppen, also auch aus Minoritäten — gibt es sehr unterschiedliche Möglichkeiten, individuelle und institutionelle Interessen bei

³⁵ Vgl. z. B. FLOUND und HALSEY, 1963, und HOLTZMAN, 1971, 546. Durch die Verwendung objektiver Tests ist allerdings erst der Nachweis der beträchtlichen Überlappung der Testergebnisse verschiedener Bevölkerungsgruppen, in USA z. B. ethnischer Minoritäten, geglückt.

³⁶ Vgl. SHULMAN, 1970. Die erhebliche Verbesserung von Prognosen auf der Grundlage solcher Umweltmaße hat u. a. WOLF, 1964, demonstriert. Vgl. auch OEVERMANN, 1969, 302 f.

³⁷ Vgl. HALSEY, 1972; FREEBERG, 1970; GOODLAD, 1972; HOLTZMAN, 1971, 551 f.

³⁸ Vgl. auch CRONBACH und SUPPES, 1969, 261.

der Auslese zu berücksichtigen. COLE (1972) hat eine Reihe möglicher Modelle diskutiert, mit dem derzeit vorwiegend verwendeten Regressionsmodell verglichen und drastische Unterschiede bezüglich der Berücksichtigung von Bewerbern festgestellt, die aus Gruppen mit traditionell niedriger Aufnahmequote stammen. Das von ihr vorgeschlagene „Equal Opportunity Model“ gibt auch Angehörigen von Minoritäten bessere Chancen, ohne die Erfolgsrate unter den Aufgenommenen gegenüber dem Regressionsmodell spürbar zu reduzieren; die individuellen Rechte werden dabei nicht mehr, wie üblich, den institutionellen Interessen völlig untergeordnet. Selbst bei Verwendung von Tests, die an der Majorität geeicht wurden, läßt sich also offenbar mit Hilfe eines geeigneten Modells der Datengewichtung vermeiden, daß Angehörigen von Minoritäten Nachteile entstehen. Angesichts der Tatsache, daß hierzulande zahllose Entscheidungen über Schülerschicksale gefällt werden, ohne daß die diesem Vorgang zugrunde liegenden Modelle je auf ihre Konsequenzen überprüft worden wären, müssen Forschungen in diesem Bereich als vorrangig bezeichnet werden. Zumindest der Anschein spricht dafür, daß derzeit sowohl beim Übergang zum Gymnasium wie beim Numerus clausus das nur die institutionellen Interessen berücksichtigende Regressionsmodell die Grundlage bildet.

d) Exkurs: Zum Numerus-clausus-Verfahren.

Wegen seiner großen Aktualität und Bedeutung sei in diesem Zusammenhang kurz auf die Problematik des Numerus-clausus-Verfahrens³⁹ eingegangen. Der Staatsvertrag über die Vergabe von Studienplätzen bestimmt, daß bei den dem Numerus clausus unterliegenden Studienfächern von den vorhandenen Studienplätzen 60% „nach Eignung und Leistung“⁴⁰ der Bewerber zu vergeben sind. Mit Eignung und Leistung ist dabei die Schulleistung gemeint, wie sie sich in der Durchschnittsnote niederschlägt, die gebildet wird „aus den Noten des Reifezeugnisses und den nicht im Reifezeugnis enthaltenen Noten der am Ende des 11. und 12. Schuljahres abgeschlossenen Fächer“⁴¹.

Ein solches Verfahren unterstellt eine hohe prognostische Validität der als Prädiktoren benutzten Zensuren. Daß diese Voraussagegültigkeit selbst innerhalb der Schule nicht gegeben ist, wurde oben bereits am Beispiel der Übergangsauslese für weiterführende Schulen erläutert⁴². Die Untersuchungen, die zur prognostischen Gültigkeit der Schulzensuren für die Leistungen in der Hochschule oder im Beruf unternommen worden sind, haben gezeigt, daß keine Zusammenhänge zwischen Zensuren und Berufserfolg bestehen und auch die Korrelationen zwischen Zensuren und Studienerfolg niedrig liegen. Insoweit das Studium keinen Selbstzweck darstellt, kann bei der Ausleseentscheidung nur das Kriterium Berufserfolg eine Rolle spielen. Denn nicht diejenigen Bewerber sollten z. B. zum Medizinstudium zugelassen werden, die gute Examensnoten erreichen, sondern die gute Ärzte sein werden,

³⁹ Staatsvertrag über die Vergabe von Studienplätzen vom 20. 10. 1972. Vgl. hierzu auch HOPF, 1974, mit weiterer Literatur.

⁴⁰ Anhang A, Nr. 1.1.1.

⁴¹ Anhang A, Nr. 2.2.

⁴² Vgl. auch unten, Kap. 13.3.

und der Studierenerfolg kann erst dann als Kriterium dienen, wenn nachweislich ein enger Zusammenhang zwischen Studien- und Berufserfolg besteht.

Der Nachweis solcher Zusammenhänge trifft freilich auf große Schwierigkeiten, da es strittig sein muß, welcher Indikator für den Berufserfolg stehen kann (z. B. Einkommen oder Berufszufriedenheit oder die Beurteilung durch Vorgesetzte bzw. Kollegen) und zu welchem Zeitpunkt der Berufserfolg eingeschätzt werden sollte. Auch ist es nicht vertretbar, Berufsgruppen (z. B. Juristen, Mediziner) undifferenziert zu betrachten, da offensichtlich recht unterschiedliche Qualifikationen erfordert werden, je nachdem ob jemand Psychiater oder Pathologe, Jugendrichter oder Patentanwalt ist⁴³. Daß zwischen Durchschnittszensuren in Sozialkunde, Latein, Physik etc. und komplexen, sich wandelnden Berufsleistungen – die oft ganz andere Qualifikationen erfordern, als in der Schule gefördert und bewertet werden – kein Zusammenhang gefunden wurde, ist nicht weiter verwunderlich.

Zwei Forschungsprobleme zeichnen sich in dieser Situation ab: Einerseits müssen Anstrengungen unternommen werden, die Auswahl unter den Bewerbern nach Gesichtspunkten und mit Hilfe von Verfahren vorzunehmen, die mit dem (zu definierenden) Kriteriumverhalten in einem engen Zusammenhang stehen sowie den Veränderungen im Kriteriumverhalten möglichst verzögerungsfrei folgen. Beispielsweise könnte der Versuch unternommen werden, an der *spezifischen* Bewährung der Bewerber in einem in sich sinnvollen Kurzstudium über Eignung und Motivation Aufschlüsse zu gewinnen, welche in bezug auf prognostische Gültigkeit den Schulzensuren gewiß überlegen wären (sofern gesichert ist, daß die Studieninhalte mit dem Kriteriumverhalten in Beziehung stehen). Hochschuleingangstests würden dagegen entweder ungünstige, normierende Auswirkungen auf die vorhergehende Schulzeit haben oder, sofern sie stärker auf die Erfassung von „Intelligenz“ gerichtet wären, eine Fülle sonstiger unerwünschter Folgewirkungen mit sich bringen⁴⁴.

Andererseits ist es angezeigt, die aus anderen Selektionssituationen bekannten Rückwirkungen der Auslese auf die vorhergehenden Schuljahre genauer zu untersuchen. Soweit man sehen kann, bestimmen nämlich Prüfungen oder das den Leistungsurteilen zugemessene Gewicht nachhaltig beispielsweise die Inhalte und Methoden des Unterrichts, die Interaktionsformen zwischen Schülern und Lehrern, die Kooperationsbereitschaft der Schüler untereinander, ihre Motivation zur Mitwirkung in der Selbstverwaltung oder die Bereitschaft der Lehrer zum Experimentieren mit neuen Curricula oder neuen Formen der Unterrichtsorganisation. Das Numerus-clausus-Verfahren in seiner derzeitigen Form wird daher mit hoher Wahrscheinlichkeit beklagenswerte Auswirkungen auf die letzten Schuljahre, ja sogar auf die Sekundarstufe I haben, da auch die Abschlußzensuren der Klasse 11 und 12 in die Berechnung der Durchschnittsnote eingehen. Man muß daher davon ausgehen, daß das Verfahren nicht nur den Zweck verfehlt, dem es dienen soll, sondern überdies unübersehbare Folgeschäden zeitigen wird, und sei es auch nur durch die zu erwartenden, negativen Rückwirkungen auf die Schule.

⁴³ Vgl. hierzu HOYT, 1965.

⁴⁴ Vgl. z. B. McCLELLAN, 1973.

Daß auch in diesem Bereich intensiv darauf hingearbeitet werden sollte, an die Stelle von Selektion Klassifikation treten zu lassen⁴⁵, daß also beispielsweise auch für den nicht studierenden Abiturienten passende Alternativen verfügbar sein und entsprechende Beratung und Bewährungsmöglichkeiten bestehen sollten, braucht kaum eigens betont zu werden. Freilich liegt eine solche Entwicklung noch in weiter Ferne – im Unterschied zu den Möglichkeiten, innerhalb der Schule, vor allem der Sekundarstufe I, auf Selektion kurzfristig zu verzichten, ohne dabei den Grundsatz der individuellen Förderung aufzugeben.

13.2.3. Information

Auch die Informationsfunktion der pädagogischen Diagnostik (13.2.3) sowie ihre Rolle bei der sozialwissenschaftlichen Forschung und Entwicklung (13.2.4) können im folgenden nur kurz diskutiert werden. Die Darstellung wird sich dabei vor allem auf bisher unerwähnt gebliebene Probleme, beispielsweise auf einige eher verfahrenstechnische Aspekte, konzentrieren.

Aus dem bisher Gesagten geht bereits hervor, daß eine wichtige Aufgabe der pädagogischen Diagnostik in der Information der Schüler und Lehrer über Lernergebnisse sowie über Merkmale des treatment und der weiteren Lernumwelt besteht, sei es zur Orientierung über den Lernverlauf, über Lerndefizite oder das Erreichen eines Kriteriums, sei es als Grundlage für Differenzierungsentscheidungen, sei es zur Verbesserung der Planung und Durchführung des Unterrichts.

Neben Schülern und Lehrern besitzen auch die Eltern und „externe“ Adressaten, beispielsweise aufnehmende Klassen oder Schulen, Arbeitgeber etc., ein legitimes Interesse daran, zumindest über bestimmte Aspekte des Schülerverhaltens informiert zu werden⁴⁶. Wer immer aber der Empfänger von diagnostischen Befunden sein mag, stets kommt es darauf an, daß die Informationen möglichst verläßlich, valide und für den jeweiligen Zweck relevant sind. Hier durch Forschung und Entwicklung den gegenwärtigen Stand der pädagogischen Diagnostik zu verbessern, dürfte vor allem im Interesse der Schüler selbst liegen.

