

# Welt der Schule

Zeitschrift  
für Lehrer  
in der Grundschule

Ehrenwirth  
Verlag  
München

## Herausgeber:

Professor Dr. Dr. Hans-Günther Assel  
Erste Erz.wiss. Fakultät  
d. Univ. Erlangen-Nürnberg  
Oberschulrat Wilhelm Berger  
Bremen

Dr. Alfons Brendel, München  
Seminarlehrer an Realschulen

Studienrat Hermann Bühnemann  
Oldenburg/H.

Professor Dr. Helmut Donat  
Pädagog. Hochschule Bremen

Professor Dr. Wolfgang Fischer  
Gesamthochschule Duisburg

Professor Dr. Heinz-Jürgen Ipfling  
Erziehungswiss. Hochschule  
Rheinland-Pfalz, Abt. Koblenz

Stud.-Dir. Walter H. Leicht  
Erz.wiss. Fakultät d. Univ. Würzburg

Professor Dr. Anton Neuhäusler  
Erz.wiss. Fakultät d. Univ. München

Verw.-Dir. Richard Reiniger  
München

Professor Dr. Horst Scarbath  
Universität Hamburg

Professor Dr. Hans Schiefele  
Universität München

Min. Rat Dr. Franz Otto Schmaderer  
Bayer. Staatsministerium für  
Unterricht und Kultus

Professor Dr. Erich Weber  
Philos. Fachbereich I  
Universität Augsburg

## Ständige Mitarbeiter:

Oberschulrat Alfred Angermeyer  
Neustadt a. d. Aisch, Lehrbeauftragter  
der Ersten Erz.wiss. Fakultät  
d. Univ. Erlangen-Nürnberg

Reg.-Schulrat Berthold Casper  
Ansbach

Rektor Georg Gick, München

Professor Dr. Hans Glöckel  
Dipl.-Psych., Erste Erz.wiss. Fakultät  
d. Univ. Erlangen-Nürnberg

Rektor Dr. Friedrich Lehmann  
München

Professor Dr. Hermann Oblinger  
Erz.wiss. Fachbereich  
d. Univ. Augsburg

Regierungsschuldirektor  
Max Scharnagl, München

Dr. Helmut Seiffert, Akad. Oberrat  
Universität Erlangen-Nürnberg

Rektor Franz Seilnacht,  
Lehrbeauftragter d. Erz.wiss. Fakultät  
d. Univ. München

Dr. Kurt Singer, Psychotherapeut,  
Oberstudienrat, Erz.wiss. Fakultät  
d. Univ. München

Oberlehrerin Maria Solcher  
Ausbildungslehrerin in München

Oberschulrat Fritz Vogtmann  
Ansbach

1

Januar 1974  
27. Jahrgang

## Schulleistungsmessung als pädagogisches Problem

---

### Vorbemerkung

Viele der hier wiedergegebenen Erkenntnisse beruhen auf Forschungsergebnissen der Projektgruppe Schullaufbahnberatung am Pädagogischen Institut II der Universität München und am Institut für Unterrichtsforschung der Universität Augsburg, die vom Frühjahr 1970 bis Ende 1973 im Auftrag des Bayerischen Staatsministeriums für Unterricht und Kultus und mit Unterstützung des Bundesministeriums für Bildung und Wissenschaft Methoden zur Laufbahnberatung von Schülern der bayerischen Orientierungsstufe erarbeitete, und deren Mitarbeiter der Verfasser war. Zur Abklärung der Probleme in Teil II haben vor allem die langen Gespräche und Diskussionen in der „Arbeitsgruppe Curriculare Tests“ (ACT), deren Mitglieder *K. Kunert*, *H. J. Schulte*, *G. Seisenberger* und der Verfasser sind, beigetragen.

### Teil I: Zum Stand der Diskussion um eine gerechte Schülerbeurteilung

**1. Die gegenwärtige Beurteilungspraxis der Lehrer steht im Kreuzfeuer einer kritischen Öffentlichkeit. In Kenntnis der Bedeutung von Zeugnissen und Gutachten der Schulen für das Erreichen bestimmter sozialer Positionen wird eine größere Transparenz und Nachprüfbarkeit von Schülerbeurteilungen verlangt**

Vor fast eineinhalb Jahrzehnten hat der Soziologe *Schelsky* einen inzwischen zum geflügelten Wort gewordenen Ausdruck geprägt, als er die Schule eine „entscheidende zentrale soziale Dirigierungsstelle“ und „bürokratische Zuteilungsapparatur von Lebenschancen“ (*Schelsky*, 1957, 17 f.) nannte. Gesagt soll damit werden, daß die Schule eine oft lebenslänglich wirksame Weichenstellung für das spätere Leben der Kinder vornimmt, indem die von ihr verteilten Zertifikate (Gutachten, Zeugnisse, Beurteilungen) für bestimmte Ausbildungs- und Berufszweige qualifizieren, die in unserer Gesellschaft mit unterschiedlichem Prestige behaftet sind und Einkommens- und Wohlstandsunterschiede bewirken. Verständlich erscheint der Wille der Eltern, alles in ihren Kräften stehende zu tun, um ihren Kinder die Chancen für das Erreichen höchstmöglicher sozialer Positionen zu verwirklichen; verständlich deshalb auch die Wogen des Protests, die nicht nur über die Schwellen der Schultür, sondern auch in die Spalten der Presse schwappen, wenn alljährlich die Entscheidungen für den Übertritt in weiterführende Schulen getroffen werden müssen. Bis vor kurzem stand im Zentrum der öffentlichen Kritik, z. B. in Bayern, einzig der beurteilende Lehrer der abgebenden Schule, bis zu jenem Zeitpunkt, als das „objektivere“, „gerechtere“ Verfahren mit regionalen Übertrittstests geschaffen wurde, das teilweise für eine Richtungsänderung der Kritik sorgte. Nach wie vor kommt jedoch dem Lehrer bzw. dem Kollegium die entscheidende Rolle bei der Zuweisung der Schüler in verschiedene Schularten zu.

Womit begründet sich der wachsende Unmut an der gegenwärtigen Beurteilungspraxis? Alle, die wir Schulen besucht haben oder noch besuchen, haben unsere Erfahrungen mit ungerechten Zensuren aus eigener Betroffenheit gemacht. Meist fehlte jedoch die Möglich-

keit, die subjektiv empfundene Ungerechtigkeit zu artikulieren, um eine Korrektur des Urteils zu erreichen. Dazu wäre die Kenntnis des angelegten Maßstabs notwendig gewesen, der — oft aus guten Gründen — von den wenigsten Lehrern bekanntgegeben wurde (es sei denn, man glaubte, der „Rötegrad“ einer korrigierten Schularbeit erfüllte diese Funktion). Wenn heute eine größere Transparenz und Nachprüfbarkeit in der Schülerbeurteilung gefordert wird, so geschieht dies mit Unterstützung wissenschaftlich-empirischer Forschungsergebnisse, die in den letzten zehn Jahren publiziert wurden. Davon seien nur einige erwähnt, auf die immer wieder zurückgegriffen wird, wenn die Fragwürdigkeit des Lehrerurteils zur Diskussion steht:

— Weiss (1965), der u. a. nachweisen konnte, daß Lehrer einen anonymen Aufsatz besser bewerteten, wenn sie vorher die Information bekamen, er stamme von einem sprachlich begabten Kind (Vater ist Redakteur), als wenn es hieß, ein durchschnittlicher Schüler, der gern Schundhefte liest und dessen beide Elternteile berufstätig seien, habe den Aufsatz geschrieben;

— Schröter (1970, 1971), der neunzehn Lehrern in allen Teilen der Bundesrepublik und Westberlin ein und denselben Aufsatz korrigieren ließ und mit „sehr gut“ bis „ungenügend“ beurteilt zurückbekam;

— Ingenkamp, der nach seinen zahlreichen Untersuchungen und Veröffentlichungen (etwa zur Frage der Vergleichbarkeit von Noten aus verschiedenen Klassen) als anerkannter Experte für den Problembereich „Schulleistungsmessung“ in Deutschland gilt (vgl. z. B. Ingenkamp 1969 a, b; 1971 a, b; 1973).

Es mangelt auch nicht an Theorien, die das Phänomen der unterschiedlichen Beurteilung desselben Gegenstands durch verschiedene Menschen zu erklären suchen.

## **2. Die Sozialpsychologie hat Fehler und Tendenzen der Wahrnehmung nachgewiesen, die auch auf den beurteilenden Lehrer zutreffen: Halo-Effekt; logischer Fehler; zentrale Tendenz; Leniency-Effekt; implizite Persönlichkeitstheorien**

In unserem Alltagsleben sind wir ständig zur Beurteilung unserer Umwelt aufgefordert: Wir urteilen über Ereignisse des Zeitgeschehens, über uns nahestehende Menschen, deren Kleidung, Gebaren, Aussehen und Äußerungen, über Autos, Kunstwerke, Architektur u. v. a. m. Unsere Urteile beruhen auf Beobachtungen, die sich aus einzelnen „objektiven“ Wahrnehmungen zusammensetzen. Die sozialpsychologische Forschung hat mittels Wahrnehmungsexperimenten jedoch nachgewiesen, daß es keine „objektiven“ Daten gibt. Selbst wenn dem Menschen relativ einfache Beobachtungsaufgaben gestellt werden, neigt er dazu, etwas in die Dinge hineinzusehen, d. h. sie zu interpretieren. Es würde an dieser Stelle zu weit führen, die höchst interessanten Ergebnisse jenes Forschungszweiges zu referieren, der sich mit den Problemen der sozialen Wahrnehmung (social perception) und der Personwahrnehmung (person perception) befaßt. Auf zwei verständlich geschriebene Arbeiten zu diesem Thema sei hingewiesen: *Österreich* (1969) und *Ulich & Mertens* (1973, 60—127). Begnügen wir uns mit den folgenden wissenschaftlich belegten Erkenntnissen:

● Es konnte festgestellt werden, daß unsere Wahrnehmung der Tendenz unterliegt, sich in bezug auf das Wahrgenommene eine stabile Umgebung zu schaffen und das Wahrgenommene mit Bestimmtheit zu identifizieren. Beides hängt wohl mit dem menschlichen Streben nach Sicherheit zusammen, mit dem Wunsch, vorbereitet zu sein auf das, was zu erwarten ist. Wenn der Mensch nicht klar wahrnehmen kann, beginnt er zu raten oder Reizkonfigurationen auf eine für ihn persönlich charakteristische Weise zu interpretieren.

● Die Erwartungs- oder Hypothesentheorie der Wahrnehmung besagt, daß die Art der Wahrnehmung das widerspiegelt, was der Mensch an Einstellungen internalisiert hat; daß das Wahrgenommene ein unbewußter Kompromiß zwischen dem, was faktisch an

Gegenständen und Sachverhalten der Umwelt vorfindbar ist, und dem, was der Mensch wahrzunehmen erwartet. Die Erwartung kann die Wahrnehmung in vierfacher Weise beeinflussen:

a) Sie liest aus (Selektionsfunktion).

*Beispiel:* Ein Lehrer, der positive oder negative Erwartungen bezüglich eines Schülers hat, bemerkt tatsächlich gegenteilige Reize nicht.

b) Sie gestaltet sinnvoll (Funktion der sinnvollen Interpretation).

*Beispiel:* Der Lehrer, der in einer Klasse mit rund 40 Kindern bemerkt, daß sich der Schüler Hans unruhig verhält, wird diese — beiläufig und unklar aufgenommene — Information sofort zu dem sinnvollen „Hans stört wieder“ verarbeiten, wenn er die Hypothese hat, daß Hans ein notorischer Unruhestifter ist, oder aber zu dem sinnvollen „Hans hilft seinem schwächeren Nachbarn“, wenn seine Hypothese über Hans „ein zuverlässiger, hilfreicher Schüler“ lautet.

c) Sie setzt Schwerpunkte (Akzentuierungsfunktion).

*Beispiel:* Ein Lehrer, dessen Hobby die Botanik ist, wird auf einem Klassenwandertag seinen Kindern wohl andere Beobachtungsaufträge erteilen als einer, der dieses Steckenpferd nicht reitet.

d) Sie legt fest (Fixierungsfunktion).

*Beispiel:* Ein Schulleiter, der die vorgefaßte Meinung hat, die Junglehrerin X habe kein pädagogisches Geschick, wird seine Einstellung immer wieder bestätigt sehen, wenn er an der Klassenzimmertür der besagten Erzieherin vorbeigeht und Lärm vernimmt. Sollte aber dennoch gerade Ruhe hinter dieser Tür herrschen, so wird dies entweder gar nicht wahrgenommen oder die Informationserfassung wird zumindest erschwert.

● Gerade das letzte Beispiel hat auf die enge Verknüpfung von Urteil und Wahrnehmung verwiesen. Daß es persönlichkeitspezifische Tendenzen auch bei Beurteilungsprozessen gibt, liegt auf der Hand. Vier der wichtigsten Beurteilungsfehler, wie sie die psychologische Forschung aufgewiesen hat, sollten zur Problematisierung ausreichen:

a) Der Halo-Effekt (Hof- oder Heiligenschein-Effekt): Ein Gesamteindruck wird für die Beurteilung von spezifischen Eigenschaften herangezogen.

*Beispiel:* Ein Lehrer hat allgemein einen guten Eindruck von einem Schüler (nettes Äußeres, ordentliche Umgangsformen), kennt ihn nicht genau, kennzeichnet aber ohne weitere Nachforschung einzelne Charakterzüge des Schülers für ein Schülergutachten positiv.

b) Der logische Fehler: Der Beurteiler hat eine eigene, laienhafte Persönlichkeitstheorie und hält es für logisch richtig, daß bestimmte Persönlichkeitsmerkmale immer miteinander gekoppelt sind.

