

Die Messung schulischen Erfolgs : eine Einführung in die Probleme der Kontrolle des Unterrichtserfolgs durch den Lehrer

Autor(en): **Messner, Rudolf**

Objektyp: **Article**

Zeitschrift: **Schweizer Schule**

Band (Jahr): **57 (1970)**

Heft 16: **Die Messungen schulischen Erfolgs**

PDF erstellt am: **28.07.2024**

Persistenter Link: <https://doi.org/10.5169/seals-533054>

Nutzungsbedingungen

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern. Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden. Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

Haftungsausschluss

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

Die Messung schulischen Erfolgs

Eine Einführung in die Probleme der Kontrolle des Unterrichtserfolgs durch den Lehrer

Rudolf Messner



Rudolf MESSNER (1941), 1960—1963 Primarschullehrer in Österreich. Mitherausgeber von zwei österreichischen Elementarschullesebüchern. 1967 Dr. phil. (Innsbruck: Erziehungswissenschaft und Psychologie). Zur Zeit Wissenschaftlicher Assistent im Fachbereich Erziehungswissenschaft der Universität Konstanz. Arbeitsgebiete: Verhaltenstheorie und Elementarschuldidaktik.*

Wer hat, dem wird gegeben werden;
wer nichts hat, dem wird auch das
wenige genommen.

Wenn man erwachsenen Personen die Frage vorlegt, woran sich bei Kindern der Erfolg ihres Schulbesuchs zeige, so wird man eine Reihe häufig wiederkehrender Antworten erhalten. Man wird vermutlich hören, daß schulischer Erfolg darin zum Ausdruck komme, daß Kinder Lesen, Schreiben, Rechnen und eine Fremdsprache gelernt hätten; vielleicht wird auch darauf hingewiesen, daß Schüler, die lange genug die Schule besucht haben, gelernt haben müßten, längere Zeit still zu sitzen und anderen Personen, vor allem einem Lehrer, zuzuhören, Erwachsene höflich zu grüßen und auf das Vaterland stolz zu sein. Es könnte auch sein, daß wir auf unsere Frage die kluge Antwort erhielten, schulischer Erfolg zeige sich darin, daß junge Menschen gelernt hätten, selbständig zu lernen.

Viel dürftiger werden die Antworten jedoch vermutlich ausfallen, wenn nicht mehr nach dem schulischen Erfolg im allgemeinen gefragt wird, sondern beispielsweise danach, welche schulischen Fortschritte einen Absolventen der dritten Klasse der Elementarschule von einem der vierten Klasse unter-

scheiden. Dies weist auf eine erste Schwierigkeit bei der Bestimmung «schulischen Erfolgs» hin. Die Erwartungen, die an die Schule gestellt werden, können zumeist nur als langfristig zu verwirklichende globale Ziele formuliert werden. Selbst Lehrer dürften auf Befragen kaum Sequenzen von Verhaltensänderungen nennen können, auf Grund derer sich an Schülern ein kontinuierlicher Erfolg der Schule ablesen ließe. Eine zweite Schwierigkeit bei der Bestimmung dessen, was als «schulischer Erfolg» angesehen werden kann, liegt darin, daß Unterricht, also systematisches Lehren, lediglich einen der Faktoren — wenn auch den wichtigsten — darstellt, durch den bei Heranwachsenden Lernvorgänge und Verhaltensänderungen hervorgerufen werden. Die Schülerleistungen, die den Lehrern durch den Lehrplan als Lehrziel vorgeschrieben sind und deren Realisierung durch schriftliche und mündliche Prüfungen kontrolliert wird, werden häufig — man denke vor allem an die Schüler der Höheren Schule — nicht nur durch den *schulischen* Unterricht, sondern durch Nachhilfelehrer und verzweifelte elterliche Bemühungen erreicht, die sich als kompensatorisches System um das oft wenig zielstrebige, aber in seinen

Anforderungen unerbittliche primäre Unterrichtssystem gruppieren. Als «schulischer Erfolg» im strengen Sinne kann jedoch nur jener Lernzuwachs angesehen werden, der durch schulische Lehre erzielt wird.

Die größte Schwierigkeit in der Bestimmung schulischen Erfolges liegt jedoch darin, daß sich die Wirkung des Unterrichts in relativ überdauernden psychischen Dispositionen (vgl. zum Begriff der Disposition STROHAL 1961) niederschlägt, also nicht direkt beobachtet, sondern lediglich durch eine Art Indizienbeweis aus dem Verhalten erschlossen werden kann. Wenn man davon spricht, daß ein Schüler eine «positive Einstellung zu seinem Vaterland» erworben habe — man könnte auch von «Vaterlandsliebe» sprechen —, dann wird auf eine «Disposition» hingewiesen. Dies bedeutet, daß man *einen Aspekt* im Verhalten des betreffenden Schülers beschreibt, von dem *vorhergesagt* werden kann, daß er sich in vielen künftigen Handlungen ausdrücken wird, etwa darin, daß der Schüler die wichtigsten Industriebetriebe und Naturschönheiten seiner Heimat aufzählen kann, daß er sich zum Geburtstag ein Buch über die Geschichte seines Vaterlandes wünscht oder einen Familienausflug zu einer historischen Stätte anregt.

Im genannten Fall kann der Erfolg der schulischen Bemühungen tatsächlich beurteilt werden, weil zwischen der schulischen Beeinflussung und der Situation, in der das Vorhandensein von «Vaterlandsliebe» beobachtet werden kann, nur eine geringe zeitliche Distanz besteht. Zentrale Ziele der Schule im Bereich der Fähigkeiten, Einstellungen und Haltungen (Fähigkeit zum kritischen Denken, Kameradschaftlichkeit, staatsbürgerliches Bewußtsein) verwirklichen sich jedoch erst in der *Lebensbewahrung des einzelnen Schülers*. Auf die Schwierigkeiten, die sich daraus für die Beurteilung der Auswirkungen von Erziehungsmaßnahmen ergeben, hat schon HERBART hingewiesen: «Der Rückstand der pädagogischen Experimente sind die Fehler des Zöglings im Mannesalter. Der Zeitraum für ein einziges dieser Experimente ist also aufs wenigste ein halbes Menschenleben» (HERBART 3/1960, 8).

Zunächst kann dieser Sachverhalt als Warnung für alle Erzieher und Lehrer gedeutet

werden, sich ihrer Erfolge nicht schon deshalb gewiß zu sein, weil gute, idealistisch motivierte Absichten am Werke sind (vgl. BREZINKA 1969)¹. Darüber hinaus sollten jedoch die angedeuteten Schwierigkeiten bei der Bestimmung schulischen Erfolges dazu führen, daß jene gedanklichen Schritte vollzogen werden, die notwendig sind, um erkennen zu können, wann im Unterricht Erfolge erzielt werden und wann nicht. Um dies leisten zu können, muß man lernen, Unterricht als einen Vorgang zu betrachten, in dem durch den Einsatz einer variablen Palette von Lernhilfen versucht wird, das Verhalten von Schülern in *der* Weise zu verändern, daß es sich den Zielen annähert, die man sich vor Beginn des Unterrichts gesetzt hat.