So wäre es beispielsweise vordringlich, über den Prozeß, die Validität und die Konsequenzen der Selbstbeurteilung von Schülern im Verlauf des Lernvorganges Untersuchungen anzustellen. Dies folgt einerseits aus der Tatsache, daß die Diagnose selbst weniger Variablen des oben in Abschn. 13.2.1.1. diskutierten Lernmodells den Lehrer rasch an die Grenzen seiner Kapazität bringen und eher nachteilige Effekte auf den Unterricht ausüben würde. Andererseits läßt sich beobachten, wie Kinder bei flexibler Unterrichtsorganisation und einem hohen Maß an Selbstbestimmung ihrer Tätigkeiten⁴⁷ sehr wohl in der Lage zu sein scheinen, ohne Zuhilfenahme formalisierter diagnostischer Verfahren aus den ihnen angebotenen Aufgaben und

⁴⁵ Vgl. zur Entwicklung in den USA oben Anm. 21 und S. 238.

⁴⁶ Vgl. zu diesem Kapitel besonders SCHREINER, 1970.

⁴⁷ Vgl. z. B. COOK und MACK, 1971.

Materialien genau diejenigen herauszusuchen, die ihren Vorkenntnissen, ihrer Motivationslage und ihren Fähigkeiten gleichermaßen entsprechen. Möglichkeiten und Grenzen der Selbsteinschätzung der Schüler sowie Art und Ausmaß der dabei wirksam werdenden Urteilsverzerrungen sind weitgehend unbekannt; ihre systematische Aufklärung könnte weitreichende Folgen für Lernen und Unterricht haben.

Im Zusammenhang der Diskussion über Selektion, Klassifikation und Prognose (Abschn. 13.2.2.) war bereits die Notwendigkeit der Bestimmung differentieller Validitätsindizes festgestellt worden. So hat sich beispielsweise gezeigt, daß die prognostische Gültigkeit von College-Eingangstests in den USA in sehr weiten Grenzen schwankt, je nach dem College, für das der Test den Erfolg eines Studenten voraussagen soll⁴⁸. Auch ist es ein geläufiger Befund, daß ein bestimmtes diagnostisches Instrument nur für bestimmte, nicht aber für andere Bevölkerungsteile aussagefähig ist⁴⁹. Die Interpretation von Schülerbeurteilungen auf der Grundlage eines globalen Validitätskoeffizienten führt daher notgedrungen zu zahlreichen Fehlentscheidungen. Da die herkömmlichen Formen der pädagogischen Diagnostik diesen Anforderungen nicht genügen, bedarf es der Entwicklung (und ständigen Revision) von Validitäts-Matrizen, die zwischen Populationen und zwischen Situationen (beziehungsweise Institutionen), auf welche die Prognosen zielen, unterscheiden. Auch gilt es, für die Interpretation von diagnostischen Befunden, zum Beispiel von Testergebnissen, differenzierte Normen bereitzustellen, wobei zugleich die Problematik übergreifender Standards dem Interpretieren nicht vorenthalten werden darf. Kriterienbezogene Normen, die sich auf Grundkenntnisse und Grundfertigkeiten erstrecken, dürften im Zuge der Curriculumrevision und der Ausbreitung des „Mastery Learning“ an Bedeutung gewinnen; auch hier sind wichtige Fragen bislang ungelöst.

Neben der prognostischen Validität sei hier auch auf die Inhaltsgültigkeit hingewiesen, die gerade für die pädagogische Diagnostik von besonderer Bedeutung ist. Offene Probleme bestehen hier nicht nur darin, zwischen Unterricht und Diagnoseverfahren Passung herzustellen, sondern, worauf ANDERSON (1972, 161) aufmerksam gemacht hat, in einem strengeren Sinne inhaltsgültige Verfahren überhaupt erst einmal zu entwickeln, beispielsweise also zu klären, welche kognitiven Prozesse bei der Lösung von Testaufgaben, die nach Ansicht ihrer Konstrukteure der Prüfung eines bestimmten Lernzieles dienen, ablaufen. Auch läßt sich der einleuchtende Vorschlag, mit Hilfe noch zu konstruierender Regeln ein „Universum“ von

⁴⁸ Vgl. z. B. CRONBACH und SUPPES, 1969, 82 ff.

⁴⁹ SHULMAN, 1970, 380, berichtet beispielsweise, daß der Stanford-Binet zwar für weiße, geistig zurückgebliebene Probanden, nicht aber für schwarze, Gültigkeit besitzt. Die in Deutschland verfügbaren Tests geben über solche Fragen in der Regel keine Auskunft; selbst bei den geläufigsten Intelligenztests weiß man nur zum Teil, was ihre Ergebnisse z. B. in speziellen Gruppen behinderter Kinder bedeuten bzw. ob sie dort überhaupt verwendbar sind. HOLTZMAN, 1971, S. 551 f., weist auf die neuerdings unternommenen Versuche hin, bei denen auf die sozialen, kulturellen und linguistischen Spezifika von Subkulturen bei der Konstruktion der Untersuchungsverfahren besondere Rücksicht genommen wird (vgl. z. B. FREEBERG, 1970); er diskutiert Vorzüge und Probleme dieses Ansatzes, der auch in Deutschland weiter verfolgt werden sollte, z. B. im Zusammenhang der dringend gebotenen Entwicklung einer für die Gastarbeiterkinder angemessenen pädagogischen Diagnostik.

Items für einen gegebenen Bereich bereitzustellen, um daraus durch Ziehung von Zufallsstichproben inhaltsgültige diagnostische Verfahren herzustellen, erst nach umfangreicher Forschungs- und Entwicklungsarbeit, die beispielsweise auf die Konstruktion einer umfassenden, allgemein zugänglichen Itembank gerichtet sein könnte, verwirklichen.

Befunde der pädagogischen Diagnostik, die nur eine Momentaufnahme des bestehenden Zustandes wiedergeben, stellen unzureichende Informationen für eine sinnvolle Beeinflussung des Lernens, für eine Beratung oder eine Entscheidung bereit. Nicht weniger wichtig als der status quo eines Schülers oder gar dessen Vergleich mit einer Bezugsgruppe dürften Kenntnisse darüber sein, unter welchen Bedingungen ein bestimmtes Verhalten, zum Beispiel eine Lernleistung, zustande gekommen ist; ob sie einen Punkt auf einer steil ansteigenden oder einer ausschwingenden Lernkurve darstellt, das heißt also, wie sie sich in einer Längsschnittbetrachtung darstellt; welches ihr intraindividueller Stellenwert im Vergleich zu anderen Merkmalen ist etc. Die mit Profilanalysen, Zuwachsmessungen⁵⁰ und der treatment-Beurteilung verbundenen Probleme bedürfen daher dringend der weiteren Bearbeitung. Zu einer umfassenden Information über treatments oder – in weiterem Sinne – über die Merkmale von Lernumwelten – gehört dabei auch die Aufklärung des Schülers über die Eigenschaften derjenigen Situationen (Berufe, Bildungsgänge, Ausbildungsinstitutionen), auf die er sich einzulassen beabsichtigt. Denn der Schüler hat keinen geringeren Anspruch auf Information über die Merkmale der ihn erwartenden „Umwelt“, als diese Anspruch auf die Kenntnis seiner Individualdaten erhebt. Gerade angesichts der erheblichen Unterschiede zwischen den möglichen Alternativen, beispielsweise zwischen verschiedenen Wahlkursen in der Gesamtschule oder zwischen verschiedenen Universitäten, kann der Schüler eine informierte Entscheidung nur treffen, wenn er mehr als nur gerüchtweise über die Eigenschaften der Situation, die ihn erwartet, etwas erfährt. Die pädagogische Diagnostik ist auf den Gebieten der Berufs- und Bildungsberatung jedoch noch gänzlich unentwickelt. Nicht nur, daß keine Modelle oder Programme existieren, die durch Zusammenführung von individuellen und externen Daten die Grundlage für eine aufgeklärte Entscheidung durch den Schüler bereitstellen, sondern es fehlt überhaupt an Kenntnissen über die relevanten Merkmale von treatments und Lernumwelten sowie an Instrumenten, mit denen diese erfaßt werden könnten⁵¹.

Im Zusammenhang der Informationsfunktion der pädagogischen Diagnostik sei schließlich besonders auf die bislang unzureichenden Kenntnisse über die Mög-

⁵⁰ CHOPPIN, in: WOOD und SKURNIK, 1969, 134–140, hat im Zusammenhang der Entwicklung einer Itembank beispielsweise ein Itemanalyseverfahren vorgeschlagen, das es erlaubt, die Fortschritte eines Schülers während seiner ganzen Schulzeit auf eine einzige Skala zu beziehen und somit die Leistungsdaten direkt miteinander vergleichbar zu machen. Problematisch dürfte hierbei allerdings das Kriterium sein, an welchem die Daten verankert werden. Vgl. auch die vom Educational Testing Service durchgeführte National Test-Equating Study in Reading, über die JAEGER, ANGOFF, HANSEN, LORET und BIANCHINI beim AERA-Meeting 1973 berichtet haben.

⁵¹ Vgl. hierzu u. a. ETS Annual Report 1969–1971, 19 ff.; CRONBACH und SUPPES, 1969, 85 ff. Einige verstreute Vorarbeiten auf diesem Gebiet werden weiter unten in Abschnitt 13.4.4. erwähnt. Vgl. auch Abschn. 13.2.2 b.

lichkeiten einer verständlichen und angemessenen Interpretation der Urteile über Schüler sowie auf die mit steigender Präzision und Reichweite der Urteile wachsende Notwendigkeit des Schutzes der Daten vor mißbräuchlicher Benutzung hingewiesen.

Das Bedürfnis nach Absicherung gegen Fehlinterpretation wird besonders ausgeprägt sein, wenn die Urteile aus der Schule herausdringen und beispielsweise von anderen Ausbildungsinstitutionen oder von Arbeitgebern zur Kenntnis genommen werden und die Grundlage weitreichender und oft irreversibler Entscheidungen bilden. Der Verhütung von Fehlinterpretationen kann dabei schon die Beachtung längst bekannter psychometrischer Eigenschaften von Beurteilungen dienen. Beispielsweise müßten Mitteilungsförmlichkeiten erprobt werden, bei denen sich die Reliabilität des jeweiligen Beurteilungsverfahrens unmittelbar im Urteil niederschlägt; die übliche naive Art und Weise, in der Zensuren ohne die wegen ihrer geringen Reliabilität gebotene Relativierung gelesen und verwertet werden, bei der beispielsweise die 2 des einen Schülers als besser gilt als die 4 des von einer anderen Schule kommenden Schülers, bei welcher also der immerhin etwa \pm eine Zensur betragende Standardmeßfehler unbeachtet bleibt, dürfte schon viel Unheil angerichtet haben⁵². In ähnlicher Weise gilt es, auch andere Aspekte der Unsicherheit pädagogischer Diagnosen auszuschließen oder in die Urteile selbst eingehen zu lassen, so daß auch der naive Adressat vor groben Fehldeutungen bewahrt bleibt. So ließen sich, neben den genannten „Meßbändern“, entsprechende „Schätzbänder“ denken, welche die prognostische Validität widerspiegeln und bei der Beratung Verwendung finden könnten; die Inhaltsgültigkeit von Beurteilungsverfahren ließe sich durch Verzicht auf standardisierte Tests, an deren Stelle informelle Verfahren oder, besser, auf die spezifischen Bedürfnisse des Unterrichts zugeschnittene Aufgaben aus einer Itembank treten könnten, wesentlich verbessern; oder es könnten inhaltlich gefüllte, verständliche Beurteilungen die üblichen, nichtssagenden und zur Überinterpretation einladenden Ziffernzensuren ablösen, indem die Zeugnisse inhaltliche Angaben über die ganz oder teilweise bewältigten Lernziele enthielten (man beispielsweise also nicht erfahre, daß ein Schüler so gut wie der Durchschnitt seiner Klasse lesen kann, sondern daß er einen einfachen Illustriertentext mit einer Geschwindigkeit von 120 Wörtern pro Minute lesen und 80 Prozent des Inhalts korrekt weitergeben kann)⁵³. Welche Maßnahmen man aber auch ergreift, es wäre eine Illusion zu glauben, dadurch ließe sich die Fehlinterpretation der schulischen Urteile gänzlich verhindern; auch mit lernzielorientierten „Mastery-Tests“ oder mit Beurteilungen, die als Meßbänder und nicht als Punktwerte vorliegen, läßt sich rigide Selektion betreiben⁵⁴. Daher ist es dringend notwendig, nicht nur die Lehrer, sondern auch die Schüler, Eltern und sonstigen Adressaten über die Möglichkeiten und Grenzen der päd-

⁵² Die Veröffentlichung von Urteilen in Form von „Bändern“, nicht als Punktwerte, dürfte eine Möglichkeit darstellen, diesem Dilemma abzuhelfen.