*Beispiel:* Viele Lehrer glauben, daß Kinder aus gebildeten Familien bei schlechten Leistungen „eigentlich mehr leisten könnten, wenn sie nur wollten“, während bei Kindern aus einfachen Familienverhältnissen eine schwache Leistung oft als normal und angemessen angesehen wird.

c) Die zentrale Tendenz: Der Beurteiler geht den Extremen einer gegebenen Beurteilungsskala aus dem Weg und fällt seine Urteile überwiegend im Mittelbereich.

**Beispiel:** Die Noten 1 und 6 werden u. a. deshalb weniger häufig verteilt als die Noten 3 oder 4, da sie vom Lehrer eher eine Rechtfertigung abverlangen.

d) Der Leniency-Effekt (leniency, engl. = Milde): Der Beurteiler gibt zu gute Urteile ab, wenn ihm der Beurteilte gut bekannt oder sympathisch ist. Bei bewußter Antipathie kommt es zu ausgleichenden Einzelurteilen. Es scheint bequemer, etwas Gutes über jemanden zu sagen, als zu differenzierenden Aussagen zu gelangen.

**Beispiel:** Der Lehrer versucht, bei einem Schüler, dem er wohlgesonnen ist, im Schulabgangszeugnis die schlechten Noten durch Hinzufügung einer erläuternden Bemerkung abzumildern, weil er weiß, daß u. U. das Zustandekommen eines Lehrvertrages davon abhängt.

● Eine weitere Einflußgröße auf das Urteilsverhalten ist die „implizite Persönlichkeitstheorie“ der Beurteilenden. Darunter ist ein kognitives Orientierungskonzept aus Erwartungshaltungen zu verstehen, die während der Lerngeschichte, der Sozialisation erworben werden. Solche Erwartungen bestehen in bestimmten Auffassungen „1. über die ‚Natur des Menschen‘, über Charaktertypen usw.; 2. über das Vorkommen von bestimmten Eigenschaften bei bestimmten Leuten, vom nationalen Vorurteil bis zum Stereotyp des ‚schlechten Schülers‘; 3. über bestimmte Ursache-Wirkungsverhältnisse zwischen Verhaltensäußerungen und dem zugrundeliegenden ‚Charakter‘; 4. über ebenfalls ‚gesetzmäßige‘ Zusammenhänge zwischen bestimmten Aktionen und Reaktionen (beim Faulen hilft nur Druck, usw.)“ (Ulich & Mertens, 1973, 102). Daß solche Auffassungen der objektiven Überprüfung nicht standhalten können, muß nicht eigens betont werden.

Wer sich auf die Suche nach einem Erklärungsprinzip für die Entstehung von Wahrnehmungs- und Urteilstendenzen begibt, stößt zwangsläufig auf einen Bereich menschlichen Verhaltens, der, bedingt durch den starken Einfluß des Behaviorismus, jahrzehntelang ein Stiefkind der psychologischen Forschung war: die emotionale Dimension der Persönlichkeit. Der Grund für die Vernachlässigung liegt sicher in der Schwierigkeit, Gefühlsphänomene zu beschreiben und von anderen psychischen Phänomenen zu unterscheiden, da alle unsere Handlungen von Gefühlen begleitet sind. Ein gefühlsorientierter Verhaltensimpuls muß als Antwort auf die Auseinandersetzung mit der Umwelt, die als gut oder schädlich für die Befriedigung von vitalen, geistigen und sozialen Bedürfnissen bewertet wurde, angesehen werden (vgl. Roth, 1971, 320). Der gefühlsmäßige Handlungsimpuls wird entweder ausgeführt oder unterdrückt, je nachdem, wie das sekundäre verständnismäßige Urteil ausfällt. Die Tatsache, daß der vom Gefühl geleitete Impuls gleichsam die Barriere des rationalen Werturteils passiert, und die intuitive Einschätzung dadurch unwillkürlich beeinflusst wird, deutet auf die Erziehbarkeit von Gefühlen hin. Im Zusammenhang mit der Beurteilung von Schülern durch den Lehrer ließe sich nun folgende Hypothese aufstellen: Das Wissen um menschliche Wahrnehmungstendenzen und Urteilsfehler bewirkt subtilere Werturteile, die indirekt die Einstellung gegenüber dem Beurteilten verändern können.

Die Reflexion der gegenwärtigen Beurteilungspraxis vor dem Hintergrund der Mechanismen sozialer und personspezifischer Wahrnehmung und Beurteilung läßt es im Moment jedoch recht und billig erscheinen, dem Lehrer die Kompetenz für eine gerechte Beurteilung abzusprechen.

### 3. Psychometrische Verfahren (Tests) aller Art werden zusehends häufiger in den Dienst der Schule gestellt. Sie können „Gütekriterien“ aufweisen, die das Lehrerurteil angeblich nicht hat: Objektivität; Gültigkeit (Validität); Zuverlässigkeit (Reliabilität); Norm; Ökonomie

Eine vorurteils- und fehlerfreie Kennzeichnung der Persönlichkeit ist von jeher die vornehmliche Aufgabe der Psychologie. Diese Wissenschaft hat eine unüberschaubare Menge psychometrischer Verfahren produziert, mit deren Hilfe — zu welchem Zweck auch immer — Aufschluß über die Persönlichkeitsstruktur des Menschen gewonnen werden soll. Da gibt es Intelligenz- und Begabungstests, Interessentests, Einstellungstests, Angsttests, Motivationstests, Eigenschaftstests, Charaktertests, Typentests, motorische und sensorische Leistungstests, Berufseignungstests, Kreativitätstests, Materialbearbeitungstests usf. Es wurde eine mathematisch-statistische Teildisziplin der Psychologie entwickelt: die Testtheorie, die allein mit den Problemen der Testentwicklung und Testanalyse beschäftigt ist. Ein guter Test soll gewissen Gütekriterien genügen. Diese Kriterien der Testgüte sind in der Hauptsache: Objektivität, Validität, Reliabilität, Norm und Ökonomie (vgl. hierzu *Lienert, 1969*).

*Objektivität* besagt, daß der Test unabhängig von der Person des Durchführenden, des Auswertenden und des Interpretierenden sein soll, d. h. daß verschiedene Untersucher bei seiner Anwendung bei ein und derselben Person zu übereinstimmenden Ergebnissen und Interpretationen kommen müssen. Objektivität wird dadurch erreicht, daß die Anleitung zur Durchführung schriftlich festgelegt und somit die Untersuchungssituation möglichst standardisiert wird; daß genaue Schlüssel für die Auswertung beigegeben werden; daß Angaben über die Bewertung einzelner Ergebnisse gemacht werden; daß die Ergebnisse eindeutig und verständlich formuliert werden.

*Validität* (Gültigkeit) besagt, daß der Test auch wirklich das mißt, wofür er konstruiert wurde, d. h. daß etwa ein Rechentest mit sog. Textaufgaben wirklich nur die mathematische und nicht die Lesefertigkeit oder das Sprachverständnis erfaßt. Zum Nachweis der inhaltlichen Gültigkeit werden die Meinungen von Experten eingeholt, außerdem werden Vergleiche mit anderen, für denselben Zweck bereits entwickelten und als geeignet empfundenen Instrumenten gezogen (Kriteriumsvalidität).

*Reliabilität* (Zuverlässigkeit) bezieht sich auf die Meßgenauigkeit des Tests, d. h. es wird gesagt, inwieweit ein erzielt Ergebnis fehlerlos beschrieben oder auf einer Skala lokalisiert werden kann. Das bedeutet auch, daß bei wiederholtem Testen unter gleichen Bedingungen bei denselben Versuchspersonen dieselben Meßwerte erzielt werden müssen. Nachgewiesen wird die Zuverlässigkeit durch parallele Testformen und/oder Testwiederholung und/oder Konsistenzanalyse.

*Norm* heißt, daß für jedes Testergebnis eine Einordnung in ein bestimmtes Bezugssystem möglich ist, daß jedem Testrohwert ein bestimmter Standardwert entspricht, an dem abgelesen werden kann, welche Position der Getestete im Vergleich zu anderen einnimmt. Notwendig hierzu ist das Wissen um die Verteilung der Testwerte einer größeren, nach Möglichkeit repräsentativen Gruppe von Untersuchten.

*Ökonomie* betrifft den Aufwand an Zeit und Material in Relation zum Zweck seiner Anwendung. Gefordert werden einfache Handhabbarkeit, schnelle Auswertbarkeit.

Wenn nun bereits Tests existieren, die objektiv, valide und reliabel sind, liegt es da nicht nahe, sie für den schulischen Sektor nutzbar zu machen? Die bereits erwähnte Kritik am schulischen Beurteilungswesen sowie der verständliche Wunsch der Lehrerschaft nach Ent-

lastung bei der Beurteilung von 30, 40 und mehr Kindern in einer Klasse haben in den sechziger Jahren dazu geführt, daß immer mehr standardisierte Tests in den Dienst der Schule gestellt wurden. Diese Entwicklung erfolgte unter dem Schlagwort „objektivierte Schulleistungsmessung“. Neben der Kritik am schulischen Prüfungswesen an sich und dem Verlangen der Lehrer nach Arbeitserleichterung trugen weitere Faktoren zur Beschleunigung dieses Prozesses bei:

- die Wende von der geisteswissenschaftlich-normativen Pädagogik zur empirischen Erziehungswissenschaft, die sich als eine Sozialwissenschaft versteht;
- der Einfluß der angelsächsischen Pädagogischen Psychologie auf unsere Pädagogen und Schulpolitiker: in den USA gibt es z. B. einen gewaltigen Dienstleistungsapparat für Tests in Schule und Hochschule, den „Educational Testing Service“ (ETS), der dafür sorgt, daß sich jeder in der Ausbildung stehende amerikanische Jugendliche durchschnittlich bis zu fünf standardisierten Tests im Jahr unterziehen muß (vgl. *Kirkland*, 1971);
- der Ruf nach Ausschöpfung der Begabungsreserven und der hohe Anteil an Versagern im weiterführenden Schulwesen, was in der Forderung nach höherer prognostischer Effizienz von Übertrittsgutachten sich ausdrückte;
- die Forderung nach Verbesserung der Chancengleichheit im Bildungswesen.

Inzwischen haben auch hierzulande Tests verschiedener Art ihren festen Platz im Schulalltag eingenommen und weitere Verlage (neben den auf dem Testsektor schon immer publizierenden Verlagen wie Hogrefe und Beltz seit neuerer Zeit auch Westermann, Klett und Wolf) zur Diversifikation ihres Programms angeregt. Diese Tatsache sollte nicht darüber hinwegtäuschen, daß nicht alles sinnvoll und pädagogisch vertretbar ist, was in unseren Schulen an Testverfahren zur Anwendung kommt. Der Beweis für diese Behauptung soll mit Hilfe einiger testtheoretischer Erwägungen angetreten werden (zur ausführlichen Kritik s. *Frenz, Krüger & Tröger*, 1973 und *Projektgruppe Schullaufbahnberatung*, 1973).

#### 4. Eine kritische Analyse herkömmlicher Tests und ihrer testtheoretischen Grundlagen läßt die Überlegenheit gegenüber dem Lehrerurteil fragwürdig erscheinen

##### *Testobjektivität schränkt die Schulwirklichkeit ein*

Zunächst einmal ist festzuhalten, daß jeder Test aus der unendlichen Vielzahl möglicher Verhaltensweisen eines Menschen nur einen kleinen Ausschnitt herausgreift und zu beschreiben versucht. Zur Beschreibung ist es notwendig, komplexes, vielschichtiges Verhalten möglichst so in Einzelbeobachtungen zu zerlegen, daß es intersubjektiv eindeutig kommunizierbar, d. h. objektiv wird. Weiterhin wird angestrebt, störende Faktoren auszuschalten, d. h. die interessierende Verhaltensstichprobe möglichst in Reinform, frei von im Moment nebensächlichen Verhaltensweisen zu erheben. Erreicht wird solche Testobjektivität durch eine restriktive Formulierung der Testaufgaben (indem etwa Antworten zur Auswahl vorgegeben werden) und durch standardisierte Verhaltensvorschriften (indem etwa Möglichkeiten der Rückfrage durch Probanden unterbunden oder indem genaue Markierungs- bzw. Antwortanleitungen getroffen werden). Übersehen wird dabei meist, daß derlei Restriktionen mit einer konsequenten Entfremdung von natürlichen Situationen des Verhaltens einhergehen, ein Umstand, den der kritische Psychologe *Holzkamp* (1972) in seinen Überlegungen „Zum Problem der Relevanz psychologischer Forschung für die Praxis“ hervorgehoben hat. Für die in den Schulen angewandten objektiven Prüfverfahren trifft dies in besonderem Maße zu. Eine Analyse gängiger Schulleistungstests zeigt, daß hier zugunsten der Testobjektivität meist nur recht einfache und leicht abfragbare Kenntnisse und Fertigkeiten abgedeckt werden, und

das oft in so „gekünstelten“ Situationen, daß es dem Lehrer schwerfallen wird, Parallelen zu seinem eigenen Unterricht zu ziehen. Außerdem findet er während seiner Arbeit mit den Kindern keine wirklichkeitsnahe Gelegenheit, frei von Störfaktoren seine Beobachtungen anstellen zu können. Die Objektivität des Tests ist demnach eine ganz andere als diejenige, die dem Lehrer abgesprochen wird: Die Testobjektivität bezieht sich auf eine verkürzte, eingeschränkte Schulwirklichkeit, die unabhängig von der Angemessenheit des Inhalts betrachtet werden kann. Ausgesprochen schädlich wäre, wenn eine vom Lehrer mehr oder weniger bewußt wahrgenommene Diskrepanz zwischen Tests und Unterricht zum Anlaß diene, den Unterricht auf die Bedürfnisse des Tests hin abzustellen anstatt, was m. E. sinnvoller und für nachfolgende pädagogische Maßnahmen fruchtbarer erschiene, den umgekehrten Weg zu beschreiten.