Obwohl seine Messung erhebliche Schwierigkeiten bereitet, ist der Lehrer auf eine kontinuierliche und detaillierte Kontrolle des schulischen Erfolges angewiesen, um das unterrichtliche Vorgehen planen und die Leistungen der Schüler beurteilen zu können. Mittel dieser kontinuierlichen Erfolgskontrolle sind im gegenwärtigen Schulsystem fortwährendes Abfragen der Schüler durch den Lehrer, die Beobachtung der Schüler im Unterricht, insbesondere bei der Lösung von Aufgaben, sowie mündliche Prüfungen, Hausaufgaben und Klassenarbeiten. Dazu kommen, besonders im Hinblick auf die Beurteilung des sozialen Verhaltens und der Motivation («Fleiß»), Beobachtungen des disziplinären Verhaltens der Schüler, Wahrnehmungen über ihr Benehmen während der Pause und Informationen, die Mitschüler gewollt oder unbeabsichtigt über ihre Kameraden geben. Aus einer großen Zahl derartiger Detailinformationen setzt der Lehrer das «Leistungsbild» des einzelnen Schülers zusammen und bringt es als Notenprofil oder verbale Charakterisierung in Zeugnissen zum Ausdruck.

Im folgenden soll zunächst auf zwei Mängel der gegenwärtigen Praxis der schulischen Erfolgskontrolle aufmerksam gemacht werden, die zu einer Reihe von erzieherisch bedenklichen Konsequenzen führen.

Dadurch soll der Blick für eine Analyse der Funktionen des schulischen Kontrollsystems (Abschnitt 2) bzw. für einige Vorschläge zur

¹ Vgl. die Zusammenfassung dieser Schrift in «Schweizer Schule» 1970/7/8, S. 282 ff.

Beseitigung der krassesten Mängel der derzeitigen Beurteilungspraxis (Abschnitt 3) geschärft werden.

1. Mängel des gegenwärtigen Systems der Lernkontrolle

Einen ersten Einblick in die Probleme der gegenwärtigen Praxis schulischer Erfolgskontrolle gewinnt man durch die Analyse einschlägiger empirischer Befunde.

1.1. R. WEISS: Das Fehlen präziser Beurteilungsmaßstäbe

Rudolf WEISS legte vor einigen Jahren 92 oberösterreichischen Lehrern der vierten Schulstufe im Rahmen von Arbeitsgemeinschaften zwei Aufsätze und zwei Rechenschularbeiten zur Beurteilung vor (vgl. WEISS 1965, 148—173).

Einer der beiden Aufsätze hatte folgenden Wortlaut:

Ein Hund allein auf der Straße

Meine Mutter und ich gingen einmal spazieren. Plötzlich stolperte ich. Die Mutter lachte und sagte: «Du bist ja über einen Hund gestolpert!» Es war ein lieber, kleiner Budel. Da sahen wir, das der arme Hund ganz allein auf der Straße ist. Er schaute uns ganz treuherzig an. Wir gingen dann zur Polizei und lieferten ihn ab. Ich dachte, er muß sich verirrt haben.

Zu bewerten waren *Rechtschreiben, Stil, Inhalt* und die *Gesamtleistung*.

Die 92 Versuchspersonen beurteilten den Aufsatz in folgender Weise:

<i>Rechtschreiben:</i>		<i>Stil:</i>	
Note 1	10 %	Note 1	23 %
2	18 %	2	52 %
3	41 %	3	15 %
4	24 % AM: 2,99	4	10 % AM: 2,12
5	7 % STR: 1,04	5	— STR: 0,96
<i>Inhalt:</i>		<i>Gesamtnote:</i>	
Note 1	40 %	Note 1	8 %
2	46 %	2	46 %
3	10 %	3	39 %
4	4 % AM: 1,68	4	7 % AM: 2,45
5	— STR: 0,79	5	— STR: 0,80

(AM: Arithmetisches Mittel; STR: Streuung)

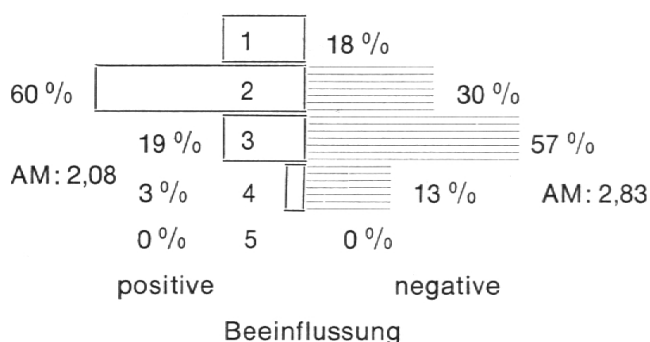
Eine Betrachtung dieser Ergebnisse zeigt: Die Beurteilungen streuen über vier (bei Rechtschreiben sogar über fünf) Stufen der Notenskala. Am relativ besten wurde der Inhalt des Aufsatzes beurteilt, am relativ schlechtesten die Rechtschreibleistung. Hier ist auch die Streuung (1,04) am größten. Für

die Gesamtnote ist offensichtlich die Rechtschreibnote maßgebender als die Bewertung des Inhalts.

Wie kann man diese Ergebnisse interpretieren? Zeigen sie nicht das Fehlen jeglicher objektiver Maßstäbe im schulischen Prüfungssystem, da dieselbe Arbeit z. B. hinsichtlich ihrer Rechtschreibung sowohl mit «sehr gut» als auch mit «nicht genügend» beurteilt werden kann? Die Interpretation wird etwas vorsichtiger erfolgen müssen. Geprüft wurde im vorliegenden Fall ja nicht, wie ein Prüfungsaufsatz im Rahmen des Unterrichts bewertet wird, sondern lediglich wie die Benotung *isoliert außerhalb des Unterrichts erfolgt*. Es ist anzunehmen, daß sich bei einer Beurteilung im Rahmen einer Schularbeit durch den Vergleich mit anderen Arbeiten bzw. auf Grund früherer Leistungen ein klasseninterner Maßstab einpendelt, der die Unsicherheit eingrenzt. Dennoch legt das vorliegende Experiment das Urteil nahe, daß die Kriterien der Leistungsbeurteilung auf bestimmten Schulstufen und für bestimmte Fächer viel zu wenig präzisiert sind, daß also ungeklärt ist, welchen Umfang, welche stilistischen und sprachlichen Eigenschaften ein «sehr guter» Aufsatz im vierten Schuljahr aufweisen soll. Dies hängt damit zusammen, daß die Kriterien guter bzw. schlechter schulischer Leistungen in den Richtlinien für Lehrer nur sehr vage beschrieben sind. Ein Beispiel dafür liefert die Interpretation des Skalenwertes «sehr gut», der bestmöglichen Note, die den österreichischen Lehrern von der Schulbehörde gegeben wird: «Als ‚sehr gut‘ sind die Leistungen anzusehen, die aufweisen, daß ein Schüler den Sinn seiner Aufgabe vollkommen und mit deutlich merkbarer Selbständigkeit erfaßt hat oder den durchgenommenen Lehrstoff ganz und sicher beherrscht.» Diese Kennzeichnung muß wegen ihres allgemeinen, viele Fächer betreffenden Charakters notwendig vage sein. Was fehlt, sind die Präzisierungen für einzelne Fächer und Schulstufen. Wann ist bei einem Aufsatz die «Aufgabe vollkommen und mit deutlich merkbarer Selbständigkeit erfaßt», wann der durchgenommene Lehrstoff «ganz und sicher beherrscht»?