⁵³ Die Vorzüge derartiger Urteile im Vergleich zu den üblichen, normenorientierten Aussagen lassen sich an den in rascher Folge publizierten Befunden der Studie National Assessment of the Educational Progress (NAEP) besonders gut erkennen.

⁵⁴ Vgl. hierzu z. B. HIERONYMUS, 1972, 62.

agogischen Diagnostik sowie über die wichtigsten Interpretationstechniken gründlich aufzuklären. Dies ist auch unter dem Gesichtspunkt geboten, daß mit zunehmender Verwissenschaftlichung der pädagogischen Diagnostik die Bereitschaft der Betroffenen abnehmen wird, die Urteile infrage zu stellen⁵⁵, wenn nicht durch ausführliche Information der Wissenschafts- und Testgläubigkeit der Boden entzogen wird⁵⁶.

Zugleich mit dem Bedürfnis nach Information über Schüler- und Umweltmerkmale gibt es ein Bedürfnis nach dem Schutz solcher Daten vor Mißbrauch. Je weiter sich die pädagogische Diagnostik entwickelt und je leichter es wird, mit Hilfe von Computern Informationen über Schüler, Lehrer und Institutionen rasch und umfassend abzurufen, desto dringender stellt sich die Frage nach der Legitimität des Zugriffs. Hier ist zunächst zu fordern, daß nur solche Daten zugänglich sein dürfen, die für den jeweiligen Zweck notwendig sind. Auch muß geklärt werden, welche Daten der jeweilige Adressat verarbeiten kann, und auch unter diesem Gesichtspunkt eine Beschränkung der Informationsweitergabe erfolgen. Darüber hinaus wäre es ein wichtiger Forschungsgegenstand, Prinzipien herauszuarbeiten, die den Zugang zu den Daten regeln, beispielsweise also begründen, welche Informationen nur für die Schüler, nicht aber die Eltern, welche für die Eltern, welche für die Schulbehörde⁵⁷, welche für die Bildungsforschung zugänglich sein sollen⁵⁸. Auch bedarf es der Klärung der Frage, ob zentralisierte Datensammlungen nicht unter Umständen besser gegen Mißbrauch zu sichern sind als die Archive der einzelnen Schule⁵⁹. Andere Vorschläge zielen auf die scharfe Kontrolle der Qualifikationen der Benutzer diagnostischer Verfahren; auf die Notwendigkeit zu verhindern, daß sich an die Testkonstruktion ökonomische Interessen heften (BLAKE, 1971), sowie auf die Aufstellung und kontrollierte Einhaltung von Verhaltens- und Konstruktionskodizes⁶⁰.

Forschungen zur Informationsfunktion der pädagogischen Diagnostik betreffen ein Feld, in welchem die Interdependenz des Gegenstandes mit anderen Faktoren besonders klar zutage tritt⁶¹. Daher dürfte es kaum möglich sein, einzelne Fragen isoliert zu untersuchen, sondern es ergibt sich die Notwendigkeit zu komplexen Schulversuchen, bei denen mehrere Faktoren zugleich variiert werden können. Begleituntersuchungen hätten dabei die Aufgabe, sowohl die außerschulischen Auswirkungen, beispielsweise auf Eltern, Arbeitgeber, weiterführende Bildungsinstitutionen, als auch die innerschulischen Effekte, beispielsweise auf die Lernmotivation der Schüler oder auf die Unterrichtsführung, in die Betrachtung einzubeziehen.

⁵⁵ Vgl. z. B. GOSLIN, 1971.

⁵⁶ So wäre es z. B. wichtig, in Erfahrung zu bringen, welche Folgen es hat, wenn man Schüler über die Tragfähigkeit und die zulässige Interpretation der Urteile unterrichtet, denen sie sich täglich ausgesetzt sehen. Vielleicht wäre dies ein erster Schritt zur Freisetzung des Orientierungswertes der Zensuren und zur Entlastung von ihrer Straf- und Belohnungsfunktion.

⁵⁷ Vgl. hierzu z. B. die Diskussion über die sogenannte accountability bei HIERONYMUS, 1972, 64 f.

⁵⁸ Vgl. z. B. GOSLIN, 1971; QUILTICH und SCHÖBER, 1972.

⁵⁹ HOLTZMAN, 1971, 550 f., berichtet über einen Großversuch zu dieser Frage, bei dem durch das sogenannte Link-System die Identifikation und die biografischen und sonstigen Daten der Studenten auf getrennten Datenbändern verzeichnet wurden und nur einer unabhängigen Vertragsfirma im Ausland die Zuordnung der beiden Sätze möglich war.

⁶⁰ Vgl. z. B. REGLEMENT, 1971; Educational Researcher, 2, 1973, No. 2, 3—5; FINDLEY, 1968.

⁶¹ Vgl. hierzu auch MESSICK, 1972, 71 f.; ETS Annual Report 1969—1971, 39.

13.2.4. Forschung und Entwicklung im Bildungswesen

Es kann nicht Aufgabe dieses Gutachtens sein, die mannigfachen Forschungsschwerpunkte der pädagogischen Diagnostik zu benennen, die sich im Zusammenhang der Bildungsforschung insgesamt stellen können. Lediglich am Beispiel der Curriculumforschung und -revision, die für die Gesamtschule zentrale Bedeutung besitzt, seien einige Gesichtspunkte skizziert, welche bei der Forschung und Entwicklung der pädagogischen Diagnostik eine Rolle spielen.

Nach dem geläufigen Verständnis ist die pädagogische Diagnostik, sei es bei der formativen oder der summativen Evaluation, sei es bei der Ermittlung von Lernvoraussetzungen, wichtiger Bestandteil der Curriculumforschung und -entwicklung⁶². Dabei wird gewöhnlich der Primat der curricularen Gesamtkonzeption, insbesondere den Lern- und Bildungszielen eines Curriculum, zuerkannt; diese stellen die Vorgaben dar, an welchen sich die pädagogische Diagnostik orientiert. Dies gilt gerade dann, wenn neue curriculare Konzepte erprobt werden: Gleichgültig, ob es sich um Versuche mit individualisiertem Unterricht, um die Erprobung von Formen des „entdeckenden“ Lernens, um Lernsequenzen, bei welchen sich die Schüler ihre Ziele selbst setzen, oder um neue Inhalte handelt, stets ist die pädagogische Diagnostik den curricularen Intentionen nachgeordnet, da sich auf diese Weise am ehesten sichern läßt, daß die für die Konstruktion und Revision notwendigen Informationen bereitstehen, vor allem aber, daß die pädagogische Diagnostik keine den Intentionen des Curriculum zuwiderlaufenden Rückwirkungen auf die Schüler ausübt. Derartige „backwash“ Effekte können nämlich, wie man vielfach beobachtet hat⁶³, dazu führen, daß das Curriculum durch die pädagogische Diagnostik – und nicht umgekehrt – determiniert wird, was im günstigsten Falle auf eine Verarmung des Unterrichts hinausläuft.

So wichtig derartige Überlegungen auch sein mögen, die ausschließliche Nachordnung der pädagogischen Diagnostik gegenüber den Zielen und Inhalten des Curriculum kann den Ansprüchen an eine umfassende, ihren Aufgaben gerecht werdende pädagogische Diagnostik nicht genügen. Zum einen ist es notwendig, auch die möglichen, nicht nur die intendierten Auswirkungen eines Curriculum sowie die latenten Funktionen des Unterrichts zu erfassen, die beispielsweise in der Entwicklung inadäquater Arbeitstechniken, in der unbemerkten Minderung der Selbständigkeit, in der Steigerung der Abhängigkeit vom Lehrer, in der Übernahme der für den jeweiligen Lehrer typischen Ausprägung der Ambiguitätstoleranz und des Präzisionsbedürfnisses oder in der Schwächung des Selbstvertrauens der Schüler bestehen können. Ferner enthalten Curricula oft unausgesprochene oder nicht hinreichend präzise beschriebene Ziele, die bislang in der Curriculumtheorie unzureichend behandelt worden sind und daher für die Curriculumkonstruktoren keine geläufigen Denkmuster darstellen, wie etwa die von EISNER (1972, 580 f.) identifizierten „expres-

⁶² Vgl. z. B. DUNN, in: INGENKAMP und MARSOLEK, 1968.

⁶³ Vgl. z. B. DUNN, 1968, 1023 f., der u. a. über TYLERS 8 year study berichtet; VERNON, 1957, 62 f.; BLOOM, 1968, 1 f. Vgl. auch oben, Kap. 13.2.2 d.

sive objectives“ und „type III-objectives“. Schließlich muß man davon ausgehen, daß sich in der Schule auch neue, von den geläufigen Vorstellungen abweichende Begabungen entwickeln, die es zu erkennen und zu stimulieren gilt, auch wenn sie nicht den gängigen Fähigkeiten entsprechen, deren Förderung man derzeit von der Schule erwartet⁶⁴.