*Inhaltliche und pädagogische Angemessenheit sind nicht immer gewährleistet*

In den Überlegungen zur Testobjektivität ist bereits angeklungen, daß den Aufgaben vieler herkömmlicher Tests die Strukturähnlichkeit mit den Zielsetzungen des Unterrichts fehlt. Das gilt insgesamt für solche Prüfverfahren, die sogenannte allgemeine Fähigkeiten testen (Intelligenztests, Lesetests, Wortschatztests, Begabungstests, Analogie-tests u. ä.), und teilweise auch für Schulleistungstests, die wegen ihrem überregionalen Anwendungsbereich wenig Rücksicht auf die Curricula einzelner Bundesländer oder gar einzelner Landkreise und Schulen nehmen können. Nun läßt sich mit Recht einwenden, die Notwendigkeit des engen inhaltlichen Zusammenhangs zwischen Tests und Unterricht müsse erst einmal aufgezeigt werden; denn schließlich könnten mit den genannten Tests doch tatsächlich „gute“ von „schlechten“ Schülern unterschieden werden. Dieses Problem kann nicht unabhängig von der Entscheidung über den Zweck des Testens erörtert werden.

Soll mit Hilfe des Tests irgendeine Aufteilung der Schüler (etwa in bestimmte Schulzweige) vorgenommen werden, so läßt sich das aufgrund seiner Trennschärfe mit Leichtigkeit erreichen. Übrig bleibt allerdings der Zweifel, ob das vom Test anvisierte Kriterium (z. B. Intelligenz, Analogiebildungsfähigkeit, Wortschatz, Lesefertigkeit, Rechenbegabung) dasjenige ist, das den erfolgreichen Besuch des einen oder anderen Schulzweiges garantiert. In Testmanualen werden zur Ausräumung solchen Argwohns Koeffizienten angeführt, die sich durch Korrelationsrechnung zwischen den Testergebnissen und z. B. den Resultaten ähnlicher Tests oder den Zeugnisnoten in späteren Jahren ergeben. Welchen geringen Aussagewert jedoch solche Bewährungskoeffizienten haben, wurde an anderer Stelle am Beispiel des Regensburger „Schulleistungs-Prüfsystems auf Testbasis“ veranschaulicht (vgl. *Mauermann*, 1973, 332 f.).

Soll der Test Auskunft über die Stärken und Schwächen der einzelnen Schüler geben, damit kompensatorische und fördernde Maßnahmen für das zielerreichende Lernen getroffen werden können, dann muß er so eng wie möglich auf die Belange des stattgefundenen Unterrichts hin zugeschnitten sein (s. *Kutscher, Raeithel & Stevens*, 1973). Nur so kann der Lehrer Zusammenhänge zwischen dem Lernerfolg seiner Schüler und dem von ihm geplanten und durchgeführten Unterricht aufdecken und eventuell notwendig werdende Differenzierungen in den Methoden mit größerer Aussicht auf Effizienz vornehmen. Gleichzeitig wäre so ein Instrument, das als „curricularer Test“ bezeichnet werden soll, ein gutes Hilfsmittel für die Verbesserung des Unterrichts (Curriculumevaluation). Für solche, eher pädagogisch zu nennende Zwecke sind herkömmliche Testverfahren jedoch nicht oder nur sehr bedingt geeignet.

### *Koeffizienten für Zuverlässigkeit lassen sich künstlich in die Höhe schrauben*

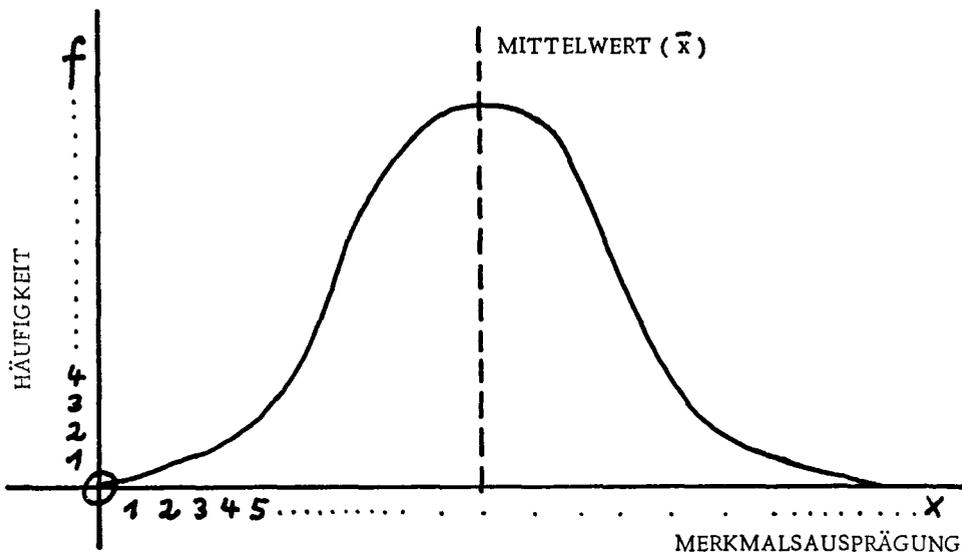
Wenden wir uns nach diesen Überlegungen zur Gültigkeit und Angemessenheit von Tests der Frage nach ihrer Zuverlässigkeit zu. Es ist in diesem Zusammenhang kaum angebracht, auf die subtilen Methoden der Reliabilitätsbestimmung von Tests einzugehen. Auf ein einfaches Rechenexempel kann verwiesen werden (s. *Mauermann*, 1973, 332). Ein Blick in die „Trickkiste“ des Testkonstruktors möge die Aufmerksamkeit auf die Entstehungszusammenhänge von Reliabilitätsschätzungen lenken. Die wenigsten Testverfahren sind bei ihrer Ersterprobung auf Antrieb reliabel. Dem kann aber rein testtechnisch abgeholfen werden (vgl. *Lienert*, 1969, 242—246). Zunächst einmal läßt sich ein besserer Koeffizient durch die Erhöhung der Aufgabenzahl erreichen. Wenn z. B. ein Test mit 20 Aufgaben einen  $r_{tt} = 0,70$  erbracht hat, so kann mittels einer einfachen Formel errechnet werden, daß in unserem Beispiel die Zahl der Aufgaben auf 77 (von der gleichen Art wie die vorhandenen) erhöht werden müßte, damit ein voll befriedigender  $r_{tt}$  von 0,90 zustandekommt. Weitere Möglichkeiten zur Verbesserung der Zuverlässigkeit sind: noch stärkere Objektivierung von Durchführung und Auswertung, Verwendung von Items mittlerer Schwierigkeit, Zufallskorrektur der Rohwerte. In Kenntnis dieser Manipulationsmöglichkeiten und dem damit verbundenen Relevanzschwund für die Praxis sollten Reliabilitätskoeffizienten doch etwas nüchterner betrachtet werden.

### *Normentabellen stehen im Widerspruch zur rasch fortschreitenden Entwicklung des Unterrichts*

Die in den Testmanualen abgedruckten Normentabellen geben die Ergebnisse für die sogenannte Eichstichprobe wieder, die repräsentativ für eine bestimmte Population ist. Mit Hilfe der Normentabellen wird es möglich, jedes Testergebnis in Relation zu seiner Bezugsgruppe zu sehen (wenn etwa nach Alter, Geschlecht, sozialer Herkunft, Region, Schultyp normiert worden ist). Selbst wenn wir davon absehen, daß die Information über den individuellen Rangplatz dem Lehrer keine Aussage über Bedingungen für das Zustandekommen der Leistung und damit keinen Anhaltspunkt für differenzierende Unterrichtsmaßnahmen gibt, steht eine Normierung im Widerspruch zur ständig fortschreitenden Entwicklung etwa im Bereich der Unterrichtstechnologie, die ja ihren Sinn verlöre, wenn man sich von ihr keine Effektivitätssteigerungen des Lernens erwartete. Mit anderen Worten, die Normen veralten gerade im Schulbereich sehr rasch. Es ist empirisch nachgewiesen, daß sowohl Schulleistungstests als auch Intelligenztests häufig schon wenige Jahre nach ihrer Normierung andere Durchschnittsleistungen erbringen (vgl. *Frenz, Krüger & Tröger*, 1973, 86 f.). Ein am Profit interessierter Testverlag wird verständlicherweise die Kosten für aufwendige Neueichung der Normentabellen scheuen, zumindest solange sich das Produkt gut verkauft.

### *Die GAUSSsche Normalverteilungskurve als Richtmaß für Schülerleistung hat fatale Konsequenzen*

Ein zentrales Konzept der Testtheorie ist die *Gaußsche* Normalverteilung. Von Testdaten wird erwartet, daß sie sich im Mittelbereich häufen, während in den Extrembereichen geringere Häufigkeit auftreten soll. In graphischer Darstellung bildet sich der Zusammenhang zwischen Merkmalsausprägung und Häufigkeit in Gestalt der sogenannten Glockenkurve ab:



Die Glockenkurve gründet auf Beobachtungen naturgegebener Phänomene. Beispielsweise wird sich die Statistik der Schuhgröße, des Gewichts oder der Körperlänge einer Zufallsstichprobe von Menschen einer annähernd normalen Verteilung anpassen. Ob das aber für geistige Fähigkeiten genauso gilt, scheint nicht bewiesen. Die neueren Ergebnisse der Sozialisationsforschung (vgl. *Gottschalch, Neumann-Schönwetter & Soukup, 1971* oder *Skowronek, 1973*) lassen die Grenzen der individuellen Lernfähigkeit offen. Versuche mit einer Strategie des „zielerreichenden Lernens“ (mastery learning) haben gezeigt, daß unter idealisierten Bedingungen Lernerfolg für 80 und mehr Prozent aller Schüler möglich ist (s. *Bloom, 1970; Carroll, 1972; Block, 1971*). Einen Lehrer, der seinen Unterricht rational plant und durchführt zum Zweck, möglichst alle Schüler die gesetzten Lernziele erreichen zu lassen, muß die Erwartung normalverteilter Schülerleistungen geradezu entmutigen. Am Rande sei vermerkt, daß die Testtheorie für den Fall anormal verteilter Ergebnisse beim Test Rezepte zur „Normalisierung“ bereithält (Items mittlerer Schwierigkeit bevorzugen, Aufgaben umgruppieren, Items verschieden gewichten usw.). Auch hier besteht also die Gefahr, daß ein mathematisch-statistisches Modell der Wirklichkeit aufgedrängt wird.

### *Ökonomie wird durch den Verzicht auf pädagogisch bedeutsame Information erkaufte*

Damit ein Test wenig Durchführungs- und Korrekturzeit erforderlich macht, für die Versuchspersonen leicht handhabbar wird und wenig Material verbraucht, werden verschiedene Maßnahmen getroffen:

überwiegend werden geschlossene Fragen (multiple choice) verwendet (über Vor- und Nachteile einzelner Itemtypen soll in Teil II noch die Rede sein);

die Testhefte werden möglichst eng bedruckt;

das Ankreuzen richtiger Antworten geschieht teilweise auf maschinenlesbaren Bögen;

die Auswertung wird mittels EDV vorgenommen;

wo keine EDV möglich ist, werden Schablonen mitgeliefert, die das Auswerten erleichtern. Für den Schüler haben derlei Ökonomiebestrebungen mitunter negative Konsequenzen, etwa daß sich irrelevante Faktoren im Testergebnis niederschlagen: die Lesbarkeit der Testfragen, das Geschick mit multiple-choice-Aufgaben umzugehen (test wiseness) und die Konzentration bei der Übertragung auf Antwortblätter und Markierungsbögen.

Für den Lehrer stellt sich am Ende das Schülerverhalten in Form von schwarzen Flächen, Kreuzchen oder Ziffern dar. Je abstrakter jedoch die Abbildung der Schülerleistung erfolgt, desto schwerer ist es, daraus Informationen zu entnehmen, die über den Prozeß des Zustandekommens der Antwort Aufschluß geben. Solche Informationen sind für eine Interpretation der Lernleistung, für eventuell notwendig werdende Korrekturen im Lehr- und Lernprozeß aber unabdingbar.

Aus den vorangegangenen testtheoretischen Ausführungen sollte deutlich geworden sein, daß sich aus den Testgütekriterien keineswegs eine so stichhaltige Überlegenheit von Testurteilen gegenüber Lehrerurteilen ableiten läßt, wie dies häufig postuliert wird. Vom pädagogischen Standpunkt aus müssen herkömmliche Testverfahren abgelehnt werden, da sie keinen Beitrag für eine individuell gerechte, förderungsorientierte Diagnostik leisten. Sie sind eher ein Rückschritt gegenüber dem subjektiven Lehrerurteil; denn vom Lehrerurteil kann angenommen werden, daß es weit mehr Daten über individuelle Voraussetzungen und Bedingungen des Schülers für das Lernen in der Schule berücksichtigen kann als dies jegliche Testbatterie tut. Das soll nun nicht heißen, daß die Diagnostik in der Schule nicht verbesserungsbedürftig ist. Der Ansatzpunkt zur Verbesserung sollte jedoch m. E. nicht ein überregional entwickeltes und überprüftes, und landeseinheitlich angewandtes Diagnoseinstrumentarium sein, sondern die kritische Rückbesinnung der Lehrerschaft auf ihre diagnostischen Qualitäten, wozu ihr neuartige, auf die Belange des Unterrichts eingehende Verfahren angeboten werden sollen. Darüberhinaus müßte in der institutionalisierten Lehreraus- und -weiterbildung eine derzeit klaffende Lücke im Bereich der pädagogischen Diagnostik geschlossen werden.