Im zweiten Teil des von WEISS durchgeführten Experiments wurde einer der beiden Lehrergruppen zugleich mit den Aufsätzen

folgende erläuternde Bemerkung vorgelegt: «Zwei Aufsätze aus der vierten Klasse einer Volksschule. Der erste stammt von einem durchschnittlichen Schüler (beide Elternteile berufstätig, liest gerne Schundhefte), der zweite von einem sprachlich begabten Buben (Vater Redakteur bei einer großen Linzer Tageszeitung).» Der andere Teil der Lehrer las dieselbe Einleitung, aber mit vertauschtem Bezug (WEISS 1965, 149 ff.). Die Beurteilung der Aufsätze unter positiver und negativer Beeinflussung ergab folgende Streuungen der Gesamtnote:



Was ist hier geschehen? Kein Mensch ist imstande, das für ihn bedeutsame soziale Geschehen in seiner ganzen Komplexität und Differenziertheit zu erfassen. Der Mensch muß sich jedoch im alltäglichen Leben trotz seines beschränkten Wissens zurechtfinden. Dies ist nur möglich, wenn er auf vereinfachende Annahmen über die Regelmäßigkeiten des sozialen Lebens vertraut. Das Verhalten anderer Menschen wird aus diesem Grunde «typisiert», d. h. geringe Unterschiede werden in starkem Maß übertrieben und unter Vernachlässigung individueller Variationen allen Angehörigen einer Gruppe (*der Italiener, der Berner, der Schulschwänzer*) zugeschrieben. Derartige «*Stereotypisierungen*» steuern in gewissem Maße auch die Beurteilung einzelner Schüler durch ihre Lehrer. Wer zur Gruppe jener Schüler gehört, dessen Eltern berufstätig sind und der gerne «Schundhefte liest», von dem wird, obwohl der Schluß im Einzelfall widerlegbar sein mag, eine schlechtere Leistung erwartet als von einem Schüler, dessen Vater «Redakteur bei einer großen Tageszeitung» ist. Da im vorliegenden Experiment Vergleichsmöglichkeiten fehlen, wird bei der Suche nach zusätzlichen Kriterien

auf die angebotenen Zusatzinformationen zurückgegriffen.

Damit ist auf einen generellen Mangel der schulischen Prüfungspraxis hingewiesen. Die Bedingungen, unter denen gegenwärtig schulische Beurteilungen vorgenommen werden, sind nur in geringem Maße standardisiert, d. h. es werden verschiedenen Schülern verschieden viele und verschieden schwierige Aufgaben vorgelegt. Selbst dann, wenn in einer schriftlichen Arbeit allen dieselben Aufgaben gestellt werden, fehlen zu meist exakte Auswertungsschlüssel, so daß für die Vergabe der Noten noch ein breiter Spielraum für eine wohlwollende oder nicht wohlwollende Interpretation bleibt. Man denke z. B. an die oft sehr willkürliche Art, in der Fehler in Lateinschularbeiten als «schwere» oder «leichte» klassifiziert werden. Man erinnere sich an die Stufungen der Richtigkeit der Lösung einer Mathematikaufgabe als «teilrichtig», «halbrichtig», «richtiger Lösungsgang», die es trotz standardisierter Aufgabenstellung häufig erlauben, beim «guten» Schüler ein Versehen in der Durchführung, beim «schlechten» Schüler aber fehlende Einsicht in den Lösungsweg zu vermuten und dementsprechend die Zensuren zu differenzieren.

Die fehlende Standardisierung der Prüfungsbedingungen läßt subjektive Faktoren in der Person des Prüfers Einfluß auf die Beurteilung gewinnen. Der Einfluß von «Stereotypen» ist nur einer unter mehreren, welche in sozialpsychologischen Untersuchungen erforscht worden sind.

Besonders interessant ist eine Arbeit von ZILLIG, in der nachgewiesen werden konnte, daß die Einstellung zu einem Schüler und die daraus resultierende Erwartung des Lehrers von der Leistung des Schülers bereits die Wahrnehmung beeinflussen. ZILLIG stellte fest, daß Lehrer bei der Korrektur von schriftlichen Arbeiten «guter» Schüler mehr Fehler übersehen hatten als bei der Durchsicht von Heften «schlechter» Schüler (ZILLIG 1928).

Die Einschätzung eines Schülers als «gut» oder «schlecht» kommt keineswegs nur auf Grund früherer Leistungen des Schülers zustande. In vielen Untersuchungen wird als sogenannter «Haloefekt» von der Tendenz berichtet, Leistungen in *einem* Verhaltensbereich als generelle Indikatoren für die

Leistungsfähigkeit in anderen Verhaltensbereichen zu interpretieren. Hierher gehört die Erscheinung, daß Schüler mit schlampiger Kleidung oder Schüler mit den Merkmalen niedriger sozialer Schichten (z. B. fehlender sprachlicher Gewandtheit) unterschätzt zu werden pflegen, während umgekehrt positiv gewertete soziale Eigenschaften oder die erfreuliche Erfolgsbilanz in einem bestimmten Fach auf die Beurteilung der übrigen Leistungen «ausstrahlen» (vgl. HÖHN 1967). Es dürfte z. B. sehr häufig vorkommen, daß disziplinäre Schwierigkeiten, die ein Schüler bereitet, zumindest durch eine kritischere Beurteilung seiner Leistungen «aufgerechnet» werden.