Schon an diesen wenigen Beispielen wird deutlich, daß die pädagogische Diagnostik wichtige, über die primären Erfordernisse der Curriculumforschung und -entwicklung hinausgehende Aufgaben zu erfüllen hat und aufgrund dieser Eigenständigkeit durch eine vollständige Unterordnung unter die Bedürfnisse der Curriculumkonstruktion in wesentlichen Funktionen beschnitten würde. Für die partielle Unabhängigkeit der pädagogischen Diagnostik spricht aber auch die Befürchtung, daß eine streng curriculumgebundene pädagogische Diagnostik zu einer Behinderung oder gar Blockierung innovativer Tendenzen führen kann, wenn die diagnostischen Instrumente nicht fortlaufend den sich ändernden Zielen und Inhalten angepaßt werden, sondern, etwa als standardisierte Tests, die Unterrichtsinhalte beizubehalten zwingen⁶⁵. Ein vordringliches Anliegen von Forschung und Entwicklung im Bereich pädagogischer Diagnostik besteht daher in der Schaffung breiter, vielseitiger, auch auf seltene Inhalte gerichteter, flexibler Instrumente, die geeignet sind, die aufgrund unabhängiger Analysen zu erwartenden manifesten und latenten Funktionen von Curricula zu erkennen sowie ihren Veränderungen verzögerungsfrei zu folgen. Wie weit die pädagogische Diagnostik von einem solchen Zustand entfernt ist, zeigen die Schulversuche, denen besonders dann keinerlei diagnostische Hilfen zur Verfügung stehen, wenn ihr Programm etwas stärker von den gewohnten Formen des Unterrichts abweicht⁶⁶.

Eine wichtige Forschungsaufgabe besteht in der Bestimmung der Gleichwertigkeit von Lernzielen. Dies Problem besteht beispielsweise bei sämtlichen Abschlußprüfungen unseres Schulsystems, bei welchen die Prüfungsinhalte (und die dem vorausgehenden Unterricht zugrunde liegenden Lernziele) von Schule zu Schule und von Ort zu Ort stark variieren, dennoch aber Gleichwertigkeit (zum Beispiel „allgemeine Hochschulreife“) unterstellt wird. WOOD und SKURNIK (1969) haben für das englische CSE-Abschlußexamen ein auf der Basis einer Itembank beruhendes Verfahren entwickelt, welches wert wäre, weiter verfolgt zu werden. Gleichwertigkeitsüberlegungen lassen sich auch auf der Ebene der Lernziele ansiedeln, wie sie in der BLOOMschen Taxonomie vorliegen. So hat ROGERS⁶⁷ vorgeschlagen, die diagnostischen Verfahren lernzielspezifisch zu konstruieren und den Schülern zu demselben Inhalt Tests zur Auswahl zu stellen, welche zum Beispiel entweder nur „Kenntnisse“ oder nur „Anwendung“ prüfen („Montagstests“ versus „Dienstagstests“), die Ergebnisse aber als gleichwertige Grundlage der Zensur zu betrachten. Besondere Bedeutung besitzt diese Frage hinsichtlich der Bestimmung von äquivalenten

⁶⁴ Vgl. auch SCHREINER, 1970, 230.

⁶⁵ Vgl. z. B. BLAKE, 1971.

⁶⁶ Vgl. z. B. den Bericht über Countesthorpe, in: Gesamtschule Informationen 3, 1972, 29.

⁶⁷ In einem Vortrag, gehalten auf dem IPN-Symposium in Kiel, März 1972. Vgl. hierzu auch WOOD und SKURNIK, 1969, 184.

Elementen in beruflichen und allgemeinen Bildungsgängen, deren Integration eine der wichtigsten künftigen Aufgaben der Schulreform darstellt.

Für die Curriculumforschung und -entwicklung würde es eine wesentliche Entscheidungshilfe darstellen, wenn Informationen über Langzeiteffekte des schulischen Lernens zur Verfügung stünden. Wenn man wüßte, welche Kenntnisse und Fertigkeiten, Einsichten und Problemlösungsstrategien, Einstellungen und Werte, die in der Schule erworben werden, auch später noch vorhanden sind und bis zu welchem Grade sie dem einzelnen noch zu Gebote stehen, ließe sich über die Relevanz von Inhalten und Zielen für spätere Verwendungssituationen, über ihren Voraussetzungswert für späteres Lernen sowie über die Effektivität von Unterrichtsmethoden besser diskutieren. Jedenfalls reicht es nicht aus zu überprüfen, ob beispielsweise ein bestimmtes Curriculum seine Ziele erreicht hat; die wichtigere Frage besteht darin, ob der treatment-Effekt andauert, ob die erworbenen Kenntnisse auch benutzt werden, ob es überhaupt einen Unterschied macht, daß der Schüler sie einmal erworben hat (beispielsweise indem sich der Zeitaufwand für das Wiederlernen verkürzt), etc.⁶⁸

Die herkömmliche pädagogische Diagnostik ist nicht gerüstet, auf solche Fragen eine Antwort zu geben. Es wäre daher wichtig, Instrumente zu entwickeln und Techniken zu adaptieren, wie sie derzeit beispielsweise im Projekt „National Assessment of the Educational Progress“ (NAEP) in den USA Verwendung finden⁶⁹; die laufend publizierten Befunde zeigen überzeugend den Wert einer solche Studie, die zuletzt für die Intensivierung der öffentlichen Diskussion über die Bildungs- und Lernziele der Schule.

Schließlich sei noch auf ein Forschungsdefizit hingewiesen, auf welches insbesondere STAKE (1970) aufmerksam gemacht hat. Bildungs- und Lernziele, aber auch andere „*judgment data*“ wie Werte, Standards, wahrgenommene Normen, Prioritäten, sind zwar logisch-philosophischen Analysen unterzogen worden, empirische Untersuchungen hierüber fehlen jedoch noch weitgehend. Angesichts des Gewichtes, mit dem solche Daten in die täglichen Entscheidungen der einzelnen Schule und des Unterrichts eingehen, sollte der Versuch unternommen werden, mittels in der psychologischen Diagnostik entwickelter und neu zu konstruierender Verfahren die Informationsqualität in diesem Bereich zu verbessern⁷⁰.

13.3. Zur Lage der gegenwärtigen pädagogischen Diagnostik in der BRD

Die bestehende pädagogische Diagnostik, vor allem Zensuren und Zeugnisse, erneut einer kritischen Analyse unterziehen hieße die bereits geleistete Mängelanalyse⁷¹ im

⁶⁸ Vgl. hierzu z. B. CARROLL, 1965; EISNER, 1972, 570.

⁶⁹ Vgl. u. a. STAKE, 1971, 65.

⁷⁰ Vgl. im einzelnen STAKE, 1970, der auch einige die pädagogische Diagnostik berührende Fragen diskutiert.

⁷¹ Vgl. vor allem INGENKAMP, 1969 und 1971, aber auch die Darstellungen von SCHREINER, 1970, WEIS, 1971, und vielen anderen.

wesentlichen wiederholen, welche zur Identifikation zahlreicher, schwerwiegender Defizite geführt hat (beispielsweise zum Aufweis der mangelhaften Reliabilität der Zensuren, ihres geringen diagnostischen Wertes, ihrer negativen Wirkungen auf die Lernmotivation, ihrer Multidimensionalität⁷², ihrer fragwürdigen „pädagogischen“ Funktion⁷³ sowie ihrer mannigfach determinierten Verzerrungen⁷⁴). Auch bei der Darstellung der vier Funktionen der pädagogischen Diagnostik dürfte deutlich geworden sein, daß die herkömmlichen Verfahren den Ansprüchen in keiner Weise genügen. Es ist merkwürdig festzustellen, wie wenig die schulische Praxis auf eine so ausgedehnte und eindeutige Mängelanalyse reagiert hat.

Bedenkt man freilich, wie beim derzeitigen Stand der Forschung und Entwicklung kurzfristig realisierbare Alternativen zur Zensurengebung aussehen könnten, so treten auch gewisse Vorzüge des tradierten Systems zutage. Dies trifft vor allem zu, wenn man sich die Zensuren durch standardisierte Tests ersetzt denkt: Selbst wenn es für alle Unterrichtsziele und -inhalte derartige Tests gäbe, wäre die Relativität des Lehrerurteils den übergreifenden Normen standardisierter Tests im allgemeinen vorzuziehen, zumindest solange Jahrgangsklassen mit Versetzungen sowie das traditionelle Berechtigungswesen existieren. Bei der Testverwendung müßte man nämlich erwarten, daß Klassen, die längere Zeit schlecht unterrichtet worden sind und wenig Anregung hatten, überwiegend mit schlechten Zensuren belegt würden, während das Lehrerurteil eine Rangordnung unter Berücksichtigung der Bedingungen, unter welchen in der jeweiligen Klasse gelernt wurde, herstellen würde. Kein Schüler einer solchen Schule würde daher bei der ausschließlichen Anwendung standardisierter Tests die von guten Zensuren abhängige Chance erhalten, sich einer anderen, anregenderen Lernumwelt auszusetzen⁷⁵. Das Lehrerurteil gewährt also dem Schüler eine gewisse Unabhängigkeit von den zufälligen Standards der Klasse, der er angehört⁷⁶, führt freilich auch umgekehrt aufgrund der Normalverteilungstendenz zur Unterschätzung hoher Leistungen in besonders günstig zusammengesetzten und optimal unterrichteten Klassen⁷⁷. Hier zeigt sich erneut, daß die Beurteilung eines Lernergebnisses nicht ohne Berücksichtigung des treatment und der weiteren Lernumwelt erfolgen darf, wozu es der Entwicklung von „Umweltmaßen“ bedarf, die an Detailliertheit den Instrumenten zur Messung individueller Unterschiede nicht nachstehen⁷⁸.

⁷² Vgl. z. B. ROEDER, 1974.

⁷³ Vgl. z. B. SCHREINER, 1970, 227 f.

⁷⁴ Vgl. z. B. ULICH und MERTENS, 1973, besonders S. 60 ff.

⁷⁵ Daß sich bei einem solchen Wechsel geradezu sprunghafte Intelligenzanstiege ergeben können, wurde in zahlreichen Untersuchungen bei Schülern, die ins College übertraten, festgestellt, vgl. z. B. die Übersicht bei BLOOM, 1964, 82 f.

⁷⁶ Daß dieses Problem auch in anderen Ländern gesehen wird, zeigen u. a. WOOD und SKURNIK, 1969, 66.

⁷⁷ Sofern standardisierte Tests nur als Bezugstests benutzt werden, geben sie andererseits dem Lehrer nützliche Orientierungshilfen, vgl. INGENKAMP 1969, 424. Diese Funktion läßt sich allerdings durch in vollem Umfange inhaltsgültige Verfahren, wie sie durch eine Itembank generiert werden könnten, wesentlich besser wahrnehmen.

⁷⁸ Vgl. hierzu SHULMAN, 1970, 373 ff.

Es wäre allerdings ein Mißverständnis, wenn diese Bemerkungen als ein Votum für die Beibehaltung der derzeitigen pädagogischen Diagnostik verstanden würden. Notwendig ist vielmehr die rasche und auf mehreren Fronten verlaufende Erforschung dieses Bereichs und die Entwicklung von brauchbaren Alternativen, welche die dringend erforderliche Ablösung des derzeitigen Zensurenwesens ermöglichen. Wie stark das Bedürfnis nach Veränderungen auf diesem Sektor ist, zeigen sinnfällig auch die Schwierigkeiten, die bei einigen Gesamtschulversuchen aufgrund der mangelhaften Möglichkeiten auch einer versuchsweise revidierten pädagogischen Diagnostik auftreten. So hat beispielsweise die konsequente Verwendung informeller Tests und ein daran geknüpftes Differenzierungsmodell an einer Berliner Gesamtschule eher beklagenswerte Konsequenzen gezeitigt⁷⁹; Versuche an der Gesamtschule Weinheim, die herkömmlichen Zeugnisse durch lernzielorientierte Berichte zu ersetzen, welche nur wenige Qualitätsstufen enthalten (und die geradezu ein Modell einer wesentlich verbesserten pädagogischen Diagnostik darstellen), stoßen u. a. auf den Widerstand externer Instanzen. Dies zeigt die Notwendigkeit, solche Versuche auf der Grundlage vorgängiger Analysen des Feldes, in dem sie ablaufen sollen, umfassend zu planen.