## **5. Einige Neuansätze zur pädagogischen Diagnostik versprechen eine Verbesserung der Schülerbeurteilung: Orientierungsstufe; Diagnosebogen; Item-Bank; lernzielorientierte Tests**

Nachdem sich 1969 mit dem Erscheinen des von *H. Roth* herausgegebenen Sammelbandes „Begabung und Lernen“ ein dynamischer Begabungsbegriff in Deutschland durchsetzte, wonach Begabung nicht mehr nur als statische Größe, sondern auch als Ergebnis von Herausforderungen der Umwelt angesehen wird, wurde die frühzeitige Selektionsentscheidung für weiterführende Schulen, die sich auf eine punktuelle Überprüfung von Lernleistungen am Ende des 4. Schuljahres bezog, vollends fragwürdig. Behutsameren Übergang von der Primar- zur Sekundarstufe verspricht man sich von einer zweijährigen Beobachtungsstufe im 5. und 6. Schuljahr.

### ***Orientierungsstufe***

Auf Anregung des „Deutschen Ausschusses für das Erziehungs- und Bildungswesen“ wurden „Förderstufen“ und „Eingangsstufen“ probeweise in Hessen und Niedersachsen eingeführt. Während die Förderstufe organisatorisch mit der Grund- und Hauptschule verbunden war, sich auf alle Kinder der 5. und 6. Schuljahre erstreckte und durch Kern-

und Kursunterricht eine Leistungsförderung auf verschiedenem Niveau vornahm, war die Eingangsstufe organisatorisch an die Realschule oder das Gymnasium gebunden und sollte dem auf Probe aufgenommenen Schüler zeigen, ob er den schultypischen Anforderungen gewachsen ist. Die Erkenntnisse aus den Schulversuchen wurden in einem Konzept für eine „Orientierungsstufe“ zusammengefaßt. Der Zweck der Orientierungsstufe ist — nach den Vorstellungen des Strukturplans des Deutschen Bildungsrates — die Beobachtung und Ausbildung der individuellen Begabungsrichtungen und Leistungsstärken bis zur eigentlichen Entscheidung über die weiterführende Schullaufbahn für die 5./6. Schülerjahrgänge; „bei größtmöglichem Ausbalancieren von schichtspezifischen Barrieren sollen durch differenzierenden Unterricht weder die Leistungsspitzen vernachlässigt noch die Leistungsschwachen überfordert werden, um Fehlentwicklungen zu verhindern“ (Keck, 1973, 114 f.). Ähnliches wird auch in den Zielsetzungen für die Orientierungsstufe in Bayern formuliert (*Schulreform in Bayern*, 1970). Die Schule hat den offiziellen Auftrag, während dieser Zeit förderungsorientiert zu arbeiten und frühzeitige Entscheidungen über künftige Schullaufbahnen zu vermeiden. In der Praxis sieht es jedoch anders aus. Die bayerische Lösung bevorzugt eine an den Schultyp gebundene Orientierungsstufe, d. h. daß bereits nach dem 4. Schuljahr eine Vorselektion getroffen wird, ob ein Schüler die Orientierungsstufe am Gymnasium oder in der Hauptschule besuchen darf. Damit ist allerdings die Chance für eine sozial gerechtere Laufbahnenlenkung weitgehend vertan. Lehrplanangleichungen und gelegentlicher Lehreraustausch haben — zugestanden — zur Erweiterung des Problembewußtseins aller Beteiligten beigetragen. Anzuerkennen ist auch der nur mit großem Idealismus zu bewältigende Mehraufwand an Zeit der Lehrkräfte an den Versuchsschulen. Der Effekt der Anstrengungen verleitet jedoch eher zur Resignation denn zur Hoffnung auf verbesserte Begabungsfindung und -förderung: Die Erfahrungen mit den Leistungskursen in den Kernfächern haben gezeigt, daß die Durchlässigkeit nahezu ausschließlich in eine Richtung geht: nach unten. Die Gründe für das Scheitern dieser Reformidee in Bayern sind sicher vielzählig. Neben den auferlegten institutionellen Einschränkungen wird wohl als Hauptgrund das Festhalten an einem überkommenen Leistungsprinzip zu sehen sein, das in einer demokratisch sich verstehenden Schule als nicht mehr angemessen erscheint (vgl. hierzu: Moser, 1972 und Klafki, 1974). Halten wir fest: Die Orientierungsstufe könnte entsprechend ihrer Konzeption den institutionellen Rahmen für eine pädagogisch vertretbare Schülerbeurteilung bieten, vermag jedoch in der praktizierten Form dieser Aufgabe noch nicht gerecht zu werden.

### *Diagnosebogen*

Um von der diagnostisch nichtssagenden Ziffernbenotung wegzukommen (zur Kritik am normbezogenen Zensierungsmodell s. a. Schreiner, 1970 a und Weis, 1971), werden derzeit unterrichtsbegleitende Diagnosebögen erprobt. Darauf werden verfügbare Daten über den Schüler gesammelt, die Aufschluß über seine Lerngeschichte, seine soziale Herkunft, sein Gruppenverhalten, seine Stellung in der Klasse, das Erreichen/Nichterreichen fachbezogener Unterrichtsziele, die Höhe der kognitiven Verarbeitung des Unterrichtsstoffes geben. Dem Lehrer wird damit ein Hilfsmittel an die Hand gegeben, das ihm erlaubt, Lernprozesse seiner Schüler zu verfolgen, Lernbedingungen einzelner Schüler zu bedenken und Hypothesen über etwaige Fördermaßnahmen anzustellen. Nach dem Hessischen Konzept des Diagnosebogens (s. betrifft: erziehung, Heft 7/1973) soll damit das übliche Zeugnis ersetzt werden. Die Erfahrungen in Hessen haben gelehrt, daß noch gewisse Mängel zu beseitigen sind, die auch den bislang verfaßten Gutachten in den Schüler-

bögen anhaften: Manche zur Beschreibung verwendeten Begriffe sind entweder so abstrakt (etwa „transformatorische“, „analytische“ oder „imitatorische“ Denkfähigkeit), daß ihr Informationsgehalt — aufgrund unterschiedlicher Interpretation bei Verfasser und Abnehmer des Gutachtens — gering ist, oder so allgemein (etwa „ein lebhaftes Kind“), daß die Gefahr der Umdeutbarkeit in Richtung auf eine bestimmte Vorannahme hinsichtlich einer Schülerpersönlichkeit besteht. Einen Diagnosebogen, der das Schülerverhalten in konkreten Unterrichtssituationen zu klassifizieren und zu identifizieren sucht und dabei gleichzeitig Daten über die Qualität des Unterrichts erbringt, hat die *Projektgruppe Schullaufbahnberatung* (1973) entworfen. Zur Praktikierbarkeit dieses Konzepts konnten leider keine Untersuchungen mehr angestellt werden. Es darf angenommen werden, daß ein sorgfältig geplanter, nicht zu stark schematisierter Beobachtungsbogen ein wertvolles Mittel der pädagogischen Diagnostik sein kann.

### *Itembank*

Als weiterer Versuch, dem Lehrer bei seiner diagnostischen Tätigkeit zu helfen, kann das Projekt „Itembank“ genannt werden, das vom KM des Landes Schleswig-Holstein finanziert wird. In der Itembank werden Unterrichtsvorbereitungen gesammelt und zwar getrennt nach dem zugrundeliegenden Basaltext (wie bei Programmierter Unterweisung), den enthaltenen Lernzielen und den dazugehörigen Testaufgaben. Nach Anfrage der Lehrer können gewünschte Unterrichtsvorbereitungen abgerufen werden. Hiermit ist also zunächst einmal eine Arbeitserleichterung für den Lehrer geschaffen. Sodann kann der Unterrichtende aufgrund seines speziellen Lernzielkatalogs die dazugehörigen Testaufgaben anfordern, die ihm vom Computer zusammengestellt, anschließend vervielfältigt und nach Rücksendung auch ausgewertet werden. Sollte dieses System — das bis 1976 projektiert ist — einmal reibungslos funktionieren, so könnte für jede Lerngruppe ein spezifisches, lernzielbezogenes Instrument zusammengestellt werden (vgl. *Royl*, 1972). Dieses Projekt mutet wie Utopia an, und es ist sicher, daß es noch geraume Zeit dauern wird, bis alle damit anfallenden Probleme bewältigt sind. Als Variante der Itembank kann eine Errungenschaft der amerikanischen Erziehungswissenschaft gelten: In Lehrerzeitschriften bieten diverse Forschungszentren ihre Lernzielformulierungen und erprobten Testaufgaben in Buchform für wenig Geld an (über 5000 Items um 3,50 Dollar; Fundort: Phi Delta Kappan, Januarheft 1974, S. 369).

Übersehen darf nicht werden, daß die Lernziele einer Itembank u. ä. nahezu ausschließlich kognitives Lernen abdecken können. Als einziges Mittel zur Beurteilung schulischen Lernens kommt sie demnach nicht in Frage.

### *Lernzielorientierte Tests*

In dem Maße, wie die Curriculumdiskussion neue Erkenntnisse für die Lehr- und Lernplanung brachte, gewann ein Prüfverfahren an Bedeutung, das Aufschluß über das Lernverhalten der Schüler im curricularen Kontext und über die Qualität des Curriculums gibt: der lernzielorientierte Test (LOT), auch kriteriumsorientierter Test genannt. Seine Aufgaben beziehen sich auf das durch Unterricht beim Schüler angestrebte Verhalten. Ein LOT gibt Auskunft, ob ein Schüler ein bestimmtes Lernziel erreicht hat oder nicht, beschränkt sich also auf qualitative Aussagen (während herkömmliche Tests quantitativ beurteilen). Mit seiner Hilfe wird der gezielte Einsatz individueller Lernhilfen vorbereitet. Er ist ein Instrument der „formativen“ Beurteilung. Bislang vorliegende Informationen über Definition, Prinzipien, Konstruktion und Einsatzmög-

lichkeiten LOTs vermittelt der Sammelband von *Strittmatter* (1973). Beiträge mehr theoretischer Natur (z. B. Meßprobleme) enthalten die Arbeiten von *Heipcke* (1972) und *Klawer u. a.* (1972). Einige Berichte über praktische Erfahrungen mit der Entwicklung und Erprobung von LOTs in verschiedenen Schulfächern liegen bereits vor (z. B. *Hagener*, 1972 — für den politischen Unterricht; *Henning u. a.*, 1972 — für Sozialkunde; *Mauermann, Schulte & Seisenberger*, 1973 und *Projektgruppe Schullaufbahnberatung*, 1973 — für Mathematik, Physik und Deutsch; *Borries*, 1937 — für Geschichte). Einen speziellen Service für lernzielorientierte Tests bietet mittlerweile der Klett-Verlag Stuttgart an: Zu dem im Verlag erscheinenden Englisch-Lehrbuch werden LOTs mit multiple-choice-Aufgaben angeboten (für Mathematik und Sprachlehre sind LOTs in Vorbereitung), die sich ganz strikt an die Lektionen des Lehrbuchs anlehnen. Die Schüler kreuzen auf einem maschinenlesbaren Markierungsbogen die richtige Antwort an. Die ausgefüllten Antwortblätter werden an den Verlag geschickt, der eine umgehende Auswertung und Rückmeldung garantiert. Aus der zurückgesandten Ergebnisliste geht hervor, in welchen Bereichen der einzelne Schüler Stärken bzw. Schwächen hat. Einzuwenden ist gegen diese LOTs,

- daß sie lediglich leicht abfragbares Wissen prüfen;
- daß die Items nach herkömmlichen Verfahren ausgewählt wurden (nach Trennschärfe, Schwierigkeitsgrad);
- daß sie eine peinlich genaue Einhaltung der Buchlektionen erfordern;
- daß sie kaum Hinweise für den Ursachenzusammenhang beim Entstehen von Lernaussfällen geben können;
- daß durch die Vielzahl der Stationen bei der Auswertung (Lehrer, Schüler, Kodierungsstelle, Einlesestation, Rechner, Schnelldrucker, Versandstelle) die Fehleranfälligkeit größer wird (s. *Küffner*, 1973).

Ist ein Lehrer von der Notwendigkeit lernzielorientierter Verfahren für die Verbesserung von Lernen und Lehren überzeugt, und will er sich in seiner Unterrichtsplanung durch vorhandene LOTs nicht gängeln lassen, so bleibt ihm die Möglichkeit, allein oder in Kooperation mit Kollegen einen eigenen LOT zu erstellen. Praktische Hinweise hierfür sollen im folgenden Teil II gegeben werden.

## Teil II: Schulleistungsmessung durch lernzielorientierte Tests (LOTs)

**1. Ein LOT, den der Lehrer aufgrund der in seinem Unterricht angestrebten Lernziele selbst oder in Teamarbeit mit Kollegen erstellt, vermindert den Mangel an Unterrichtsbezogenheit, den herkömmliche Testverfahren aufweisen**

Wie in der vorausgegangenen Kritik der herkömmlichen Diagnostik bereits ausführlich dargestellt wurde, liegt einer der Hauptmängel dieser Tests (seien es Intelligenztests oder normierte Schulleistungstests) in ihrer unstatthafter Trennung von Unterricht und Schülerbeurteilung. Diese Trennung zu überwinden vermag nur der Lehrer, der weiß,

- wie die eine oder andere Unterrichtseinheit abgelaufen ist,
- wie die Schüler sich an diesem Unterricht beteiligt haben,
- welche Störungen durch Unterrichtsausfälle, Erkrankungen u. ä. während des Lernprozesses auftraten,
- welche Unterrichtsmittel eingesetzt oder nicht verwendet werden konnten,
- welche Lerngeschichte der einzelnen Schüler aufzuweisen hat,
- welchem familiären Milieu der Schüler entstammt.