Wenn man in Betracht zieht, daß die Beurteilung eines Schülers durch den Lehrer eine der wichtigsten Determinanten der Lernmotivation und damit der weiteren Lernanstrengungen des Schülers darstellt, dann wird der «Teufelskreis» ersichtlich, in den ein auf Grund eines Stereotyps nicht sachgerecht beurteilter Schüler gerät. Durch das Erlebnis des Mißerfolgs sinken das Selbstvertrauen, die Lernmotivation und damit die objektive Leistungsfähigkeit — was wiederum zu einem verstärkten Erlebnis des Mißerfolgs führt. Dazu kommt, daß die der eigenen Leistung zugeschriebenen Mängel vom Lehrer und vom Schüler nicht auf objektive individuelle Gegebenheiten (vorhandene Lernvoraussetzungen, Intelligenz) zurückgeführt, sondern als persönlicher Defekt interpretiert werden, für den man sich zu schämen hat (ASCH 1951, vgl. auch WEISS 1969, 39 f.). So läßt sich mit wenigen Sätzen die Beziehung zwischen Mängeln der gegenwärtigen Prüfungspraxis und jenem Bild des schulbeladenen Versagers skizzieren, das bei jeder kritischen Analyse mit aller Deutlichkeit als Negativposten des gegenwärtigen Schulsystems hervortritt.

Dieser Effekt ist besonders treffend als «self-fulfilling-prophecy», d. h. als eine durch sich selbst in Erfüllung gehende Prophezeiung gekennzeichnet worden. Sie zeigt sich überall dort, wo Schüler unter dem Druck der positiven oder negativen Erwartungen, die ihre Lehrer und Mitschüler von ihnen haben, sich in ihren Leistungen auf ein hohes oder niedriges Niveau einpendeln. Ein Lehrerwechsel und die damit verbundene Veränderung der Leistungserwartung kann

— offensichtlich in Zusammenhang mit diesem Phänomen — gelegentlich zu auffälligen Leistungsveränderungen einzelner Schüler führen¹.

1.2. Karlheinz INGENKAMP:

Klasseninterne Normierung

Dadurch, daß die einzelne Schulklasse die soziale Einheit darstellt, innerhalb derer der Lehrer, isoliert von der Beurteilungspraxis anderer Klassen, die Notenskala anwenden muß, entwickeln sich außerordentlich unterschiedliche klasseninterne Beurteilungssysteme. Zur Illustration sei auf eine Untersuchung von INGENKAMP verwiesen, in der die Rechenleistungen von 37 sechsten Klassen verglichen worden sind². Die drei besten (1, 2, 3) und die drei schlechtesten Klassen (4, 5, 6) erreichten im Test folgende durchschnittliche Punktzahlen:

1	46,47	4	25,43
2	41,62	5	25,33
3	40,68	6	21,76

Die Mittelwerte (Arithmetisches Mittel) der Rechenzensuren in denselben Klassen ergeben folgendes Bild (6-stufige Notenskala):

1	3,21	4	3,17
2	3,17	5	3,47
3	3,34	6	3,26

In der besten Klasse (1) liegen 75 % der Schüler mit ihrer Testpunktzahl über dem besten Ergebnis der schlechtesten Klasse. (37). Dennoch weichen die erteilten Zensuren in beiden Klassen kaum voneinander ab. Sollten sich derartige Ergebnisse auch in anderen Untersuchungen bestätigen, so wäre dies eine Illustration dafür, wie sehr der schulischen Leistungskontrolle eine regionale Normierung fehlt. Der einzelne Lehrer orientiert sich bei der Beurteilung an einem klasseninternen Maßstab. Dies führt in der genannten Untersuchung im Extremfall (Vergleich der Klassen 1 und 37) dazu, daß Leistungen, die in der einen Klasse mit Note 1 oder 2 beurteilt werden, in einer anderen Klasse die Note 4 ergeben. Diese Unter-

¹ Weitere Literatur zum Problem der Objektivität von schulischen Prüfungspraktiken ist ausführlich referiert in INGENKAMP 1969.

² Vgl. Schultests: Mode oder pädagogische Notwendigkeit. Weinheim o. J. — Vgl. auch INGENKAMP 1968 a.

schiede können weitreichende Konsequenzen haben, z. B. wenn sie dazu führen, daß Schüler, die nach Absolvierung der einen Klasse als geeignet zum Besuch einer weiterführenden Schule erklärt werden, in einer anderen Klasse oder Schule diese Qualifikation nicht erhalten, obwohl sie — objektiv gesehen — qualifizierter wären¹.

2. Die Überforderung des gegenwärtigen schulischen Kontrollsystems

Die referierten Befunde geben einen ersten Einblick in das Bedingungsgefüge schulischer Erfolgskontrolle. Es entstand der Eindruck, daß es sich um ein außerordentlich labiles, durch affektive Faktoren störbares Instrument handelt. Gerade dieses Werkzeug soll jedoch, wie der folgende Überblick zeigt, einer großen Zahl sehr unterschiedlicher Zwecke dienen (vgl. CHAUNCEY und DOBBIN 1968).

2.1. Funktionen schulischer Kontrolle

(1) Beurteilung des Leistungsniveaus einzelner Schüler

Schulische Prüfungen dienen hauptsächlich dazu — so scheint es zunächst —, diese

¹ In einer 1969 erschienenen Arbeit — «Zur Problematik der Jahrgangsklasse» — stellt INGENKAMP seine empirischen Befunde zum Problem der klasseninternen Normierung von schulischen Zensuren in einem größeren Zusammenhang dar (INGENKAMP 1969 b, 227—235). Unter anderem zeigt sich, daß die Zensuren, die von Lehrern gegeben werden, mit den Testrohpunkten, welche dieselben Schüler in Schultests erreichen, in dem Sinne übereinstimmen, daß Schüler mit guten Zensuren im allgemeinen auch gute Testleistungen aufweisen und umgekehrt. Außerordentlich unterschiedlich sind jedoch — wie erwähnt — die Zensuren, welche Schüler mit derselben Testleistung in verschiedenen Klassen bekommen. Damit bestätigt sich eine fatale Konsequenz des gegenwärtigen Beurteilungssystems: die Chance der Aufnahme in weiterführende Schulen hängt häufig von der zufallsbedingten Zugehörigkeit zu einer bestimmten Schulklasse ab. Die hier aus Deutschland berichteten Ergebnisse werden durch Untersuchungen, die in Österreich durchgeführt worden sind, bestätigt (HOLZINGER 1955). Auch hier bediente man sich eines Vergleichs der Leistungen von Schülern mit unterschiedlichen Schulnoten in einem Intelligenztest. Ein bemerkenswertes Resultat: Es gab mit «nicht genügend» beurteilte Schüler, deren Testleistung über dem Durchschnitt der mit «sehr gut» bewerteten lag (HOLZINGER 1955, 80 ff.).