Auch zum Stand der hiesigen Forschung über die pädagogische Diagnostik braucht hier nicht ausführlich gesprochen zu werden, da zwei voneinander unabhängige Analysen zur Zeit im Entstehen begriffen sind⁸⁰. Selbst wenn man die gelegentlich masochistisch wirkende Neigung zur Selbstkritik mancher Diagnostiker in Rechnung stellt, wird man dem Urteil zustimmen müssen, daß die Forschung, gemessen an den theoretischen und praktischen Bedürfnissen, noch ganz in den Anfängen steckt. Dies zeigt auch eine Durchsicht der wichtigsten deutschsprachigen psychologischen und pädagogischen Zeitschriften beispielsweise des Jahrgangs 1972 oder der Dokumentation über die Forschungsarbeiten in den Sozialwissenschaften von 1971⁸¹. Abgesehen von wenigen Arbeiten findet man die bekannten kritischen Argumente gegen die herkömmlichen Zensuren sowie Diskussionen über deren Selektionsfunktion, nicht jedoch über ihre Funktion der Lernförderung. Zudem gewinnt man den Eindruck, daß die berichteten Forschungsarbeiten ganz unsystematisch und unkoordiniert die Interessen der einzelnen Forscher widerspiegeln. Auch sind die bisherigen Versuche, neue Modelle der pädagogischen Diagnostik zu entwickeln und zu erproben, vereinzelt geblieben oder stecken noch in den Anfängen, wenn man einmal von dem über Jahrzehnte bewährten, freilich weithin unbekanntem Modell absieht, welches an den freien Waldorfschulen praktiziert wird und welches Eigenschaften aufweist, die geeignet wären, einige der oben diskutierten, wichtigen Funktionen der pädagogischen Diagnostik zu erfüllen⁸².

⁷⁹ Vgl. SCHLÖMERKEMPNER, 1971.

⁸⁰ INGENKAMP im Auftrag des Europarates und LÜKING im Auftrag der Stiftung Volkswagenwerk.

⁸¹ Forschungsarbeiten 1971 in den Sozialwissenschaften. Dokumentation des Informationszentrums für sozialwissenschaftliche Forschung, München, Berlin 1972.

⁸² Vgl. z. B. RAUER, 1973.

13.4. Mittelfristige Forschungsschwerpunkte der pädagogischen Diagnostik

Die in Abschn. 13.2. bezeichneten Forschungsdesiderata sind zu umfangreich, als daß sie angesichts der verfügbaren finanziellen und personellen Möglichkeiten in absehbarer Zeit erfüllt werden könnten. Es gilt daher, unter den Gesichtspunkten der Dringlichkeit und der Machbarkeit einige Schwerpunkte zu identifizieren. Hierbei werden, wie in der Einleitung begründet, die Bedürfnisse der Sekundarstufe und der sich entwickelnden Gesamtschulen die Auswahl anleiten, während andere Bereiche, beispielsweise das vorschulische Lernen oder die Weiterbildung, nicht diskutiert werden. Eine Reduzierung des Schwerpunktkataloges erfolgt darüber hinaus aufgrund der vorliegenden Mängelanalysen, die massive negative Auswirkungen der herkömmlichen, vorwiegend auf die Leistungsbeurteilung beschränkten pädagogischen Diagnostik nachgewiesen haben. Eine Konzentration auf Forschungs- und Entwicklungsvorhaben, die hier Abhilfe schaffen könnten, erscheint somit angezeigt. Dabei muß man sich der Gefahr bewußt sein, daß allein schon das Vorhandensein diagnostischer Verfahren den Gegenständen, auf die sie gerichtet sind, zusätzliches Gewicht verleiht. Daher darf die Entwicklung einer auf andere Phänomene als das kognitive Lernen gerichteten pädagogischen Diagnostik nicht aus den Augen verloren werden.

Innerhalb der Sekundarstufe und der Gesamtschule ergibt sich ein breites Spektrum von Funktionen und Verwendungsmöglichkeiten der pädagogischen Diagnostik. So bedürfte beispielsweise ein Differenzierungsmodell, wie es in der „Individually Prescribed Instruction“ (IPI) vorliegt, einer Fülle in das Curriculum integrierter Beurteilungsverfahren, welche die notwendige Steuerung des Lernens gewährleisten; RODERS (1974) Vorschlag der Differenzierung nach Leistungsbereichen beziehungsweise -dimensionen erfordert ebenfalls eine recht umfangreiche pädagogische Diagnostik. Formen flexibler Unterrichtsorganisation, wie man sie in englischen Primarschulen antrifft, scheinen dagegen nur auf wenige diagnostische Verfahren angewiesen zu sein. Je nach Differenzierungsmodell wird man demnach Diagnoseverfahren in sehr verschiedenem Umfang und von unterschiedlicher Art benötigen. Es ist eine wichtige Forschungsfrage, das für die Verwirklichung der Schul- und Unterrichtsziele notwendige *Minimum an pädagogischer Diagnostik, welches bei geeigneten curricularen und unterrichtsorganisatorischen Maßnahmen, die die Selbststeuerung des Lernens durch die Schüler begünstigen, weit unter dem derzeit üblichen Ausmaß an Schülerbeurteilung liegen kann*, sowie die den Zielen am ehesten entsprechende Form der Verfahren und die Art ihrer Verwendung zu bestimmen. Darüber hinaus muß sich die Akzentverlagerung eines Unterrichts, bei dem es mehr auf die Förderung des Lernens als auf kontinuierliche Auslese ankommt, in Form und Verwendungsweise der pädagogischen Diagnostik widerspiegeln.

Bei aller Variation zwischen den möglichen Differenzierungsformen läßt sich jedoch erkennen, daß die benötigten Instrumente der pädagogischen Diagnostik außerordentlich flexibel sein müssen, insofern sie sowohl den Bedürfnissen zunehmend individualisierten Lernens genügen als auch genau zu den je spezifischen Zielen und Inhalten, welche dem Unterricht zugrunde liegen, passen müssen. Standardisierte,

aus einer fixierten Menge von Items bestehende Tests oder sonstige Verfahren können daher nicht als geeignet gelten, die Probleme des Lernens, der Klassifikation, der Information und der Curriculumentwicklung zu lösen. Im Gegenteil: Bei häufiger Verwendung legen sie für lange Zeit fest, welche Qualifikationen vernachlässigt beziehungsweise gefördert werden, und erschweren die flexible Weiterentwicklung der Curricula.

Neben der allgemeinen Forderung nach Flexibilität einer zu entwickelnden pädagogischen Diagnostik steht die Notwendigkeit ihrer Theoriebezogenheit. Wenn man diagnostische Befunde sinnvoll interpretieren und Förderungsmaßnahmen aus ihnen ableiten möchte, benötigt man Einblick in die Ursachen von Mißerfolgen, Lerndefiziten, Motivationsverlusten etc. Forschung und Entwicklung in der pädagogischen Diagnostik lassen sich daher ohne engen Bezug zur sonstigen Bildungsforschung nicht sinnvoll betreiben. So ist es beispielsweise wichtig, eine sorgfältige psychologische Analyse der Lern- und Problemlösungsprozesse vorzunehmen, deren Resultate mit den Verfahren der Leistungsbeurteilung erfaßt werden; die genaue Prüfung der einzelnen Items eines Beurteilungsverfahrens unter diesem Gesichtspunkt⁸³ liefert dabei wertvolle Hinweise für die Planung des Unterrichts. Sinnvoller als die Erhöhung von Koeffizienten der prognostischen Gültigkeit dürfte der Versuch sein, eine möglichst präzise, der theoretischen Weiterentwicklung folgende Beziehung des Instruments auf das Konstrukt herzustellen, welches es messen soll (zum Beispiel Schulleistung oder Intelligenz). Die notwendige Ausrichtung der pädagogischen Diagnostik auf die Lernziele und -inhalte, die bei der Curriculumkonstruktion und der Unterrichtsplanung herausgearbeitet werden, darf nicht dazu führen, daß die Erforschung beispielsweise der kognitiven Stile oder des „mathematischen“ Verhaltens⁸⁴ vernachlässigt wird oder daß die Ergebnisse solcher Untersuchungen bei der Fortentwicklung der pädagogischen Diagnostik unberücksichtigt bleiben.

Es erscheint angezeigt, an dieser Stelle an die auch jetzt schon bestehenden Möglichkeiten zu einer kurzfristigen Verbesserung der herkömmlichen pädagogischen Diagnostik zu erinnern. Beispiele bereits geleisteter Forschung auf diesem Gebiet, die lediglich der Implementierung bedürfen, finden sich vor allem in Abschnitt 13.2.3. Die im folgenden aufgeführten sechs Forschungsschwerpunkte der pädagogischen Diagnostik beruhen, ohne daß die einzelnen Begründungen wiederholt werden, auf den oben angestellten Überlegungen, der Rezeption der einschlägigen Literatur sowie auf der Einschätzung der derzeitigen Forschungslage und der Bedürfnisse von Gesamtschulen.

⁸³ Vgl. hierzu ANDERSON, 1972.

⁸⁴ Vgl. ROTHKOPF, 1973.

13.4.1. Kriteriumbezogene Verfahren

In diesem Schwerpunktbereich liegen eine Reihe einschlägiger deutscher Arbeiten vor⁸⁵, die es voranzutreiben und zu erweitern gilt. Es geht dabei nicht nur um die Konstruktion „lernzielorientierter“ Tests und die Lösung der bei ihnen auftretenden statistischen Probleme, sondern das Kriterium, auf welches die zu erarbeitenden Verfahren bezogen sein sollen, muß selbst unter psychologischen und pädagogischen Gesichtspunkten ausdifferenziert werden. So dürfte es beispielsweise um der Förderung des Lernens willen langfristig wichtiger sein, auf die Identifikation von Dimensionen schulischer Leistung⁸⁶ hinzuarbeiten und kriterienbezogene Maße hierfür zu entwickeln, statt sich auf die von Lehrern bestimmten Lernziele zu beschränken. Darüber hinaus müssen Modelle für die Interpretation von Ergebnissen aus kriterienbezogenen Messungen entwickelt und erprobt werden.