Vor diesem Hintergrund kann er das beobachtete Lernverhalten in Beziehung zu Lernvoraussetzungen setzen und Erklärungen dafür suchen und er könnte, so es die organisatorischen Gegebenheiten zuließen, differenzierende pädagogische Maßnahmen treffen.

Die Beobachtung des Schülers während des Unterrichts allein genügt nicht für eine umfassende Skizzierung des Lernverhaltens. (Es ginge unter Umständen, wenn der Lehrer eine kleine Gruppe vor sich hätte anstelle einer Klasse mit 30—50 Schülern.)

Eine Methode, wie der Lehrer einige Bereiche des Lernverhaltens — überwiegend sind dabei kognitive Lernprozesse betroffen — überprüfen kann, ist der LOT, den der Lehrer aufgrund der im Unterricht angestrebten Lernziele selbst konstruiert und in seiner Klasse einsetzt. Mit diesem Instrument wird der Mangel an Unterrichtsbezogenheit, wie er den herkömmlichen Testverfahren vorgeworfen wird, reduziert. Es sei an dieser Stelle bereits darauf hingewiesen, daß es sich bei LOTs nur um *eine* mögliche und keineswegs hinreichende Methode zur Schülerdiagnose handelt.

## **2. Grundlage für die Konstruktion eines LOT ist die Struktur der Lerninhalte, die nach sachlogischen und lernpsychologischen Ordnungsprinzipien zu erstellen ist**

Bedingung für eine erfolgsversprechende Konstruktion und Anwendung eines LOT ist eine lernzielorientierte Unterrichtsplanung. Sie verlangt eine intensive vorbereitende Beschäftigung mit den angestrebten Unterrichtszielen. Ein Lehrer, der sich kurz vor Unterrichtsbeginn vornimmt, die Tulpe als Beispiel einer Gartenblume durchzunehmen und deshalb aus dem Lehrmittelzimmer das zerlegbare Plastikmodell einer Tulpe mit ins Klassenzimmer schleppt, mag vielleicht einen recht brauchbaren Biologieunterricht halten. Er wird es jedoch — nach eventuell festgestellten Lerndefiziten — sehr schwer haben, den Unterricht so zu rekonstruieren, daß eine Suche nach den Ursachen von Lernschwierigkeiten Erfolg hat. Es lassen sich gewiß noch weitere Gründe für die Notwendigkeit zielorientierten Unterrichts finden. Unser simples Beispiel soll in diesem Zusammenhang genügen. In der Praxis anwendbare Ratschläge und Modelle zielorientierter Unterrichtsplanung geben u. a. *Peterßen (1972)*, *Kunert (1972)* und *Peterßen (1974)*.

Erste Überlegungen des Lehrers für die Unterrichtsplanung gelten meistens, wenn nicht nahezu ausschließlich, dem zu übermittelnden Lehr-/Lernstoff. Daß hier ein Umdenken bezüglich der Prioritäten von Zielüberlegungen im Rahmen einer pädagogischen Diagnostik notwendig ist, wird verschiedentlich betont (z. B. *Schreiner, 1972*; *Klafki, 1974*; *Projektgruppe Schullaufbahnberatung, 1973*). Häufig stellt sich die Unterrichtsplanung so dar, daß der Lehrplan das Thema liefert, das Schülerbuch die Lerninhalte und das Lehrerhandbuch die methodisch-didaktischen Ratschläge beisteuern. Oft wird jedoch nicht einsichtig, weshalb die eine oder andere Vorgehensweise einzuschlagen ist. Abhilfe von fehlender Transparenz schafft eine *Strukturierung der Lerninhalte*. Hierbei handelt es sich um eine Anordnung von Lernstoffen mit hierarchischer Beziehung: Das Lernen eines vorstehenden Lerninhalts ist Voraussetzung für das erfolgreiche Aneignen des in der Sequenz nachfolgenden Inhalts. Ein Beispiel für eine Inhaltsstruktur zeigt Abbildung 1. Weitere Beispiele finden sich bei *Gagné (1969)*, *Bloom, Hastings & Madaus (1971)* und *Mauermann, Schulte & Seisenberger (1973)*.

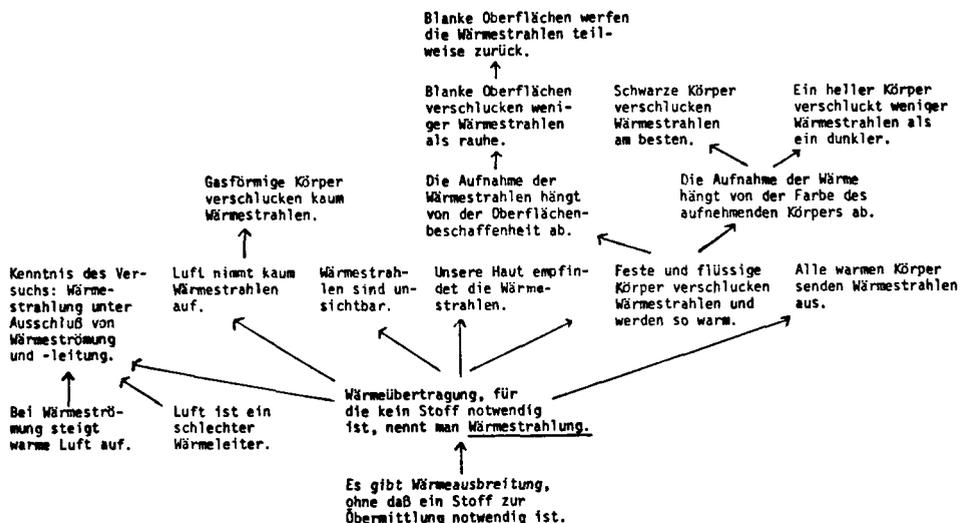


Abb. 1: Beispiel einer Struktur der Lerninhalte für den propädeutischen Physikunterricht „Wärmestrahlung“ der 5. Jahrgangsstufe (Quelle: Modellcurriculum „Wärmeausbreitung“ der Projektgruppe Schullaufbahnberatung. München, 1972, unveröffentl. Manuskript).

Aus dem abgebildeten Beispiel wird ersichtlich, daß die Anordnung aufgrund sachlogischer Überlegungen erfolgt ist. Als anderes Ordnungskriterium neben der Sachlogik können lernpsychologische Überlegungen dienen (besonders bei der Grobstrukturierung): Motivationale oder soziale Gegebenheiten legen oft bestimmte sequentielle Abfolgen nahe. (So wird ein für die Schüler interessant gestalteter Unterricht wohl kaum in der Reihenfolge, wie sie durch die Pfeilrichtung in Abbildung 1 angedeutet ist, die einzelnen Lerninhalte vermitteln. Am Anfang der Unterrichtseinheit kann z. B. die gemeinsam getroffene Beobachtung stehen, daß im Sommer überwiegend hellere Kleidung getragen wird als im Winter. Die anschließenden Überlegungen können die Farbe als Ursache verschiedener Wärmeaufnahme herausarbeiten und die Sonnenstrahlen als Ursache der Erwärmung entdecken. Darauf ließe sich das Phänomen der Wärmestrahlung von denjenigen der Wärmeleitung und Wärmeströmung abgrenzen. Dieses Vorgehen entspräche dann jedoch nicht mehr der sachlogischen Abfolge, sondern wäre auf die Bedürfnisse der Schüler und deren Erlebniswelt abgestimmt.)

**3. Unmittelbarer Ausgangspunkt für die Aufgabenerstellung zum LOT ist die Lernzielformulierung. Die materiale (stoffliche) Komponente wird ihr über die Lerninhaltsstruktur, die formale (Verhaltens-) Komponente mit Hilfe von Lernziel-Taxonomien zugewiesen**

Die Gliederung des Lernstoffes allein ist noch keine ausreichende Basis für die Erstellung eines LOT. Der Lehrer hat zwar dadurch komplette Informationen über die Lernmaterie, es fehlen aber Angaben darüber, auf welchem Niveau der geistigen Verarbeitung der Schüler deren Aneignung unter Beweis stellen soll: Genügt das bloße Wiedererinnern eines bestimmten Wissens oder sollen die im Unterricht erworbenen Kenntnisse auf ein

Problem des Alltags angewendet werden? D. h. also, daß nun zusätzlich der „formale Bildungsgehalt“ des Lernstoffes reflektiert werden muß.

Der amerikanische Psychologe und Erziehungswissenschaftler *B. Bloom* hat zusammen mit seinem Mitarbeiterstab das zur Zeit wohl bekannteste Klassifikationsschema für formale Lernziele geschaffen. *Bloom* u. a. gliedern das menschliche Verhalten in drei Bereiche: kognitives, sozial-emotionales und psycho-motorisches Verhalten. Für die ersten beiden Bereiche liegen ausgearbeitete Taxonomien — wie diese Klassifikationsschemata auch genannt werden — vor (*Bloom*, 1956 — deutsch 1972; *Krathwohl* u. a., 1964). Was unter „Niveau der geistigen Verarbeitung“ zu verstehen ist, sei an ein paar Beispielen veranschaulicht:

1. *Kenntnisse*: Der Schüler soll angeben können, daß Luft etwa 1/5 Sauerstoff und etwa 4/5 Stickstoff enthält.
  2. *Verständnis*: Der Schüler soll die chemische Gleichung  $C + O_2 \rightarrow CO_2$  in die verbale Form übertragen können: Kohlenstoff und Sauerstoff verbinden sich zu Kohlendioxyd.
  3. *Anwendung*: Der Schüler soll die Kenntnis über die Notwendigkeit von Sauerstoff in Verbrennung von Wachs für die Erklärung des Verlöschens einer brennenden Kerze in einem geschlossenen Behälter heranziehen können.
  4. *Analyse*: Der Schüler soll bei einer vom Lehrer aufgebauten Versuchsanordnung zum Nachweis des Sauerstoffgehalts der Luft die wesentlichen Elemente (Wasserbehälter, Kerze, Wasserspiegel, Wassersäule, Reagenzglas usw.) unterscheiden und ihre Funktion im Experiment benennen können.
  5. *Synthese*: Der Schüler soll mit Hilfe der an der Schule vorhandenen physikalischen Experimentiergeräte einen Versuch zum Nachweis des Sauerstoffgehaltes eines unbekanntes Gasgemisches aufbauen.
  6. *Beurteilung*: Der Schüler soll aus zwei Versuchsaufbauten zum Sauerstoffnachweis bei Gasgemischen den zweckmäßigeren (nach Kriterien wie Zeit- und Materialaufwand, Gefährlichkeit) benennen.
- (Vgl. auch *Mauermann, Schulte & Seisenberger*, 1973, 103—107)

Aus den Beispielen ist ersichtlich, daß eine materiale und eine formale Komponente (eine Inhalts- und eine Verhaltensdimension) miteinander zu verknüpfen sind, um schließlich eine präzise Formulierung des Lernziels zu erhalten. Solche Lernzielformulierungen sind die eigentliche Grundlage des zielorientierten Unterrichts und auch des LOT.

#### **4. Die Beachtung der in der Taxonomie höher stehenden (komplexeren) Ziele wirkt dem in den herkömmlichen Leistungstests übergewichteten Kenntnisabfragen entgegen**

Die Anwendung der Taxonomie formaler Lernziele bei der Bestimmung von Unterrichtszielen erweist sich deshalb als wertvoll, weil sie den Lehrer auf kognitive Prozesse aufmerksam macht, die im gegenwärtigen Unterricht unterrepräsentiert sind. Daß dieser Mangel mit Hilfe der Taxonomie reduziert werden kann, wird deutlich, wenn man Curricula, Lernzielkataloge oder LOTs analysiert, bei deren Entwicklung die Strategie der Taxonomisierung angewendet wurde (z. B. *Kunert*, 1974; *Borries*, 1973; *Hagener*, 1972), wengleich oft eine recht leichtfertige Zuordnung der Verhaltensdimensionen getroffen wird (etwa bei *Henning* u. a., 1972).

Das Festlegen von Lernzielen auf höheren Ebenen der Hierarchie erweist sich als relativ schwierig, zumal die *Bloomschen* Kategorien nicht eindeutig voneinander zu trennen sind. Besondere Schwierigkeiten machen die Kategorien „Verständnis“ und „Analyse“, die nach der vorliegenden Beschreibung kaum strikt abzugrenzen sind.

*Bloom* u. a. sind zwar nicht die einzigen, die sich mit der Kategorisierung von formalen

Lernzielen befaßten, sie haben aber die bisher umfangreichste und detaillierteste Taxonomie vorgelegt. Wem die *Bloomsche* Taxonomie zu kompliziert erscheint, kann sich auch mit schlichteren Klassifizierungen bescheiden: *Schulz* (1971) schlägt drei Kategorien vor, die er unter dem Aspekt der „Intentionalität“ des Unterrichts zusammenfaßt und die bei ihm „Kenntniserwerb“, „Reproduktion von Erkenntnissen“ und „Produktives Denken“ heißen. Im „Strukturplan für das Bildungswesen“ des Deutschen Bildungsrats werden vier formale Lernzielstufen unterschieden: Wiedergabe von Gelerntem, Reorganisation des Gelernten, Transfer des Gelernten auf neue Situationen und problemlösendes Denken/entdeckende Denkverfahren. Hier, wie auch zu der Einteilung von *Schulz*, ließen sich entsprechende Kategorien bei *Bloom* finden.

Gehen wir davon aus, daß die Stufen 3 bis 6 bei *Bloom* ein Transferieren der erworbenen Kenntnisse beinhalten und die Stufen 1 und 2 sehr eng beieinanderliegen, wäre auch eine grobe Zwei-Stufen-Klassifizierung denkbar: Kenntnisswissen und problemlösendes Verhalten (zur Information über weitere Taxonomien s. *Meyer*, 1974).