Aufgabe zu erfüllen, z. B. festzustellen, ob der Schüler A Dreisatzrechnungen richtig lösen, den Unterschied zwischen Weizen und Roggen nennen oder sämtliche drei Strophen eines Liedes vorsingen kann. Um ein begründetes Urteil über die Leistungshöhe eines Schülers auf einem Gebiet fällen zu können, ist es notwendig, das Vorhandensein der erwünschten Fähigkeit durch mehrere gleichwertige Aufgaben zu prüfen, also beispielsweise dem Schüler nicht nur *eine* Dreisatzrechnung vorzulegen, sondern ihn eine Reihe von Aufgaben mit steigendem Schwierigkeitsgrad lösen zu lassen, in denen verschiedene Aspekte der komplexen Fähigkeit «Lösen von Dreisatzrechnungen» geprüft werden. Dieses Prinzip wird in der schulischen Prüfungspraxis wenig beachtet.

(2) Beurteilung des Lernzuwachses

Wenn die Leistung eines Schülers zu verschiedenen Zeitpunkten gemessen wird und wenn sie im Meßzeitraum angestiegen ist, kann der schulische Erfolg als *Lernzuwachs* ausgedrückt werden. Die Information über den Lernzuwachs ist von größter Bedeutung, um die individuellen Lernfortschritte abschätzen zu können. Es dürfte im Hinblick auf die Motivierung der Schüler einer der größten Mängel des gegenwärtigen Schulsystems sein, daß mit schulischen Bewertungen lediglich der absolute Leistungsstand, nicht aber der Lernzuwachs ausgezeichnet wird. Dadurch geschieht es immer wieder, daß Schüler, die auf einem Gebiet bereits große Vorkenntnisse besessen haben, auch bei minimalem Lernfortschritt gute Noten erhalten, während Schüler mit einer niedrigen Ausgangsleistung trotz größter Anstrengungen unbeachtet bleiben. Typisch dafür sind z. B. die Vorteile, welche sich Kinder aus einem kultivierten Sprachmilieu verschaffen, oder die Geringschätzung, mit der oft eine leichte Besserung in der Sauberkeit der schriftlichen Arbeiten bei einem «unordentlichen» Schüler registriert wird.

(3) Beurteilung der Lernschwierigkeiten

Eine dritte Funktion der Kontrollaktivität des Lehrers besteht darin, jene Schwierigkeiten festzustellen, die *während* des Lernprozesses auftreten. Derartige Schwierigkeiten können mit einzelnen Phasen des Lernpro-

zesses verbunden sein. Sie sind z. B. zu erwarten, wenn eine nur formal beherrschte Rechnungsart auf komplex formulierte «Testaufgaben» angewendet werden soll, die eine Reihe schwer durchschaubarer Rechenoperationen verlangen, oder wenn im Erstleseunterricht nach der Ganzwortmethode jener Moment erreicht ist, ab dem Laute und Buchstaben als Elemente von Wörtern «ausgegliedert» werden sollen. In der gegenwärtigen Beurteilungspraxis wird vermutlich kaum zwischen der Diagnose von Lernschwierigkeiten, die oft durchaus behebbare Hemmungen im Lernprozeß darstellen, und der Feststellung der schließlich von einem Schüler erreichten Leistung unterschieden. Schüler, die Schwierigkeiten beim Lernen haben, werden zumeist als weniger leistungsfähig angesehen. Leitidee ist der «*konfliktlose Lernprozeß*». Er scheint pädagogische Meisterschaft und «Begabung» der Schüler anzudeuten, obwohl er den Erfahrungen widerspricht, die jeder schöpferisch Tätige bei seinen eigenen gedanklichen Anstrengungen macht.

(4) *Beurteilung von Lehrmethoden*

Während zur Beurteilung des individuellen Leistungsniveaus die mehrmalige Messung der Leistung *eines* Individuums notwendig ist und die Intention dahin geht, individuelle Leistungsausprägungen zu erfassen, tritt bei der Beurteilung der Effizienz einer Lehrmethode die Bedeutung der differentiellen Merkmale der einzelnen Individuen zurück. Im ersten Fall handelt es sich darum, zu unterscheiden, welcher von mehreren *Schülern* erfolgreicher ist, im zweiten Fall geht es unter Vernachlässigung der individuellen Unterschiede darum, zu entscheiden, welche von mehreren *Lehrmethoden* effektiver ist. Man muß zu diesem Punkt bemerken, daß es dem einzelnen Lehrer in seinem Unterricht kaum möglich sein dürfte, objektive Unterschiede zwischen Lehrverfahren festzustellen. Obwohl zahlreiche Lehrer darauf «schwören», daß die «Ganzheitsmethode» wirksamer sei als die «synthetische Methode», ist es selbst bei minutiösem methodologischen Vorgehen bisher nicht gelungen, innerhalb des bestehenden Schulsystems derartige Urteile zu bestätigen: SCHMALOHR konnte z. B. weder im Hinblick auf die Arbeitshaltung noch bezüglich

der Schulleistungen oder des Leseverständnisses methodenspezifische Leistungsunterschiede nachweisen (SCHMALOHR 1961).

(5) *Beurteilung der Realisierung der Lernziele*

Es ist offensichtlich nicht damit getan, daß ein Lehrer darauf achtet, die Schülerleistungen möglichst genau zu messen. Es scheint ebenso wichtig, daß sich die Messung auf Verhaltensweisen bezieht, welche den *erwünschten Erziehungszielen entsprechen*. In der gegenwärtigen Unterrichtspraxis wird dieser Aspekt schulischer Prüfungen nur in geringem Ausmaß reflektiert. Man hofft vielmehr, daß sich die Zielangemessenheit der geforderten Leistungen dadurch «von selbst» einstellt, daß man dem Schematismus der Lehrpläne, Lehrbücher und Arbeitsmaterialien, vielleicht auch den Unterrichtsvorbereitungen erfahrener Kollegen folgt. Obwohl jedoch allen Lehrern und Schulen dieselben Lehrpläne vorgeschrieben sind, gibt es große Unterschiede hinsichtlich der tatsächlich angestrebten Schülerleistungen. Während der Lehrer A in Aufsätzen Rechtschreibfehler nicht sehr beachtet, falls der Inhalt lebendig und der Stil abwechslungsreich sind, ahndet der Lehrer B vor allem die falsch geschriebenen Wörter und die regelwidrig gebauten Sätze. Während der Lehrer A durch die äußere Form von Mathematikschularbeiten wenig in seinem Urteil über den Lösungsgang beeinflusst wird, neigt der Lehrer B immer dann zur besseren Note, wenn das Schriftbild «ordentlich» sowie frei von Streichungen und Kritzeleien ist.