Zahlreiche offene Fragen bestehen auch hinsichtlich einer Sonderform⁸⁷ der kriterienbezogenen pädagogischen Diagnostik, der Messung und Beurteilung des „zielreichenden Lernens“ (*Mastery learning*)⁸⁸. Nicht nur, daß die Prüfung der inneren Konsistenz oder der Stabilität solcher Verfahren aufgrund der stark reduzierten Varianz der Ergebnisse besondere Schwierigkeiten bereitet, es fehlt hier vor allem an vorgängiger Curriculumforschung und -entwicklung, bei welcher nicht nur diejenigen grundlegenden Fähigkeiten und Fertigkeiten, auf die allein zielerreichendes Lernen sinnvoll gerichtet sein kann, beschrieben und begründet werden, sondern auch die Konstruktion von Lernzielhierarchien beziehungsweise von sequentiellen Lernstufen⁸⁹ in Angriff genommen wird.

Im Unterschied zu anderen Bereichen scheint es hier auch lohnend zu sein⁹⁰, die Vor- und Nachteile computerunterstützten Testens näher zu prüfen.

13.4.2. Itembank

Die vielfach begründete Forderung nach hoher Flexibilität der pädagogischen Diagnostik, die in einem Kontext zunehmend individualisierten Lernens wirksam werden soll, läßt sich nicht mit Hilfe festgefüger, in sich geschlossener Tests oder sonstiger Verfahren erfüllen. Die grundlegenden und vielversprechenden Vorarbeiten von WOOD und SKURNIK (1969) zur Konstruktion einer Itembank sollten deshalb in Deutschland zum Ausgangspunkt weiterer Forschungs- und Entwicklungsarbeiten genommen werden⁹¹. Denn es kann kein Zweifel bestehen, daß eine gut kon-

⁸⁵ Vgl. KLAUER u. a., 1972; FRICKE, 1972; HEIPCKE, 1970; WEIS, 1971.

⁸⁶ Vgl. z. B. ROEDER, 1974.

⁸⁷ Vgl. POPHAM und HUSEK, 1969, im Unterschied beispielsweise zu FRICKE, 1972, 150 ff.

⁸⁸ Vgl. CARROLL u. a., 1971; AIRASIAN, 1971.

⁸⁹ Vgl. BLOCK, 1971.

⁹⁰ Vgl. z. B. BRENNAN, 1972, 15; HOLTZMAN, 1971, 548.

⁹¹ Das von ROYL, 1972, beschriebene Itembankprojekt weicht einerseits in wesentlichen Punkten von dem britischen Ansatz ab, hat andererseits aber auch in bezug auf die noch offenen, wichtigen psychometrischen und interpretatorischen Probleme offenbar noch keine Lösungsversuche unternommen.

struierte, breit angelegte Itembank einige fundamentale Schwierigkeiten der herkömmlichen pädagogischen Diagnostik gegenstandslos werden lassen könnte: Mit ihrer Hilfe lassen sich für jede Lerngruppe spezifische, kriterienbezogene Instrumente zusammenstellen; sie liefert aufgaben- oder lernzielspezifische Verfahren zur Beurteilung der Vorkenntnisse von Schülern; sie ermöglicht die Messung auch seltener Lernziele und Inhalte sowie der Langzeiteffekte schulischen Lernens⁹²; mit Hilfe einer Itembank lassen sich durch *item sampling* besonders wirksame und ökonomische Tests zur Beurteilung der Qualität und Auswirkung von treatments zusammenstellen, ohne daß dabei die Schüler einer individuellen Prüfung unterzogen werden; kurz, es gibt wenige Aufgaben im Bereich des kognitiven Lernens, der Beratung, der Klassifikation, der Curriculumforschung oder der Information, die nicht mittels einer Itembank besser als durch andere Hilfsmittel erfüllt werden könnten. Auch läßt sich erwarten, daß bei breiter Nutzung der flexiblen Angebote einer Itembank die zu befürchtenden, fatalen *back wash*-Effekte standardisierter Tests, von denen oben die Rede war, im wesentlichen vermieden werden können.

Dringliche Forschungsprobleme ergeben sich hier vor allem in bezug auf die bei WOOD und SKURNIK angezielte Vergleichbarkeit der Standards, die Verankerung der Items an Außenkriterien sowie die maximale Diskriminierung zwischen Individuen. Beispielsweise müßte es im Unterschied zu dem britischen Ansatz auch möglich sein, von einer Itembank lernzielbezogene Tests abzurufen, die zur Überprüfung zielerreichenden Lernens dienen und eine echte Stichprobe aus einem „Universum“ von Items zu einem bestimmten Ziel darstellen⁹³, ohne daß damit ein Vergleich mit einem Außenkriterium impliziert wäre. Auch andere Möglichkeiten der Aktivierung einer Itembank zum Zweck der Lernförderung, nicht zur Perfektionierung des normbezogenen Vergleichs, wären zu explorieren. Die Arbeiten von WOOD und SKURNIK⁹⁴ können dabei als erste, wichtige Anregung dienen, und es bedarf hier umfangreicher, auch auf die Identifikation möglicher Folgewirkungen gerichteter Forschung, nicht aber der übereilten Entwicklung einer durch ihre Effizienz bestehenden Technologie, die zur Normierung und Standardisierung des Prüfungswesens und den damit verbundenen negativen Konsequenzen führen müßte. Ob eine Itembank die jeweilige Zusammenstellung der Instrumente und Teile der Ergebnisinterpretation den Lehrern abnehmen sollte, oder ob es besser wäre, beispielsweise den Lehrern nur eine größere Auswahl qualifizierter Aufgaben anzubieten, die sie dann selbst zusammenstellen und anwenden, bedarf ebenfalls der Untersuchung.

⁹² Etwa nach dem Vorbild des Projektes National Assessment of the Educational Progress, vgl. z. B. HOLTZMAN, 1971.

⁹³ Hierzu vgl. z. B. BRENNAN, 1972.

⁹⁴ Vgl. auch WOOD, 1968.

13.4.3. Individuelle Unterschiede

Es liegt auf der Hand, daß der Bedarf an diagnostischen Verfahren, mit denen sich individuelle Unterschiede zuverlässig bestimmen lassen, in einer auf die Förderung des individualisierten Lernens gerichteten Schule erheblich zunimmt, besonders, wenn die herkömmlichen Differenzierungsformen beibehalten werden. Vor allem für die Bedürfnisse der Beratung kommt es darauf an, die individuellen Stärken und Schwächen identifizieren und interpretieren zu können, um den Schüler zu einer informierten Entscheidung zu befähigen und über das damit verbundene Risiko aufzuklären⁹⁵. Für Unterrichtsformen wie die oben (Abschn. 13.2.2.) beschriebene „flexible Unterrichtsorganisation“ bedarf es andererseits der Erforschung der Möglichkeiten und Grenzen der Selbstbeurteilung und damit der Bestimmung eines möglicherweise sehr niedrig anzusetzenden Bedarfs an diagnostischen Verfahren.

13.4.4. Umweltmaße

Es ist oben begründet worden, warum in die Leistungsurteile auch die Bedingungen eingehen müssen, unter denen gelernt worden ist. Auch gehört es zu den Voraussetzungen einer sinnvollen Klassifikation der Schüler zu wissen, welche Schüler am besten zu welchem treatment passen beziehungsweise wie man treatments verändern muß, um sie für die Schüler fruchtbar werden zu lassen. Unter den möglichen Forschungsthemen in diesem Feld kommt daher der Erforschung der im Unterricht wirksamen treatments und der Entwicklung einer entsprechenden pädagogischen Diagnostik vor allem aus Gründen der relativ kurzfristigen Modifizierbarkeit dieser Faktoren Priorität zu. Gleichwohl darf die Forschung über andere Aspekte der Lernumwelt, zum Beispiel über die engeren Determinanten von Lerndefiziten, die hinter den gebräuchlichen, groben, sozio-ökonomischen Indikatoren stehen, oder über das schulische „Lernklima“ nicht vernachlässigt werden, da solche Kenntnisse eine wichtige Grundlage für gezielte kompensatorische Bemühungen darstellen⁹⁶. Auch an das Problem der Gleichwertigkeit von Lernzielen im allgemeinbildenden und im berufsbildenden Schulwesen sei hier erinnert.

13.4.5. Pädagogische Diagnostik für den Bereich des sozialen und affektiven Verhaltens

Der Bereich des sozialen und affektiven Lernens ist aus der pädagogischen Diagnostik in der Schule bislang ebenso ausgeklammert geblieben wie die Erfassung der unmittelbar auf die Unterrichtsgestaltung einwirkenden Werte und Einstellungen der

⁹⁵ Vgl. z. B. HIERONYMUS, 1972, 65; WEIS, 1971, 553.

⁹⁶ Auf einige der verstreuten Vorarbeiten in diesem Gebiet sei hingewiesen: Educational Researcher, 2, 1973, 2, 18 f.; WALBERG, 1971; BLOOM, 1968, 12, mit weiterer Literatur; WALBERG und ANDERSON, 1972; DREW, 1971; CENTRA und ROCK, 1971; WOLF, 1964; EISNER, 1972.

Lehrer⁹⁷. Es ist daher an der Zeit, eine auf diese wichtigen Vorgänge gerichtete pädagogische Diagnostik zu entwickeln, sei es um Lernhemmnisse zu beseitigen, sei es zur Beratung, sei es für die Bedürfnisse der Curriculumrevision oder zur Feststellung der bislang latenten Nebenwirkungen von Schule und Unterricht⁹⁸.

13.4.6. Psychometrische Fragen

Zum Schluß seien einige eher technische Probleme aufgeführt, deren Lösung dringlich wäre. Sieht man einmal von den – auch – psychometrischen Schwierigkeiten der Konstruktion einer Itembank ab, ist hier vor allem zu denken an die Entwicklung neuer Verfahren zur Bestimmung der Inhaltsgültigkeit, an die Erprobung unterschiedlicher Modelle der Datengewichtung, an die Lösung der mit der Interpretation von Testprofilen verbundenen Probleme, an die Entwicklung von Instrumenten, deren Meßergebnisse die Veränderung und Entwicklung der Schüler interpretierbar wiedergeben, sowie an die Erprobung des *item-sampling* für die treatment-Messung und Curriculumevaluation.

13.4.7. Schulversuche

Die Interdependenz wichtiger schulischer Faktoren wie zum Beispiel der Differenzierung, des Curriculum und der pädagogischen Diagnostik, dürfte an mehreren Stellen deutlich geworden sein. Ein wichtiger Teil der Forschung und Entwicklung im Bereich der pädagogischen Diagnostik wird deshalb in Form komplexer Schulversuche durchgeführt werden müssen, sei es daß es darum geht, die Auswirkungen einer radikal reduzierten Anwendung diagnostischer Verfahren im Unterricht zu erproben, sei es daß Felderprobungen einer im Labor entwickelten Technologie durchgeführt werden⁹⁹, sei es daß die Rezeption veränderter Leistungsbeurteilungen auf externe Instanzen, zum Beispiel auf Eltern oder Arbeitgeber, in Erfahrung gebracht werden sollen.

13.5. Institutionalisierung

Die folgenden Vorschläge zur Institutionalisierung der Forschung und Entwicklung der pädagogischen Diagnostik beziehen sich auf die Verwirklichung der im Gutachten vorgeschlagenen Forschungsschwerpunkte¹⁰⁰. Die Institutionalisierung erfolgt

⁹⁷ Vgl. hierzu MESSICK, 1972, 79.