### **5. Zur Erschließung der lerninhaltlichen Komponenten komplexerer Lernziele (Testaufgaben) empfiehlt sich die Methode der Aufgabenanalyse (Rückwärtsanalyse) nach Gagné**

Wie bereits erwähnt, ist es relativ leicht, Lernziele auf dem Kenntnisniveau zu überprüfen (etwa: Wie heißt das Spiegelgesetz? Welche Flüsse münden bei Passau in die Donau? Wann geschah der Mord von Sarajewo? Wann wird nach einem Doppelpunkt groß geschrieben?). Komplizierter wird das Prüfen, wenn das Lernziel lautet: Der Schüler soll das Wissen um die unterschiedliche Reflexion von Wärmestrahlen beim Auftreffen auf feste Körper unterschiedlicher Farbe zur Erklärung von beobachtbaren Gegebenheiten aus der Natur und Technik seiner näheren Erlebniswelt heranziehen können.

Hier soll die Kenntnis auf einen konkreten Sachfall transferiert werden können. Dem Lehrer stellt sich zunächst die Aufgabe, denjenigen Sachverhalt auszuwählen, der geeignet ist, beim Schüler ein Problem-Lösungsverhalten hervorzurufen. Da die Problemstellung den Schüler mit einer für ihn *neuen* Situation konfrontieren soll, muß der Lehrer dafür sorgen, daß das für die Testaufgabe vorgesehene Problem während des Unterrichts nicht bereits angesprochen oder behandelt wird (ansonsten wäre es für den Schüler kein Problemlösen mehr, sondern lediglich ein Rückerkennen an einen bekannten Lösungsweg). Das konkrete Lernziel könnte in unserem Fall dann z. B. heißen:

— Der Schüler soll Vorschläge für den Umbau eines normalen Eisenbahngüterwaggons zum Transport von tiefgefrorenen Nahrungsmitteln machen und begründen können.

Die Testaufgabe ließe sich dann etwa so formulieren (Quelle: Modellcurriculum Wärmeausbreitung, a.a.O.):

— Die Bundesbahn befördert auch im Sommer tiefgefrorene Seefische von Hamburg nach München. Wie müßte ein normaler Güterwagen verändert werden, wenn er für diesen Zweck geeignet sein soll?

Gesetzt den Fall, der Lehrer erhebt die richtige Beantwortung dieser Frage zum Kriterium für den erfolgreichen Abschluß seiner Curriculumeinheit „Wärmeausbreitung“. Was kann er tun, wenn — wie zu erwarten — ein Teil seiner Schüler dieses Item nicht ausreichend lösen kann? Er hat nicht ohne weiteres eine Erklärung für das Nicht-Lösen zur Hand.

Um begründen zu können, weshalb ein Schüler versagte, bedarf er weiterer Informationen über das Lernverhalten. *Gagné* (1971) schlägt einen recht einsichtigen Weg zur Er-

schließung der lerninhaltlichen Komponenten von komplexen Lernzielen bzw. Testaufgaben vor: Er rät, vom gewünschten Endverhalten rückwärts zu fragen. Welche Lernziele muß der Schüler vorher durchlaufen haben, damit er das gewünschte Endziel erreichen kann? Welche untergeordneten Ziele sind Bedingung für das Meistern des nächsthöheren, darauf bezogenen Inhalts der Lernsequenz? Auf diese Weise lassen sich schrittweise die vorgeordneten Lernziele ermitteln, bis schließlich die relativ einfachsten Kenntnisse und Fertigkeiten feststehen.

An unserem Beispiel vom umzubauenden Güterwaggon würden die vorgeordneten Kenntnisse heißen:

**1. Ebene:** — Die Aufnahme der Wärme durch Strahlung hängt von der Farbe des aufnehmenden Körpers ab. Ein heller Körper verschluckt weniger Wärmestrahlen als ein dunkler.

— Es gibt gute und schlechte Wärmeleiter. Luft und lufthaltige Körper sind schlechte Wärmeleiter.

**2. Ebene:** — Feste und flüssige Körper verschlucken Wärmestrahlen und werden so warm. — Erwärmte Stellen in einem Körper können die Wärme an benachbarte kältere Stellen weitergeben.

Nach dieser Rückwärtsanalyse (*Gagné* nennt sie ‚task analysis‘) gilt es also, für die untergeordneten Ziele auch Testaufgaben zu konstruieren, damit erforscht werden kann, ob das Nicht-Lösen durch die fehlenden Kenntnisse über Gesetze der Wärmeausbreitung bedingt ist. Sicher werden die Informationen bezüglich der im Curriculum angestrebten Lerninhalte nicht ausreichen, um jedesmal den Lösungsweg exakt rekonstruieren zu können. Außercurriculares Wissen wird der Schüler zusätzlich zur Lösung einbringen müssen (vgl. *Mauermann*, 1974):

— Ein normaler Güterwaggon ist dunkelbraun angestrichen.

— Die Wände eines normalen Güterwaggons sind aus einfachen Brettern gefertigt.

— Wenn tiefgefrorene Fische auftauen, dann verderben sie.

Sollte sich bei der Überprüfung herausstellen, daß der Schüler wegen solcher außercurricularer Kenntnisse diese Aufgabe nicht lösen konnte, darf sie nicht als Kriterium zur Beurteilung des Schülerverhaltens in bezug auf das Curriculum verwendet werden. Die Aufgabe ist dann als nicht aus der näheren Erlebniswelt stammend anzusehen und berührt damit nicht das gesetzte Lernziel (s. o.).

Erleichtert wird diese rückwärtsfragende Aufgabenanalyse dann, wenn für das gesamte Curriculum eine Sachstruktur vorliegt.

## **6. Bei der Formulierung von LOT-Aufgaben sind die Vor- und Nachteile der unterschiedlichen Itemtypen (gebundene und freie Antworten) zu beachten**

Die folgende Darstellung beschränkt sich auf Papier-Bleistift-Verfahren, d. h. auf Testformen, die vom Schüler eine schriftliche Beantwortung verlangen. Das hat den schlicht pragmatischen Grund, daß für mündlich-akustische oder manuell-experimentelle Verfahren oft die technischen Voraussetzungen in unseren Schulen fehlen (etwa für Tonbandmitschnitte oder Schülerwerkplätze). Herkömmlicherweise unterscheidet man in der Testtheorie zwei Klassen von Testaufgaben:

a) Aufgaben mit freier Beantwortung

b) Aufgaben mit gebundener Beantwortung (multiple choice).

### *(Zu a)*

Aufgaben mit freier Beantwortung findet man wohl im überwiegenden Maße, nimmt man die Schulaufgaben oder Probearbeiten unter die Lupe, wie sie gegenwärtig an unseren Schulen durchgeführt werden. Die Freiheitsgrade bei der Beantwortung können schwanken zwischen solchen Aufgaben, die nur das Füllen einer Lücke (Lückentext) oder ein einziges Wort der Beantwortung verlangen (z. B. In welchem Jahr entdeckte Kolumbus Amerika?), und solchen, die eine komplexe Beantwortung erfordern, sog. Kurzaufsatzaufgaben (z. B. Welche Umstände führten zum Ausbruch des 1. Weltkrieges? Was hältst du von der Kriegsschuldfrage?). Gerade die letztgenannten Kurzaufsatzprüfungen stehen häufig im Kreuzfeuer der Kritik seitens der Schülerschaft und der Öffentlichkeit. Solchen Prüfungen wird vorgeworfen, sie verleiteten den Lehrer zur Einstufung der Schüler ohne objektive Bewertungsmaßstäbe. Faktoren wie Lesbarkeit der Handschrift, Rechtschreibung und Grammatik, Ausdrucksfähigkeit und andere nicht relevante Faktoren (s. Teil I) beeinflussen unbewußt die Beurteilung. Ein sprachlich geschickter Schüler vermag z. B. zu bluffen, indem er die gestellte Frage durch eine ähnliche ersetzt, die er beantworten kann, oder indem er die inhaltliche Dürftigkeit seiner Antwort durch glänzende Form überspielt.

Die Vorteile solcher Aufgaben dürfen nicht übersehen werden:

- Vom Schüler wird eine Zusammenschau der Lernstoffe verlangt (was der Entwicklung einer sinnvollen Lernmethode nur förderlich sein kann).
- Der Lehrer kann aus den Antworten erschließen, inwieweit der Schüler
  - angemessene Lernstoffe auszuwählen vermag,
  - seine Erkenntnisse zu ordnen in der Lage ist,
  - seine Meinung und Haltung artikulieren kann,
  - selbständig zu arbeiten vermag.

### *(Zu b)*

Aufgaben mit gebundener Antwortform dominieren in den sog. „objektiven Tests“. Items mit gebundener Antwortform bestehen allgemein aus einer Vorgabe (stem, Stamm) und vorformulierten Antwortmöglichkeiten, unter denen der Schüler die richtige herauszusuchen hat. Von hier häufig verwendeten Testaufgabentypen sind im Anhang Beispiele wiedergegeben:

#### *1) Richtig-Falsch-Aufgabe*

Eine Aussage soll beurteilt werden, ob sie richtig oder falsch ist. Die Vorgabe muß klar und eindeutig formuliert sein, damit eine eindeutige Stimmt/Stimmt-nicht-Entscheidung gefällt werden kann. Gegen diesen Item-Typ wird eingewendet, daß die Ratequote 50% beträgt. Für diagnostische Zwecke ist dieser Typ wenig geeignet, da er keine Hinweise für das Zustandekommen der Antwort beim Schüler gibt.

#### *2) Einfache Mehrfachwahlantwort-Aufgabe*

Der Stamm ist entweder eine Frage oder eine unvollständige Aussage, die Auswahlantworten sind entweder Antworten auf die Frage oder Ergänzungen der unvollständigen Aussage.

#### *3) Komplexe Wahlantwort-Aufgabe*

Die Vorgabe enthält mehrere Aufgaben (Aussagen oder Sätze) und die Aufforderung, diese auf ihre Richtigkeit zu überprüfen. Es werden also mehrere Richtig-Falsch-Entscheidungen verlangt. Die Kombinationen beider Wahlantworten können Aufschluß darüber geben, mit welcher der zu überprüfenden Aussage der Schüler Schwierigkeiten hatte.

#### 4) Zuordnungsaufgabe

Sie besteht aus zwei Reihen von Namen, Begriffen, Aussagen, graphischen Darstellungen o. ä. Die Begriffe, Aussagen usw. der ersten Gruppe sind den am besten zu ihnen passenden der zweiten zuzuordnen.

Gewiß haben Tests mit gebundenen Antwortformen Vorteile:

- subjektive Maßstäbe bei der Korrektur sind ausgeschlossen,
- sachfremde Faktoren können die Schülerleistung nicht beeinflussen,
- die Auswertungszeit ist relativ kurz,
- die Auswertung kann unter Umständen Maschinen überlassen werden.

Die Nachteile sollen jedoch nicht verschwiegen werden:

- Es besteht die Gefahr, daß sich der Schüler auf diese Prüfungen einstellt und einen oberflächlichen Lernstil entwickelt.
- Intelligente Schüler könnten ungerecht hohe Werte erzielen, weil sie aus der Fragestellung (bei schlechtformulierten Testfragen) Anhaltspunkte für die Beantwortung entnehmen können.
- Schwerwiegendster Nachteil ist wohl der, daß die Aktivität (Kreativität) des Schülers eingengt wird und gerade Ziele auf höherem kognitiven Niveau — im Sinne der *Bloomschen Taxonomie* — nur sehr schlecht mit Wahlantwortaufgaben überprüft werden können.

Sinnvoll scheint demnach für einen LOT eine Kombination von multiple-choice- und freien Antwortformen zu sein.

Wenn an dieser Stelle ein Hinweis auf Literatur zur Entwicklung informeller Tests gegeben wird, so deshalb, weil darin u. a. praktische Vorschläge für Testaufgabenkonstruktion enthalten sind. Für kriteriumsorientierte Tests sind die darin enthaltenen Konzepte der Trennschärfe, des mittleren Schwierigkeitsgrades und der Distraktorenkonstruktion (als Angebot zum „Hereinlegen“ von nicht ganz sicheren Prüflingen) abzulehnen. Bücher zur objektivierten Leistungsmessung haben z. B. vorgelegt: *Chauncey & Dobbin* (1968); *Ebel* (1965); *Wendeler* (1970); *Gaude & Teschner* (1970) — zur Kritik an *Gaude & Teschner* vgl. *Schreiner*, (1970b).

#### 7. Ein LOT, der als Instrument der pädagogischen Diagnostik gelten will, hat Aufgaben zu enthalten, die so konstruiert sind, daß sie eine Interpretation spezifischer Lernschwächen zulassen

Wie schon mehrmals betont worden ist, genügen dem Lehrer Informationen über das Erreichen bzw. Nichterreichen eines Lernziels durch den Schüler nicht, um pädagogisch sinnvolle Maßnahmen für den nachfolgenden Unterricht treffen zu können. Er muß auch die Ursachen für individuelle oder gruppenspezifische Lernschwächen kennen. Dazu ist notwendig, andere Konstruktionsprinzipien anzuwenden, als dies bei den herkömmlichen objektiven Lehrertests geschieht. Nicht die bestmögliche Trennschärfe zwischen guten und schlechten Schülern darf als Maßstab für die Güte eines Items zählen, sondern seine Interpretierbarkeit in Hinblick auf Lernverhalten.