Es ist zu einer der wichtigsten Aufgaben der Unterrichtsforschung geworden, die Ziele der Schule viel präziser und detaillierter zu beschreiben, als dies bisher geschehen ist. Den Anlaß dazu haben Psychologen gegeben, die sich mit Fragen der Unterrichtsplanung und der Messung des Lehrerfolges befaßten (vgl. TYLER 1949). Sie waren gewohnt, menschliche Leistungen in der Begrifflichkeit *beobachtbaren Verhaltens* zu beschreiben. Die Qualifikationen, welche die Schule vergeben will, sind jedoch zumeist in der vieldeutigen Sprache der «Lehrstoffe» und «Lehrinhalte» formuliert. AEBLI wies auf dieses Problem bereits 1951 hin: «Aber was bedeutet es, einen Gegenstand wie den ‚Hebel‘ oder einen Begriff wie den ‚gewöhn-

lichen Bruch' zu kennen? Meint man damit die Fähigkeit, eine Definition davon zu geben? Offensichtlich nicht! Will man sagen, der Schüler müsse sich den Hebel vorstellen, seinen Funktionsmechanismus vor sich sehen können?» (AEBLI 1966, 14). Wenn man Lehrziele in die Terminologie der Verhaltenswissenschaften «übersetzt», d. h. jene «*Endverhaltensweisen*» zu nennen versucht, die der Lernende am Ende einer Unterrichtseinheit zeigen können soll, betritt man das faszinierende Forschungsgebiet der Differenzierung und Klärung schulischer Zielsetzungen (vgl. MAGER 3/1969)¹.

(6) *Beurteilung von Nebenwirkungen des Unterrichts*

Das schulische Kontrollsystem sollte nicht nur erfassen, ob der Unterricht die *erwarteten* Wirkungen zeigt, sondern auch für die günstigen oder ungünstigen *Nebenwirkungen* schulischer Beeinflussung sensibel sein. Der Lehrer ist allerdings auf Grund seiner Integration in das Schulsystem kaum in der Lage, jene Akte der Distanzierung vorzunehmen, die notwendig sind, um zu erkennen, daß Schülerreaktionen, wie Gewöhnung an passives Zuhören, kritikloses Anpassen an die von Lehrern vertretenen Meinungen, Orientierung des Lernverhaltens an von außen kommenden Belohnungen und Bestrafungen, im Sinne der Ziele des Schulsystems unerwünschte Nebenwirkungen darstellen. Gegenwärtig wird in unseren Schulen vor allem danach gefragt, ob und in welchem Maße die Ziele der Schule erreicht werden. Womit die Realisierung dieser Ziele erkauft werden muß, welche unerwünschten affektiven und sozialen Nebenwirkungen eintreten — über dieses Problem beginnt man eben erst für den Bereich universitärer Lehre nachzudenken (vgl. DYER 1966).

(7) *Selektion der Schüler*

Eine Leistung, die vom schulischen Kontrollsystem mit größter Selbstverständlichkeit er-

¹ Von besonderer Bedeutung ist die Präzisierung der Lehrziele durch ihre Beschreibung in Begriffen erwarteten Schülerverhaltens für die Revision von Schulleistungstests geworden. Es scheint, daß die traditionellen Schultests, die überwiegend an konventionellen schulischen Zielsetzungen orientiert sind, durch Tests ersetzt werden, die auf präzise Lernzielbeschreibungen bezogen sind (critterion-referenced tests) (vgl. GLASER 1963).

wartet wird, ist jene der *Selektion* von Schülern. Zunächst ist es wichtig, sich bewußt zu machen, daß *diese* Funktion des schulischen Kontrollsystems durchaus von jener der *Beurteilung der Schülerleistungen* getrennt werden kann, auch wenn eine derartige Differenzierung auf Grund der gewohnten Verknüpfung zwischen den Tätigkeiten der Leistungsmessung und der sozialen Selektion gedanklich nur schwer nachvollzogen werden kann. In den meisten Lehrern dürfte die Vorstellung einer Schule ohne Zensuren, obwohl sie gelegentlich davon träumen, einen «horror vacui» hervorzurufen, vielleicht vor allem deshalb, weil die Zensuren als jenes Mittel der Steuerung des Schülerverhaltens angesehen werden, das selbst dem undiszipliniertesten Schüler gelegentlich und in verstärktem Maße, wenn sich die Tage des Zeugnisses nähern, einigen Schrecken einzujagen vermag. Es bedarf keiner näheren Erläuterung, daß hinter einer derartigen Begründung der Notwendigkeit von Zensuren eine ziemlich simple Theorie über die Stimulatoren menschlicher Antriebe steht.

Soziologische Analysen, die besonders von H. SCHELKY angeregt worden sind, haben einen tieferen Einblick in die sozialen Mechanismen gegeben, welche die Auslesefunktion der Schule hervorbringen. Demnach ist die Schule nach dem Wegfall der sozialen Auszeichnung durch Geburt («Adel») zur *zentralen* Instanz für die Zuteilung sozialen Ansehens geworden. Untersuchungen neueren Datums haben diese These SCHELKYs teilweise widerlegt. Es ist nämlich deutlich geworden, daß in das Schulsystem Auffassungen von erwünschten (und folglich positiv benoteten) Leistungen «hineinprogrammiert» sind, durch die Kinder aus der Ober- und Mittelschicht gegenüber Kindern aus der Unterschicht begünstigt werden. Wie immer dem sei, die selektive Funktion wird der Schule und ihren Lehrern von der Gesellschaft auf administrativem Wege vorgeschrieben, ohne Rücksicht darauf, daß die erste erzieherische Absicht der Schule auf eine *optimale Förderung jedes einzelnen Schülers* gerichtet ist. Lehrer, die angesichts der sozialen Konsequenzen, welche die Leistungsmessung auf Grund der mit ihr verknüpften Ausleseeffekte hat, in Gewissenskonflikte geraten, bekommen die geschil-

derte erzieherische Dysfunktionalität ihres Prüfverhaltens zu spüren.

In diesem Zusammenhang ist es von Interesse, näher zu untersuchen, welche einzelnen Aktivitäten bei der Vergabe von Zensuren erforderlich sind. Zunächst werden — wie beschrieben — die Leistungen jedes einzelnen Schülers gemessen. In einem zweiten Schritt werden diese Einzeldaten verglichen und die Schüler auf Grund dieser Operation in eine Rangreihe gebracht. Bestimmte Gruppen der gereihten Schüler werden zusammengefaßt und fünf oder mehr Notenstufen zugeordnet. Die fatalen Konsequenzen der Notengebung entstehen durch zwei Gegebenheiten: Erstens stellen die einzelnen Stufen der Notenskala ein zweipoliges Kontinuum dar, d.h. die Notenskala besitzt ein positives und ein negatives Ende. Zweitens besteht die behördlich normierte Auffassung, daß sich die Gesamtzahl der Schüler in *der* Weise über die Notenstufen verteilen müsse, daß relativ wenige Schüler sehr gute oder sehr schlechte Leistungswerte erreichen, die Mehrheit der Schüler jedoch um die mittlere verhältnismäßig neutrale Qualifikation (z. B. «befriedigend») gruppiert ist. Es wird also angenommen, daß sich Schülerleistungen nach dem Modell der «Normalverteilung» (*GAUSSsche Kurve*) verteilen. Der Kern dieser «Normalverteilungsideologie» besteht darin, daß Gruppen von Personen in «Leistungsfähige» und «Nicht-Leistungsfähige» aufgeteilt werden, d. h., daß Aufgaben in jedem Falle so zu stellen sind, daß ein bestimmter Anteil der Schüler in die Kategorie «nicht genügend» fällt. Der Automatismus dieser Ideologie führt beispielsweise dazu, daß in jedem Jahrgang des Gymnasiums ein relativ konstanter Anteil «schlechter» Schüler ausscheidet, d. h. auf jeder Stufe wird von neuem eine Gruppe von «Versagern» definiert. Die Absurdität dieser Konsequenz ist kaum mehr bewußt. Angenommen, von 40 Schülern, die die erste Klasse eines Gymnasiums besuchen, scheiden auf Grund «nicht-genügender» Lateinleistungen (nach dem Modell der Normalverteilung) bei einer fünfstufigen Notenskala 8 % (= 3) aus. Man könnte der Meinung sein, daß damit das Problem der «Versager» gelöst wäre. Was aber geschieht? In der zweiten Klasse und in jeder weiteren scheiden zumeist wie-