⁹⁸ An dieser Stelle sei nochmals an die in diesem Gutachten undiskutiert gebliebenen, wichtigen Bereiche des motorischen und ästhetischen Lernens erinnert, die ebenfalls eine Fülle forschungsbedürftiger Probleme diagnostischer Art enthalten.

⁹⁹ Vgl. hierzu CRONBACH und SUPPES, 1969, 82 ff.

¹⁰⁰ Vgl. auch die (partiell abweichenden) Institutionalisierungsvorschläge bei INGENKAMP, 1969, 428 f.

auf drei Ebenen, der nationalen, der regionalen und der lokalen, wobei der Grad der Spezialisierung in derselben Reihenfolge abnimmt, auf lokaler Ebene also beispielsweise die stärkste Verschränkung der pädagogischen Diagnostik mit anderen Bereichen, vor allem der Curriculumentwicklung, anzutreffen sein wird. Zur Forschung an den Hochschulen werden hier keine Vorschläge unterbreitet.

13.5.1. Zentralinstitut für pädagogische Diagnostik

Für eine große Zahl wichtiger Aufgaben – die Entwicklung einer Itembank, die Bearbeitung der mit kriterienorientierten Beurteilungsverfahren verbundenen Probleme, die Bearbeitung der auch für den sozial-affektiven, den motorischen und den ästhetischen Bereich erforderlichen Diagnostik, für die Koordination von Schulversuchen, die Erforschung der Langzeitwirkungen von Schule und Unterricht, die Evaluation der dezentral erfolgenden Curriculumevaluation, die durch Forschung zu leistende Umsetzung von Grundlagenerkenntnissen in die Praxis, die Bestimmung extra-curricularer Auswirkungen der Schule, die Beschreibung der Auswirkungen bestimmter diagnostischer Systeme auf die Schule und vieles andere mehr – dürfte ein Zentralinstitut für pädagogische Diagnostik besser als andere Institutionen geeignet sein, die notwendige Forschungs- und Entwicklungsarbeit zu leisten. Auch kann nur in einem Zentralinstitut der notwendige wissenschaftliche Sachverstand vereinigt sowie die erforderliche technische Ausstattung vorgesehen werden. Insbesondere durch die Itembank, die den Kern eines solchen Instituts darstellen würde, läßt sich dabei der enge Kontakt zu den Schulen und den Institutionen auf mittlerer Ebene sichern, insofern beide Seiten auf Austausch und Kooperation angewiesen sind; zugleich läßt sich dadurch die Koordination der bislang zersplitterten Arbeitsgruppen auf dem Gebiet der Curriculumforschung und -entwicklung verbessern. Hierzu muß gewährleistet sein, daß den Mitarbeitern des Zentralinstituts der Zugang zur Praxis jederzeit möglich ist.

Es gehört auch zu den Aufgaben eines zentralen Instituts, für breite Aufklärung der Öffentlichkeit über die Probleme und Implikationen der pädagogischen Diagnostik zu sorgen sowie zur Fortbildung bestimmter Personengruppen, zum Beispiel der in der Lehrerbildung tätigen Personen, beizutragen. Darüber hinaus muß für einen engen Kontakt zur sonstigen Bildungsforschung Sorge getragen werden.

Ein fester Platz im Forschungsprogramm des Zentralinstituts gebührt der Untersuchung der sozialen Konsequenzen der pädagogischen Diagnostik¹⁰¹. Wie die Entwicklung in den USA gezeigt hat, werden die mit der Diagnostik verbundenen Gefahren leicht unterschätzt¹⁰² beziehungsweise zu spät erkannt.

Wegen der bei einer zentralen Einrichtung besonders großen Gefahr möglichen Mißbrauchs ist die Unabhängigkeit des Instituts durch geeignete Maßnahmen zu sichern. Das curriculare Problem der Normierung des Unterrichts durch Tests dürfte zwar

¹⁰¹ Vgl. z. B. SIZER, 1971.

¹⁰² Vgl. z. B. HOLTZMAN, 1971.

dadurch, daß eine Itembank standardisierte Tests weitgehend ersetzt, im wesentlichen gelöst sein. Unabhängigkeit ist aber auch notwendig, wenn es beispielsweise darum geht, Verwaltungsentscheidungen kritisch zu durchleuchten, auf die Einhaltung von Standards zu achten oder Zielkonflikte aufzudecken.

Es empfiehlt sich, darüber hinaus die Möglichkeit zur Gründung einer auch vom Zentralinstitut unabhängigen, kritischen „*Diagnostiker-Vereinigung*“ zu prüfen, durch welche Kontrolle über die Einhaltung von Standards ausgeübt wird sowie auf dem Wege über Grundsatzentscheidungen zu konkreten Problemfällen Maßstäbe gesetzt werden¹⁰³.

13.5.2. Regionale Pädagogische Zentren und lokale Curriculumprojekte

Regionalen Pädagogischen Zentren (RPZ)¹⁰⁴ oder vergleichbaren Institutionen könnte die Aufgabe der Entwicklung diagnostischer Verfahren im Zusammenhang der Curriculumforschung und -entwicklung zufallen. Hierbei dürfte die Zusammenarbeit mit der zentralen Itembank außerordentliche Erleichterungen mit sich bringen, wie umgekehrt die Itembank durch die Produkte der RPZ sich laufend ergänzt und aktualisiert. Die wissenschaftliche Begleitung einzelner Schulversuche dürfte ebenfalls in den Aufgabenbereich von Institutionen auf mittlerer Ebene gehören; auch hier ist die enge Kooperation mit dem Zentralinstitut von großer Bedeutung.

Die RPZ können darüber hinaus zu wichtigen Agenturen der Lehrerweiterbildung an den einzelnen Schulen auf dem Gebiet der pädagogischen Diagnostik werden. Hier ließe sich zum Beispiel an psychometrisch ausgebildete Schulberater denken, die partiell in der Curriculumforschung mitarbeiten und am Zentralinstitut regelmäßig fortgebildet werden. Diese Personengruppe könnte auch bei den an Schulen laufenden, lokalen Curriculumprojekten für die Anwendung und Entwicklung einer angemessenen pädagogischen Diagnostik sorgen. Die von der Itembank angebotenen Hilfen dürften auch hier wiederum von größter Bedeutung sein, wie auch umgekehrt die Itembank auf Zuarbeit von der Basis angewiesen ist.

13.5.3. Lehrerbildung

Die Ansprüche einer entwickelten pädagogischen Diagnostik an die Kompetenz der Lehrer sind erheblich, zumal kein noch so sorgfältig entwickeltes Instrument den Interpreten ersetzen kann. Ebenso wichtig wie die Entwicklung einer problemadäquaten pädagogischen Diagnostik ist daher die gleichzeitige Vorbereitung der Lehrer auf angemessene Anwendung der Instrumente und auf richtige Interpretation der Befunde. Dies ist auch ein Forschungsproblem, da man derzeit nur ungenau angeben kann, über welche Kompetenzen ein Lehrer verfügen muß, um nicht nur mög-

¹⁰³ Vgl. hierzu z. B. Educational Researcher, Okt. 1972, S. 3 f.; REGLEMENT, 1971.

¹⁰⁴ Vgl. GERBAULET u. a., 1972.

lichst fehlerfrei, sondern auch umfassend urteilen zu können. Auch ist es eine offene Frage, in welcher Phase der Lehrerbildung die Einführung in die Probleme der pädagogischen Diagnostik schwerpunktmäßig erfolgen sollte.

13.6. Literatur

- AIRASIAN, P. W.: The role of evaluation in Mastery learning. In: J. H. BLOCK (Hrsg.), *Mastery learning*. New York, 1971, S. 77–88.
- ANDERSON, R. C.: How to construct achievement tests to assess comprehension. In: *Rev. Ed. Res.* 42 (2) 1972, S. 145–170.
- BLAKE, E.: Test information as a reinforcer of negative attitudes toward black Americans. In: *Proceedings of the 1970 Invitational Conference on Testing Problems*, ETS, Princeton, N. J., 1971, S. 160–169.
- BLOCK, J. H.: Criterion-referenced measurements: Potential. In: *School Review*, Febr. 1971, S. 289–298.
- BLOOM, B. S.: *Stability and change in human characteristics*. New York, 1964.
- BLOOM, B. S.: Toward a theory of testing which includes measurement-evaluation-assessment. Paper presented at the Symposium on problems in the evaluation of instruction. Los Angeles, Calif., Dec. 13–15, 1967. ERIC Reports No. ED 036878, Dep. HEW, Washington 1968.
- BLOOM, B. S.: Individual differences in school achievement: A vanishing point? New York, 1971. Deutsch: Individuelle Unterschiede in der Schulleistung: ein überholtes Problem? In: EDELSTEIN, W., und D. HOPF (Hrsg.), *Bedingungen des Bildungsprozesses. Psychologische und pädagogische Forschungen zum Lehren und Lernen in der Schule*. Stuttgart, 1973.
- BOLDT, R. F.: Anchored scaling and equating: old conceptual problems and new methods. *ETS Research Bulletin RB-72-28*, Educational Testing Service, Princeton, 1972.
- BRENNAN, R. L.: Computer assisted achievement testing in instruction. Paper presented at „Lehrsysteme 72“, Berlin, April 1972, S. 1–22.
- CARROLL, J. B.: A model of school learning. In: *Teachers College Record*, 64 (8), 1963. Deutsch: Ein Modell schulischen Lernens. In: EDELSTEIN, W., und D. HOPF (Hrsg.), *Bedingungen des Bildungsprozesses. Psychologische und pädagogische Forschungen zum Lehren und Lernen in der Schule*. Stuttgart, 1973.
- CARROLL, J. B.: School learning over the long haul. In: J. D. KRUMBOLTZ (Hrsg.), *Learning and the educational process*. Chicago, 1965. Kap. 10, S. 249–269.
- CARROLL, J. B.: Problems of measurement related to the concept of learning for mastery. In: *Educational Horizons*, Spring 1970, S. 71–80.
- CENTRA, J. A. und D. ROCK: College environments and student academic achievement. In: *Amer. Ed. Res. J.*, 8 (4), 1971, S. 623–634.
- CHOPPIN, B. H.: An item bank using sample-free calibration. In: R. WOOD und L. S. SKURNIK, *Item banking*, London 1969, S. 134–140.
- COLE, N. S.: Bias in selection. ACT research report No. 51, May 1972. The American College Testing Program, Iowa City.
- COOK, A. und H. MACK: *The teacher's role*. New York, 1971.
- CRONBACH, L. J. und P. SUPPES (Hrsg.): *Research for tomorrow's schools. Disciplined inquiry for education*, London 1969.
- DREW, C. J.: Research on the psychological-behavioral effects of the physical environment. In: *Rev. Ed. Res.*, 41 (5), 1971, S. 447–465.
- DUNN, S. S.: Die Rolle des Testens in der Lehrplanbewertung. In: K. H. INGENKAMP und Th. MARSOLEK (Hrsg.), *Möglichkeiten und Grenzen der Testanwendung in der Schule*. Weinheim, 1968.