Ein Weg, komplexes Lernverhalten in seine Bestandteile zu zerlegen, wurde bereits erwähnt. Die durch Rückwärtsanalyse ermittelten lerninhaltlichen Komponenten können mittels Items abgefragt werden. Dadurch lassen sich eventuelle Bruchstellen in der Lernsequenz diagnostizieren. Es mag z. B. für einen Schüler zutreffen, daß er zwar alle Aufgaben auf der Ebene der Kenntnisse lösen kann, aber immer dann versagt, wenn es an das Transferieren dieses Wissens auf neue Problemsituationen geht. (Diesen Fall hat wohl

jeder Mathematiklehrer beobachten können: Einige Schüler glänzen mit ihren Fertigkeiten bei mechanisch lösbaren Rechenaufgaben, kämpfen jedoch mit größten Schwierigkeiten beim Lösen von sog. Textaufgaben.)

Aufgaben, deren Lösung Aufschluß über ihr Zustandekommen geben, wurden bislang noch nicht systematisch in LOTs einbezogen. Erste zögernde Schritte in diese Richtung hat die Projektgruppe Schullaufbahnberatung in München getan (s. *Mauermann, Schulte & Seisenberger, 1973; Becker-Freyseng, Krüger & Rietbrock, 1973*).

Bei der Konstruktion der LOTs der Projektgruppe wurden zwei Prinzipien zu verwirklichen gesucht:

a) Rein verbal abgefaßte Items sollten sich mit figural-anschaulichen Aufgaben abwechseln.

Soweit es sich ermöglichen läßt, sollte parallel zu einem verbal abgefaßten Item ein anderes konstruiert werden, das mit einer Zeichnung oder einer symbolischen Darstellung das Problem veranschaulicht. Es ist anzunehmen, daß diejenigen Schüler, die rein verbal gestellte Aufgaben schlechter lösen als die figuralen, auch im Unterricht auf anschauliche Lehr- und Lernmittel angewiesen sind.

b) Wahlantworten sollten so konstruiert sein, daß sie unter Umständen Aufschluß über eine spezielle Lernschwäche geben.

Wurden früher die falschen Wahlantworten (Distraktoren) mit der Absicht formuliert, einen Schüler, der die Lernaufgabe nicht völlig beherrscht, zu verführen oder hereinzulegen, so wird ein Lehrer mit pädagogisch-diagnostischem Leitgedanken versuchen, sich in seine Schüler hineinzusetzen, um deren — u. U. auch falsche — Lösungswege zu antizipieren. Er wird deshalb Distraktoren anbieten, die Ergebnissen irrtümlicher Lösungsprozesse entsprechen. Mit der Konstruktion des Distraktors wird somit eine Hypothese über erwartetes, curricular inadäquates Lernverhalten aufgestellt, wie es das folgende Beispiel einer Physikaufgabe verdeutlicht:

Nachts wird es bei uns dunkel. Warum?

Suche die richtige Erklärung!

- a) Nachts sendet die Sonne keine Strahlen aus.
- b) Wir sind dann auf der Schattenseite der Erdkugel.
- c) Nachts kommt die Finsternis.
- d) Die Strahlen der Sonne breiten sich nach allen Richtungen aus.
- e) Dicke Wolken schieben sich vor die Sonne.

Die bei Tests mit Auswahlantworten übliche Auswertung hätte bei diesem Beispiel nun so auszusehen, daß alle Schüler, die nicht die Aussage (b) ankreuzten, zu denjenigen gezählt werden, die das Lernziel offensichtlich nicht erreicht haben und auch keine Punktwertung erhalten. Damit wird jedoch die Information, die aus dieser Aufgabe herausgelesen werden kann, stark reduziert. Es lassen sich aus der Wahl der Distraktoren nämlich auch Vermutungen über das Lösungsverhalten der Schüler ableiten. Solche Hypothesen könnten z. B. lauten:

- Wer (a) angekreuzt hat, besitzt ein falsches Wissen über die Sonne als Energiequelle.
- Wer (c) gewählt hat, scheint die Aufgabenstellung nicht verstanden zu haben, da der Distraktor mit anderen Worten nur das wiederholt, was im Stamm ausgesagt ist.
- Wer (d) gewählt hat, hat sich wahrscheinlich durch den Bekanntheitsgrad dieses im Unterricht gelernten Merksatzes irritieren lassen.
- Wer (e) angekreuzt hat, befindet sich unter Umständen noch in einem „naiven“ Verhältnis zu seiner Umwelt.

Zugegeben, solche Hypothesen klingen sehr gewagt. Es wäre sicher verfehlt, allein schon aufgrund der Beantwortung eines Items zureichende Aussagen über das Lernverhalten der Schüler zu treffen. Das wäre erst möglich, wenn der Schüler sich in weiteren, ähnlichen Prüfungssituationen analog verhielte. Die potentielle Funktion der Distraktoren dürfte jedoch einsichtig geworden sein: Sie vermögen den Lehrer für die Suche nach den Ursachen von Lernschwierigkeiten zu sensibilisieren und zu trainieren.

## 8. Ein LOT ist ein Instrument zur Lerndiagnose und trägt zur Verbesserung des Lernklimas bei

Inwieweit eine intensive Beschäftigung mit pädagogisch relevanter Testaufgabenkonstruktion und -analyse unter den derzeitigen Bedingungen in Schule und Gesellschaft durchführbar ist, muß dahingestellt bleiben. Sicher käme einer Institution wie der Itembank eine Entlastungsfunktion für den Lehrer zu. Hier hätte auch ein computerunterstützter Unterricht (CUU) sein sinnvollstes Betätigungsfeld. Solange es in der BRD nur einige wenige Versuche in dieser Richtung gibt, muß sich der Lehrer selbst behelfen. Ein Hilfsmittel dazu könnte die übersichtliche Protokollierung der Testergebnisse sein. So eine „Evaluationsstruktur“ wurde von der „Arbeitsgruppe Curriculare Tests“ in Verbindung mit einem Modellcurriculum für Physik entwickelt (s. Abbildung 2).

Die in der Evaluationsstruktur eingetragenen Zahlen (vgl. Abbildung 2) bezeichnen die Aufgabennummern des Tests, paarweise angeordnete Items sind parallel. Die oberhalb der gestrichelten Linie befindlichen Kästchen repräsentieren Teilantworten bei den in freier Beantwortung zu lösenden Problemaufgaben. Die Pfeile signalisieren vermutete Bedingungsbeziehungen zwischen den Aufgaben. Für jeden Schüler wird durch ein „+“ oder „-“ in das betreffende Kästchen die Beantwortung eingetragen. So kann nach der Korrektur individuell festgestellt werden,

- in welchen Stoffbereichen Stärken bzw. Schwächen vorliegen,
- bis zu welchem Niveau der kognitiven Verarbeitung der Schüler gelangt ist und
- wo „Bruchstellen“ im Lernprozeß sind.

Mit dieser Evaluationsstruktur wird gewährleistet, daß sich die Auswertung des LOT nicht auf bloße Richtig/Falsch-Aussagen und deren Addition beschränkt. In Verbindung mit intensiver Itemanalyse und Vergewärtigung des stattgefundenen Unterrichts werden zusätzliche Informationen zur Erklärung und Interpretation des Lernverhaltens gewonnen. So kann der LOT zu einem curricularen Test werden und Hinweise zur Gestaltung des nachfolgenden Unterrichts geben.

Nicht zuletzt für die Schüler wird der LOT positive Konsequenzen zeitigen. Sie bekommen durch die Auswertung (z. B. via Evaluationsstruktur) in präziser und differenzierter Weise Rückmeldungen über ihre Lernfortschritte und eventuellen Lernlücken, die gezieltere Lernanstrengungen ermöglichen. Diese richten sich dann nicht mehr auf eine Veränderung der Position innerhalb einer klassenspezifischen Leistungshierarchie, um „statusmarkierende Wertmarken einzuheimsen“ (Schreiner, 1972, 157), sondern auf die individuelle Distanz zum Lernziel. Diese Veränderung läßt sich lernpsychologisch umschreiben als Übergang von der sachfremden (sekundären, extrinsischen) zur sachbezogenen (primären, intrinsischen) Motivation, die für ein lebenslanges Lernen Voraussetzung ist. Der schwache Schüler gelangt durch zielorientierte Leistungsmessung auch zu Erfolgserlebnissen, da seine Leistung nicht ständig durch den Vergleich mit besseren Schülern relativiert

# Evaluationsstruktur

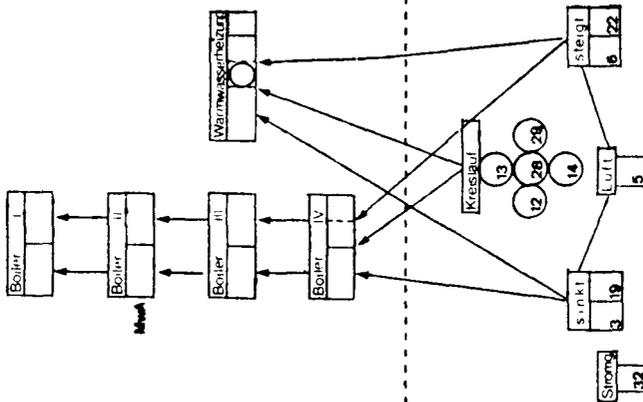
# München-Nord II

Name:

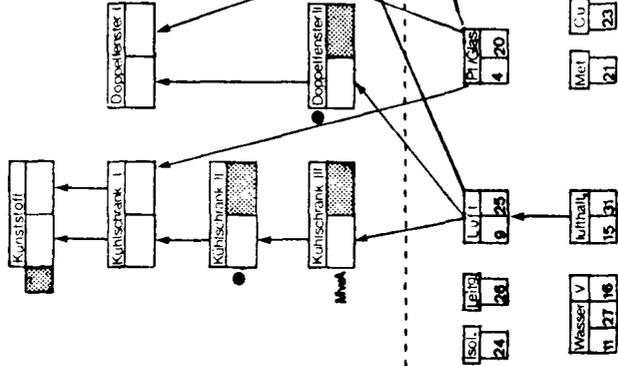
Klasse:

Form:

Wärmeströmung



Wärmeleitung



Wärmestrahlung

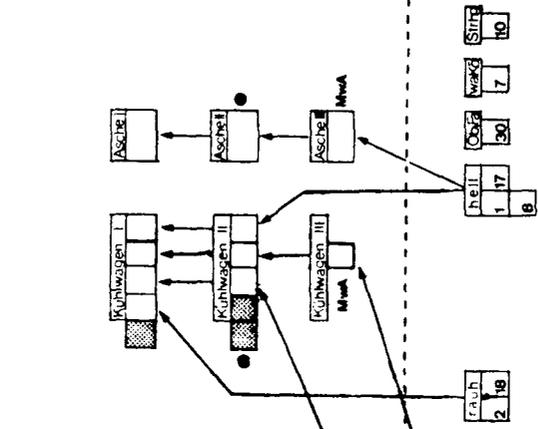


Abb. 2: Evaluationsstruktur für einen curricularen Test (Quelle: Modellcurriculum „Wärmeausbreitung“ der Projektgruppe Schullaufbahnberatung. München, 1972, unveröffentl. Manuskript).

wird. Daß damit eine emotionale Dimension der Schülerpersönlichkeit angesprochen wird, liegt auf der Hand: Erfolg oder Mißerfolg in der Schule bewirken Lust bzw. Unlust am schulischen Lernen. Als Reaktion auf die Auseinandersetzung mit seiner Umwelt, die ihm die Befriedigung vitaler, geistiger und sozialer Bedürfnisse (vgl. oben, Teil I, Abschnitt 2) gewährt oder versagt, bildet sich beim Schüler eine Handlungstendenz (= dynamische Disposition) heraus, die sein weiteres Verhalten bestimmt. Die zielorientierte Leistungsmessung vergrößert die Chancen für Erfolgserlebnisse und vermag so auf die Einstellungen der Schüler zu Schule und Lernen positiven Einfluß zu nehmen.

Die Rückmeldung über den Lernerfolg sollte zum Anlaß genommen werden, mit den Schülern über die schulischen Leistungsanforderungen zu diskutieren und sie dabei von der Fremdbeurteilung zur Selbstbeurteilung der Leistung hinzuführen, wodurch der Prozeß des Mündigwerdens gefördert wird. Mit einer vom ständigen interindividuellen Leistungsvergleich befreiten Schumatmosphäre würde auch eine Bedingung für das Lernen in jenem Bereich erfüllt, der — wohl auch aufgrund seiner Sprödigkeit gegenüber Operationalisierungsversuchen — in einer auf kognitive Fähigkeiten und Fertigkeiten fixierten Schule, trotz seiner Proklamation in Lehrplänen und Richtlinien, ständig zu kurz kommt: der Bereich des sozialen Lernens, der für eine arbeitsteilige demokratische Gesellschaft so notwendige und wichtige Zielsetzung wie „Kommunikation“ und „Kooperation“ umfaßt.

## ANHANG

### Beispiele verschiedener Testaufgabentypen

(Quelle: Lernzielorientierte Tests der Projektgruppe Schullaufbahnberatung. München, 1972)

#### 1. *Richtig-Falsch-Aufgabe:*

An Regentagen strahlt die Sonne weniger Licht aus.  
Ist diese Aussage richtig?

- a) ja
- b) nein

---

#### 2. *Einfache Mehrfachwahlantwort-Aufgaben:*

Eine weiße Zimmerwand ..... einfallende Parallelstrahlen.