derum Schüler wegen nicht-genügender Lateinkenntnisse aus: Das Modell der Gaußschen Kurve führt dazu, daß mit größter Selbstverständlichkeit Jahr für Jahr die untersten 8 bis 10 % kurzerhand zu «Versagern» deklariert werden. Die Vorstellung der erzieherisch faszinierenden Möglichkeit, daß *Qualifikationen* entscheidend sind und daß *ohne weiteres alle Schüler einer Klasse diese Qualifikation erreichen können*, geht ebenso verloren, wie die Überzeugung, daß die Ursache für schlechte Leistungen nicht nur in der fehlenden «Begabung» der Schüler liegen könnte, sondern mit viel größerer Wahrscheinlichkeit in der Qualität des Unterrichts.

Daß dieser Mechanismus durchbrochen werden kann, beweisen die Benotungen in jenen «Nebenfächern» (man beachte: die administrative Funktion der Selektion bewirkt eine Teilung der Fächer in «Haupt» und «Neben»fächer), in denen als schlechteste Note, die nur sehr selten unterschritten wird, die neutrale Notenstufe, z. B. «4», vergeben wird.

Die Notenskala ist für die meisten der genannten didaktischen Funktionen, z. B. die detaillierte Information der Schüler über ihren Leistungsstand, völlig ungeeignet. Sie erweist sich jedoch als das zentrale Instrument zur autoritären Steuerung des Anpassungsprozesses von Schülern an die durch die Schule repräsentierten Normen der Gesellschaft¹.

Je nach den Meßoperationen, die der Zensurierung vorangehen, können Noten — didaktisch gesehen — verschiedenes ausdrücken:

- den erreichten Leistungsstand zu Ende der Beurteilungsperiode;
- den Lernzuwachs in der Beurteilungsperiode;
- die Durchschnittsleistungen des Schülers innerhalb der Beurteilungsperiode;
- das Ausmaß, in dem nach Meinung des Lehrers ein Lernziel realisiert worden ist;
- die Position des Schülers innerhalb der Leistungshierarchie einer Schulklasse.

Unter anderem auf Grund dieser Vieldeutigkeit der schulischen Zensuren haben Prognosen, die mit Hilfe dieses Instruments er-

¹ Vgl. den Beitrag von Peter FÜGLISTER in der «Schweizer Schule» 1970/3, S. 105 ff.

stellt worden sind, außerordentlich unterschiedliche Vorhersagekraft (vgl. INGENKAMP 1969, WEINGARDT 1969).

(8) Motivierung der Schüler

Schließlich wird vom System der schulischen Erfolgskontrolle erwartet, daß es den Schüler über seine Leistungen informiert und zu weiteren Lernbemühungen motiviert. Dabei impliziert das derzeitige Beurteilungssystem — etwas überspitzt ausgedrückt — folgende Motivationstheorie: Gute Schüler werden am besten durch Belohnung gefördert; schlechte Schüler am zweckmäßigsten durch Bestrafung (z. B. schlechte Zensuren).

2.2. Zur Revision der schulischen Erfolgskontrolle

Der Überblick über die unterschiedlichen Leistungen, die das schulische Kontrollsystem gleichzeitig erfüllen soll, dürfte zweierlei deutlich gemacht haben:

Erstens: Der einzelne Lehrer ist weit überfordert, wenn von ihm verlangt wird, daß er alle genannten Funktionen erfüllen soll. Es übersteigt die Fähigkeiten auch eines intelligenten und gut ausgebildeten Lehrers, *zugleich* die Leistungen der Schüler im kognitiven und affektiven Bereich sowie die Lernvoraussetzungen und Lernschwierigkeiten einzelner Schüler feststellen zu wollen und zusätzlich darauf zu achten, ob der Unterricht den Lehrzielen entspricht bzw. welche Lehrverfahren sich am besten bewähren.

Zweitens: Manche der genannten Funktionen des schulischen Systems der Erfolgskontrolle schließen sich gegenseitig aus. Dies gilt vor allem für die Funktion der Selektion der Schüler und die Funktion der Lernmotivierung. Die Teilung in gute und schlechte Schüler, wie sie das zweipolige Notensystem für die Extremgruppen bedeutet, dürfte bei der negativ bewerteten Gruppe verheerende motivationale Auswirkungen haben.

In neueren Vorschlägen zur Schul- und Unterrichtsorganisation wird deshalb darauf hingewiesen, daß das Prinzip der *«mittleren Schwierigkeitsdosierung»* zentrale Bedeutung für die Sicherung der Lernmotivierung besitzt. Darunter wird verstanden, daß die Schüler durch das Unterrichtsangebot jeweils *«leicht überfordert»* werden sollen, so daß die Lernmotivation erhalten bleibt und

das Lernziel mit großer Wahrscheinlichkeit erreicht werden kann. Voraussetzung für die Realisierung dieses Prinzips ist allerdings die *«Tüchtigkeitsgruppierung»* der Schüler in jenen Fächern, in denen Kenntnisse und Fähigkeiten für den Unterrichtserfolg eine entscheidende Rolle spielen (z. B. Mathematik, Fremdsprachen). Diese wenigen Hinweise genügen, um die Richtung anzuzeigen, in der eine Antwort auf die Frage nach den Chancen der Verbesserung des schulischen Kontrollsystems gefunden werden kann. Diese Lösung kann jedenfalls nicht im Einsatz eines Allheilmittels, z. B. der Anwendung von Schultests durch den Lehrer, gefunden werden — dazu sind die Funktionen des schulischen Prüfsystems viel zu komplex. Eine Verbesserung der schulischen Erfolgskontrolle in einem heute theoretisch schon vorstellbaren Ausmaß *darf vielmehr nur im Zusammenhang mit drei Entwicklungen erwartet werden:*