- Educational Testing Service: Educational change: Implications for measurement. Proceedings of the 1971 Invitational Conference on Testing Problems. Princeton, 1972 a.
- Educational Testing Service: New needs, new responses. Annual Report 1969—71. Princeton, 1972 b.
- EISNER, E. W.: Emerging models for educational evaluation. In: *School Review*, 80 (4), 1972, S. 573—590.
- FINDLEY, W. G. und B. B. MASIA: Overview of the research on educational and psychological testing. In: *Rev. Educ. Res.*, 33, 1968, S. 5—11.
- FLOUD, J. und A. H. HALSEY: Social class, intelligence tests, and selection for secondary schools. In: A. H. HALSEY, J. FLOUD und C. A. ANDERSON (Hrsg.), *Education, economy and society*, New York, 1963.
- FREEBERG, N. E.: Assessment of disadvantaged adolescents: A different approach to research and evaluation measures. In: *J. Educ. Psych.*, 61, 1970, S. 229—240.
- FRICKE, R.: Testgütekriterien bei lernzielorientierten Tests. Ein Maß zur Bestimmung von Objektivität, Zuverlässigkeit, Gültigkeit und Trennschärfe bei lernzielorientierten Tests. In: *Zeitschrift für erziehungswissenschaftliche Forschung*, H. 3, 1972.
- GAGE, N. L.: *Teacher effectiveness and teacher education. The search for a scientific basis.* Palo Alto, 1972.
- GERBAULET, S., O. HERZ, L. HUBER, K. NEVERMANN, C. PETRY, H.-H. PISTOR, J. RASCHERT, I. RICHTER, H. RIENITS: *Schulnahe Curriculumentwicklung. Ein Vorschlag zur Errichtung Regionaler Pädagogischer Zentren mit Analysen über Innovationsprobleme in den USA, England und Schweden.* Stuttgart, 1972.
- GOODLAD, J. J.: Educational opportunity: The context and the reality. In: *Proceedings of the 1971 Invitational Conference on Testing Problems*, Princeton, 1972, S. 3—13.
- GOSLIN, D. A.: Ethical and legal aspects of the collection and use of educational information. In: *Proceedings of the 1970 Invitational Conference on Testing Problems*, Princeton, 1971, ETS, S. 149—159.
- HALSEY, A. H. (Hrsg.): *Educational priorities. Vol. 1: E. P. A. problems and policies.* Department of Education and Science, HMSO, London, 1972.
- HEIPCKE, K.: *Zur Theorie lernzielorientierter Tests. Teil I, curriculare Entscheidungen und lernzielorientierte Tests. Monographien zur Unterrichtsforschung Nr. II der Arbeitsgruppe für Unterrichtsforschung der Universität Konstanz*, 1970.
- HIERONYMUS, A. N.: Today's testing: What do we know how to do? In: *Proceedings of the 1971 Invitational Conference on Testing Problems*, ETS, Princeton, N. J., 1972, S. 57—68.
- HÖRMANN, H.: *Aussagemöglichkeiten psychologischer Diagnostik.* Göttingen, 1964.
- HOLTZMANN, W. H.: The changing world of mental measurement and its social significance. In: *American Psychologist*, 26, 1971, S. 546—553.
- HOPF, D.: *Übergangsauslese und Leistungsdifferenzierung. Eine Untersuchung am Beispiel der Grammar und Comprehensive Schools in England.* Frankfurt a. M., 1970.
- HOPF, D.: Das Numerus-clausus-Verfahren. Möglichkeiten, Grenzen, Folgewirkungen. In: *Neue Sammlung*, 14, 1974, S. 180—189.
- HOYT, D. P.: The relationship between college grades and adult achievement, a review of the literature. ACTP Res. Rep. No. 7, Iowa City, 1965.
- INGENKAMP, K. H.: Möglichkeiten und Grenzen des Lehrerurteils und des Schultests. In: H. ROTH (Hrsg.), *Begabung und Lernen.* Stuttgart, 1966, S. 407—432, 9. Aufl. 1973.
- INGENKAMP, K. H. (Hrsg.): *Die Fragwürdigkeit der Zensurengebung. Texte und Untersuchungsberichte.* Weinheim, 1971.
- JACKSON, P. W.: *Life in classrooms.* New York, 1968.
- JACKSON, P. W.: Die Welt des Schülers. In: W. EDELTEIN und D. HOPF (Hrsg.), *Bedingungen des Bildungsprozesses. Psychologische und pädagogische Forschungen zum Lehren und Lernen in der Schule.* Stuttgart, 1973.
- JONES, E. E. und H. SIGALL: The bogus pipeline: A new paradigm for measuring affect and attitude. In: *Psychological Bulletin*, 76, 1971, S. 349—364.

- KLAUER, K. J., R. FRICKE und M. HERBIG: Lernzielorientierte Tests. Beiträge zur Theorie, Konstruktion und Anwendung. Düsseldorf, 1972.
- KRAFFMANN, L.: Spiele und soziale Lernziele. In: DAUBLEBSKY, B., u. a., Spielen in der Schule. Stuttgart, 1973.
- LINDVALL, C. M. und J. O. BOLVIN: Programed instruction in the schools: An application of programing principles in „individually prescribed instruction“. In: Programed Instruction. 66th yearbook, II, NSSE, Chicago 1967.
- LOHNES, P. R.: Evaluating the schooling of intelligence. Educ. Researcher 2, 1973, No. 2, S. 6—11.
- MCCLELLAN, J. E.: Individualized instruction: A projection. In: Philosophical redirection of educational research. 71. Yearbook of the National Society for the Study of Education (NSSE), Part I, Chicago, 1972, S. 164—192.
- MCCLELLAND, D. C.: Testing for competence rather than for „intelligence“. In: American Psychol., Jan. 1973, S. 1—14.
- MESSICK, S.: Research methodology for educational change. In: Proceedings of the 1971 Invitational Conference on Testing Problems, ETS, Princeton, N. J., 1972, S. 69—81.
- NOVICK, M. R.: Bayesian considerations in educational information systems. In: Proceedings of the 1970 Invitational Conference on Testing Problems, ETS, Princeton, N. J., 1971, S. 77—88.
- OEVERMANN, U.: Schichtenspezifische Formen des Sprachgebrauchs und ihr Einfluß auf die kognitiven Prozesse. In: ROTH, H. (Hrsg.), Begabung und Lernen. Stuttgart 1974⁹, S. 297—355.
- POPHAM, W. J. und T. R. HUSEK: Implications of criterion-referenced measurement. In: J. Educ. Meas., 6, 1969, S. 1—9.
- QUILISCH, M. und R. SCHOBER: Bildungsforscher in der Schule. Grundlagen und Grenzen empirischer Sozialforschung im Schulbereich in verfassungsrechtlicher Sicht. Deutscher Bildungsrat, Gutachten und Studien der Bildungskommission, Bd. 22. Stuttgart 1972.
- RAUER, W.: Zur Beurteilung von Schülerleistungen in den Freien Waldorfschulen. In: Lebendige Schule, 28, 1973, S. 75—78.
- Reglement der Schweizerischen Testkommission zur Förderung der Testpsychologie und zum Schutz vor Testmißbrauch. In: Psychologie, 30, 1971, S. 340 ff.
- ROEDER, P. M.: Modelle der Differenzierung in Abhängigkeit von Leistungsdimensionen einzelner Fächer. In: Dimensionen der Schulleistung. Deutscher Bildungsrat, Gutachten und Studien der Bildungskommission, Bd. 21/1. Stuttgart 1974.
- ROSENSHINE, B.: Teaching behaviors and student achievement. Stanford, 1970.
- ROTHKOPF, E. Z.: Struktur und Prozeß: Die Steuerung der Lerntätigkeiten im Unterricht. In: W. EDELSTEIN und D. HOPF (Hrsg.), Bedingungen des Bildungsprozesses. Psychologische und pädagogische Forschungen zum Lehren und Lernen in der Schule. Stuttgart, 1973.
- ROYL, W.: Itembank — Aufbau und Funktion. In: Die Deutsche Schule, 64, 1972, S. 547 bis 564.
- SCHLOEMERKEMPER, J.: Gesamtschule im Wettbewerb. Zu: W. P. TESCHNER — Was leisten Leistungskurse? In: Die Deutsche Schule, 63 (10) 1971, S. 628—642.
- SCHREINER, G.: Sinn und Unsinn der schulischen Leistungsbeurteilung. In: Die Deutsche Schule, 62, 1970, S. 226—237.
- SCHWARZER, R.: Mastery Learning durch programmierte Instruktion? Eine Untersuchung der Beziehungen zwischen Lernerfolg, Intelligenz und Arbeitszeit beim programmierten Unterricht. Phil. Diss., Kiel 1972.
- SHULMAN, L. S.: Reconstruction of educational research. In: Rev. Educ. Res., 40 (3), 1970, S. 371—396.
- SIZER, T. R.: Testing: Americans' comfortable panacea. In: Proceedings of the 1970 Invitational Conference on testing problems, ETS, Princeton, 1971, S. 14—21.
- STAKE, R. E.: Objectives, priorities, and other judgment data. In: Rev. Educ. Res., 40, 1970, S. 181—212.

- STAKE, R. E.: National assessment. In: Proceedings of the 1970 Invitational Conference on Testing Problems, ETS, Princeton, N. J., 1971, S. 53—66.
- THELEN, H. A.: Classroom grouping for teachability. New York, 1967.
- ULICH, D. und W. MERTENS: Urteile über Schüler. Zur Sozialpsychologie pädagogischer Diagnostik. Weinheim, 1973.
- VERNON, P. E. (Hrsg.): Secondary school selection. A British Psychological Society inquiry. London, 1957.
- VEROFF, J.: Wie allgemein ist das Leistungsmotiv? In: W. EDELSTEIN und D. HOPF (Hrsg.), Bedingungen des Bildungsprozesses. Psychologische und pädagogische Forschungen zum Lehren und Lernen in der Schule. Stuttgart, 1973.
- WALBERG, H. J.: Models for optimizing and individualizing school learning. In: *Interchange*, 2 (3), 1971, S. 15—27.
- WALBERG, H. J. und G. J. ANDERSON: Properties of the achieving urban class. In: *J. Ed. Psych.* 63 (4), 1972, S. 381—385.
- WEIS, V.: Zensierungsmodelle und ihre pädagogischen Konsequenzen. In: *Die Deutsche Schule*, 63, 1971, S. 542—553.
- WOLF, R.: The measurement of environments. In: ANASTASI, A. (Hrsg.), *Testing problems in perspective*. Washington 1966, S. 491—503.
- WOOD, R.: The item bank project. A pilot study of an alternative method of monitoring achievement. Preliminary draft report, National Foundation of Educ. Research, London, Jan. 1968 (mimeo).
- WOOD, R. und L. S. SKURNIK: Item banking. National Foundation of Educ. Research, London, 1969.