- a) verschluckt
- b) erhellt
- c) spiegelt
- d) zerstreut
- e) absorbiert

---

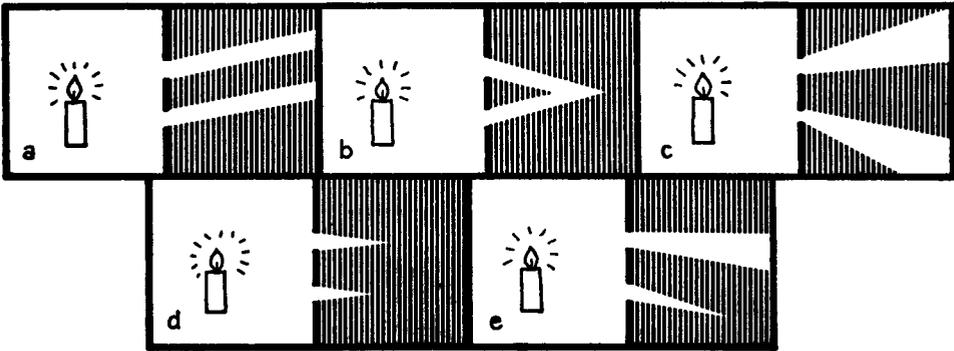
Das Zurückwerfen von Lichtstrahlen nennt man

- a) Blendung — b) Absorption — c) Reflexion — d) Brechung — e) Optik

Nachts wird es bei uns dunkel. Warum?  
Suche die richtige Erklärung!

- a) Nachts sendet die Sonne keine Strahlen aus.
- b) Wir sind dann auf der Schattenseite der Erdkugel.
- c) Nachts kommt die Finsternis.
- d) Die Strahlen der Sonne breiten sich nach allen Richtungen aus.
- e) Dicke Wolken schieben sich vor die Sonne.

Ein durchlöcherter Pappkarton wird in einem dunklen Raum vor eine Kerze gehalten.  
Welches Bild zeigt den richtigen Verlauf des Lichtes *hinter* dem Karton?



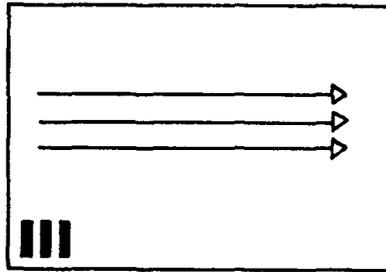
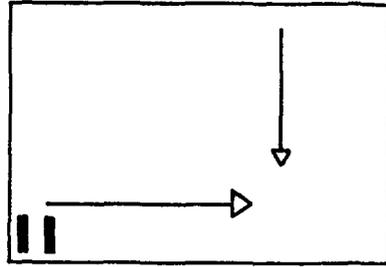
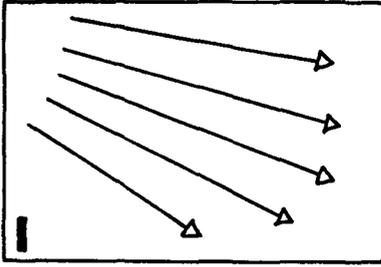
### 3. Komplexe Mehrfachwahlantwort-Aufgaben:

Prüfe folgende Aussagen!

- I. Die Sonne sendet Lichtstrahlen aus.
- II. An Regentagen strahlt die Sonne weniger Licht aus.
- III. Auch in der Nacht sendet die Sonne Strahlen aus.

Welche Aussage(n) stimmt (stimmen)?

- a) alle
- b) I und II
- c) I und III
- d) II und III
- e) nur I



Welches Bild zeigt Lichtstrahlen, die sich geradlinig ausbreiten?

- a) I, II und III
- b) nur I und II
- c) nur II und III
- d) nur III
- e) keines

#### 4. Zuordnungsaufgaben:

Am Schluß der folgenden drei Sätze fehlen die Satzzeichen.

- I. Was wünschst du dir zum Geburtstag
- II. Wünsch dir etwas zum Geburtstag
- III. Ich wünsche mir etwas zum Geburtstag

In welcher der folgenden Antworten sind alle Satzzeichen richtig zugeordnet?

- a) I? II! III.
- b) I! II. III?
- c) I. II? III!
- d) I? II. III!
- e) I! II? III.

Lies folgenden Text!

In die Lücken passen Wörter mit „a“, „ah“ oder „aa“!

Gabi darf beim Kuchenbacken helfen. Zuerst soll sie mit der Küchenmaschine Nüsse *m (I)len*. Während ein *p (II)r* in den Mund wandern, werden die anderen Haselnüsse in den Trichter geworfen. Gabi kann sich ihr Wunderwerk bereits *ausm (III)len*. Noch ist es nicht so weit! Mutter braucht zunächst die Küchen*w (IV)ge*.

Entscheide Dich für einen der Buchstaben (a) bis (e)!

- a) I = a, II = ah, III = ah, IV = aa
- b) I = ah, II = aa, III = a, IV = aa
- c) I = ah, II = a, III = aa, IV = ah
- d) I = a, II = ah, III = a, IV = aa
- e) I = aa, II = aa, III = ah, IV = a

### Literaturverzeichnis

- Becker-Freyseng, W., Krüger & Rietbrock, G.: Anhang zu Projektgruppe: Diagnostik in der Schule. München: Oldenbourg, 1973.
- Block, J. H. (ed.): Mastery learning. New York: Holt, Rineholt & Winston, 1971.
- Bloom, B. S. (ed.): Taxonomy of educational objectives. Handbook I: Cognitive domain. New York: McKay, 1956. Deutsch: Taxonomie von Lernzielen im kognitiven Bereich. Weinheim: Beltz, 1972.
- Bloom, B. S.: Alle Schüler schaffen es; betrifft: erziehung, 1970, 3, Heft 4, 15—27.
- Bloom, B. S., Hastings J. T. & Madaus, G. F.: Handbook on formative and summative evaluation of student learning. New York: McGraw Hill, 1971.
- Borries, Bodo v.: Lernziele und Testaufgaben für den Geschichtsunterricht. Stuttgart: Klett, 1973.
- Carroll, J. B.: Lernerfolge für alle; Westermanns Pädagogische Beiträge, 1972, 24, 7—12.
- Chauncey, H. & Dobbin, J. E.: Der Test im modernen Bildungswesen. Stuttgart: Klett, 1968.
- Ebel, R. L.: Measuring educational achievement. Englewood Cliffs: Prentice-Hall, 1965.
- Frenz, H.-G., Krüger, K. & Tröger, H.: Die Unangemessenheit der herkömmlichen Testdiagnostik für schulische Entscheidungen; in: Projektgruppe: Diagnostik in der Schule. München: Oldenbourg, 1973.
- Gagné, R. M.: Die Bedingungen des menschlichen Lernens. Hannover: Schroedel, 1969.
- Gagné, R. M.: Die Analyse der Unterrichtsziele für die Unterrichtsplanung; in: Glaser, R. (Hrsg.): Programmiertes Lernen und Unterrichtstechnologie. Berlin: Cornelsen, 1971.
- Gaude, P. & Teschner, W.-P.: Objektivierete Leistungsmessung in der Schule. Frankfurt: Diesterweg, 1970.
- Gottschald, W., Neumann-Schönwetter, M. & Soukup, G.: Sozialisationsforschung. Frankfurt/Main: Fischer, 1971.
- Hagener, D.: Lernzielorientierter Test als praxisnahe Curriculum-Evaluation. Zur Funktion und Reichweite lernzielorientierter Tests im politischen Unterricht; Westermanns Pädagogische Beiträge, 1972, 24, 351—364.
- Heipcke, K.: Zur Theorie lernzielorientierter Tests — curriculare Entscheidungen und lernzielorientierte Tests; Zeitschrift für erziehungswissenschaftliche Forschung, 1972, 6, 69—91.
- Henning, S., Köbberling, A., Noack, J., Sobisch, S., Taddicken B. & Wetzig, B.: Entwicklung und Erprobung eines lernzielorientierten Tests; Westermanns Pädagogische Beiträge, 1972, 24, 249—217.
- Holz kamp, K.: Kritische Psychologie. Frankfurt/Main: Fischer, 1972.
- Ingenkamp, K.: Sind Zensuren aus verschiedenen Klassen vergleichbar? betrifft: erziehung, 1969, 2, Heft 3, 11—14 (a).
- Ingenkamp, K.: Möglichkeiten und Grenzen des Lehrerurteils und der Schultests; in: Roth, H. (Hrsg.): Begabung und Lernen. Stuttgart: Klett, 1969 (b).
- Ingenkamp, K.: Probleme der schulischen Leistungsbeurteilung unter besonderer Berücksichtigung des Deutschunterrichts; Der Deutschunterricht, 1971, 23, Heft 3, 54—76 (a).
- Ingenkamp, K. (Hrsg.): Die Fragwürdigkeit der Zensurengebung. Weinheim: Beltz, 1971 (b).
- Ingenkamp, K.: Schulleistungstests; in: Meißner, O. & Zöpfl, H. (Hrsg.): Handbuch der Unterrichtspraxis, Band 1. München: Ehrenwirth, 1973.

- Keck, R.: Gelenke in der Schulorganisation; in: Nicklis, W. S. (Hrsg.): Handwörterbuch der Schulpädagogik. Bad Heilbrunn: Klinkhardt, 1973.
- Kirkland, M. C.: The effects of test on students and schools; Review of Educational Research, 1971, 41, 303—350.
- Klafki, W.: Sinn und Unsinn des Leistungsprinzips in der Erziehung; in: Sinn und Unsinn des Leistungsprinzips. Ein Symposium. München: dtv, 1974.
- Klauer, K. J., Fricke, R., Herbig, M., Rupperecht, H. & Schott, F.: Lehrzielorientierte Tests. Düsseldorf: Schwann, 1972.
- Krathwohl, D. R., Bloom, B. S. & Masia, B. B.: Taxonomy of educational objectives. Handbook II: Affective domain. New York: McKay, 1964.
- Küffner, H.: Probleme der Organisation von Testauswertdiensten. Referat anlässlich der 4. Item-Tagung in Kiel am 27. 9. 1973.
- Kunert, K.: Lehr- und lernzielorientierter Unterricht; Christ und Schule 1972, 18, 110—117.
- Kunert, K.: Curriculumentwicklung und Curriculumevaluation. Unveröffentlichte Dissertation, Universität München, 1974.
- Kutscher, J., Raeithel, A. & Stevens, U.: Entwicklungslinien einer Pädagogischen Diagnostik, in: Projektgruppe: Diagnostik in der Schule. München: Oldenbourg, 1973.
- Lienert, G. A.: Testaufbau und Testanalyse. Weinheim: Beltz, 1969<sup>3</sup>.
- Mauermann, L.: Über die fragwürdigen Gütekriterien von Schulleistungstests; Bayerische Schule, 1973, 26, 331—334.
- Mauermann, L.: Außercurriculare Faktoren unterrichtlicher Kommunikation (geplante Dissertation) München, 1974.
- Mauermann, L., Schulte, H. J. & Seisenberger, G.: Ansätze zu einer curriculumbezogenen Schülerbeurteilung; in: Projektgruppe: Diagnostik in der Schule. München: Oldenbourg, 1973.
- Meyer, H. L.: Trainingsprogramm zur Lernzielanalyse. Frankfurt/Main: Fischer-Athenäum, 1974.
- Moser, H.: Leistungsprinzip: Ökonomische Kategorie und soziale Norm; Die Deutsche Schule, 1972, 64, 202—212.
- Österreich, D.: Wie Menschen über Menschen urteilen; betrifft: erziehung, 1969, 2, Heft 11, 13—19 und Heft 12, 22—26.
- Peterßen, W. H.: Zur Bestimmung und Formulierung von Lernzielen für begrenzte Unterrichtseinheiten; Die Deutsche Schule, 1972, 64, 45—56.
- Peterßen, W. H.: Grundlagen und Praxis des lernzielorientierten Unterrichts. Ravensburg: Maier, 1974.
- Projektgruppe Schullaufbahnberatung: Schlußbericht. München 1973 (eingereicht beim Bayer. Staatsministerium für Unterricht und Kultus).
- Roth, H.: Pädagogische Anthropologie, Band 2. Hannover: Schroedel, 1971.
- Royle, W.: Itembank — Aufbau und Funktion; Die Deutsche Schule, 1972, 64, 547—564.
- Schelsky, H.: Schule und Erziehung in der industriellen Gesellschaft. Würzburg: Werkbund, 1957.
- Schreiner, G.: Sinn und Unsinn der schulischen Leistungsbeurteilung; Die Deutsche Schule, 1970, 62, 226—237 (a).
- Schreiner, G.: Leistungsmessung im Dienst des Berechtigungswesens; Westermanns Pädagogische Beiträge, 1970, 22, 479—481 (b).
- Schreiner, G.: Gegen eine verdinglichende Leistungsbeurteilung; Westermanns Pädagogische Beiträge, 1972, 24, 155—159.
- Schröter, G.: Der eine gibt „1“ und der andere „6“. Sind Deutschlehrer ungerecht? DIE ZEIT vom 4. 12. 1970, S. 65.
- Schröter, G.: Die ungerechte Aufsatzzensur. Bochum: Kamp, 1971.
- Schulreform in Bayern: Band 1, Lehrpläne für die Grundschule, Orientierungsstufe und Hauptschule. München, 1970.
- Schulz, W.: Vorw. zu Mager, R. F.: Lernziele u. Programm. Unterricht. Weinheim: Beltz, 1971.
- Skowronek, H. (Hrsg.): Umwelt und Begabung. Stuttgart: Klett, 1973.
- Strottmatter, P. (Hrsg.): Lernzielorientierte Leistungsmessung. Weinheim: Beltz, 1973.
- Ulich, D. & Mertens, W.: Urteile über Schüler. Weinheim: Beltz, 1973.
- Weis, V.: Zensierungsmodelle und ihre pädagogischen Konsequenzen; Die Deutsche Schule, 1971, 63, 542—553.
- Weiß, R.: Zensur und Zeugnis. Linz, 1965.
- Wendeler, J.: Standardarbeiten — Verfahren zur Objektivierung der Notengebung. Weinheim: Beltz, 1970<sup>3</sup>