- dem fortschreitenden Prozeß der *Entdeckung und Präzisierung von Lernzielen;*
- der Verwirklichung neuer, wesentlich *flexiblerer Formen der Schülergruppierung*, einschließlich der Entwicklung von diagnostischen Verfahren zur Beurteilung der Schüler vor Beginn eines Lehrprozesses;
- die *Differenzierung der Lehrerrolle.*

Der Lehrberuf kann — wie unter anderem die Analyse der Kontrollaufgaben ergeben hat — als berufliche Position mit einer *ungemein vielfältigen und diffusen Aufgabenstellung* angesehen werden. Mit der gebührenden Vorsicht darf jedoch die Prognose gewagt werden, daß sich in absehbarer Zukunft zu einzelnen Aufgaben der traditionellen Lehrerrolle neue didaktische Berufsrollen ausgliedern werden (vgl. FLECHSIG 1969). Man kann z. B. annehmen, daß die geeigneten Meßverfahren zur Erhebung der Lernvoraussetzungen und der Lernfortschritte von Schülern in Hinkunft von Spezialisten entwickelt und dem einzelnen Lehrer zur Verfügung gestellt werden. Dasselbe gilt für die Bewertung der Wirksamkeit von Lehrmethoden. In Zukunft werden vermutlich objektivierte Lehrsysteme, deren jeweiliger Effekt unter bestimmten Bedingungen relativ genau bekannt ist, zur Verfügung stehen.

Desgleichen dürfte die Erhebung von Lernschwierigkeiten eine Angelegenheit von Spezialisten werden.

3. Vorschläge zu einer Verbesserung der schulischen Prüfungspraktiken

Der einzelne Lehrer ist tagtäglich gezwungen, schulische Erfolge zu kontrollieren. Er kann die Ausübung dieser Tätigkeit nicht so lange unterbrechen, bis die Erziehungswissenschaft die Voraussetzungen geschaffen hat, ein neues Schulsystem zu planen. Es sollen daher abschließend einige Vorschläge gemacht werden, die dem Lehrer hier und heute eine Verbesserung seiner Prüfungspraxis erlauben. Derartige Vorschläge bedürfen allerdings des Hintergrundes der vorangegangenen Analyse, um deutlich werden zu lassen, eine wie begrenzte Funktion ihnen im Hinblick auf den gesamten Komplex der Erfolgskontrolle zukommt. Die folgenden Empfehlungen sind keine Rezepte für eine optimale Beurteilungspraxis, sie sind vielmehr als Anstoß gedacht, über bestimmte Aspekte der schulischen Kontrollpraxis nachzudenken, bzw. sich weitere Anregungen aus der Fachliteratur zu holen.

Erste Empfehlung:

Die angestrebte Schülerleistung (das erwünschte Endverhalten) soll präzise beschrieben werden.

Die Kontrolle schulischen Erfolges kann um so besser durchgeführt werden, je genauer die *Ziele* schulischer Lehre beschrieben sind. Dazu ist es — wie erwähnt — notwendig, die Ziele des Unterrichts in *Kategorien beobachtbaren Schülerverhaltens zu beschreiben*. Den Vorgang der Formulierung der Zielvorstellungen als erwartete Schülerverhaltensweisen nennt man *«Operationalisierung der Lehrziele»* (vgl. MAGER 1969; vgl. MESSNER/POSCH 1969, 31 ff.).

1. möglichst präzise Beschreibung des erwünschten *Verhaltens*;
2. möglichst genaue Beschreibung der *situativen Bedingungen*, unter denen das Endverhalten realisiert werden soll;
3. Angabe eines *«Leistungskriteriums»* für die Erreichung des Zieles, z. B. des Prozentsatzes der richtig zu lösenden Aufgaben (vgl. MAGER 1969).

Als Beispiel könnte jedes beliebige Lehrziel herangezogen werden. Angenommen, den Schülern soll die *«Beherrschung der Silbentrennung»* «beigebracht» werden.

Man sieht sogleich: *«Beherrschung der Silbentrennung»* ist eine wenig präzise Beschreibung des angezielten Endverhaltens. Die Lehrpläne enthalten viele derartige Beschreibungen. Es wird dort in Zielangaben von *«Verständnis»*, *«Bekanntheit mit . . .»*, *«Bereitschaft . . . hineinzudenken»*, *«Sinn für Naturschutz»*, *«Einblick in . . .»*, *«Gefühl für . . .»*, *«Sicherheit im Zahlenrechnen»* usw. gesprochen. Diese Beispiele ließen sich beliebig fortsetzen. Für eine exakte Lehrzielbeschreibung müssen diesen Zielformulierungen Verhaltensbegriffe mit einem geringen Interpretationsspielraum zugeordnet werden. Dies geschieht am besten durch die Formulierung von *Aufgaben*, die jene Schüler lösen können sollen, die ein Lernziel erreicht haben. Das erwünschte Verhalten wird durch Ausdrücke wie *«aufschreiben»*, *«auswendig hersagen»*, *«aufzählen»*, *«zeichnen»*, *«Aufgaben lösen»* usw. beschrieben. Beim Lehrziel Silbentrennung wäre z. B. zu fordern, daß die Schüler eine größere Zahl von Wörtern, die schriftlich vorgegeben werden, durch senkrechte Striche richtig abteilen. Zweitens sind die situativen Bedingungen anzugeben. Man entscheidet sich z. B. dafür, daß

- 50 aus dem Wörterbuch durch Zufall ausgewählte
- schriftlich vorgegebene Wörter
- ohne Hilfsmittel (z. B. Liste der Trennungsregeln)
- in 10 Minuten
- bei normalen schulischen Arbeitsbedingungen
- so oft wie möglich abgeteilt werden sollen.

Zur Kontrolle der Aufgabenstellung kann sich der Konstrukteur dieser Aufgabe folgende vier Fragen vorlegen:

1. Ist festgelegt, was dem Lernenden zur Verfügung stehen soll?
2. Ist festgestellt, was ihm verweigert werden soll?
3. Sind die äußeren Bedingungen, unter denen das Endverhalten geäußert werden soll, angegeben?
4. Gibt es bestimmte Fähigkeiten, die man *nicht* entwickeln will? Sind diese Fähig-

keiten ausgeklammert? (vgl. MAGER 1969, 27).

Besonders wichtig ist die Angabe des *Leistungskriteriums*, also die Entscheidung, wieviele richtige Lösungen verlangt, wieviele Fehler toleriert werden sollen, welche Lösungen überhaupt als Fehler gelten sollen, ob alle richtigen Lösungen gleich gewichtet werden usw. Im gegebenen Fall werden z. B. alle Trennungsmöglichkeiten gleich bewertet sowie alle Trennungen, die falsch sind, und alle Auslassungen als Fehler gerechnet. Als positiv werden alle Arbeiten anerkannt, welche 75 % richtige Lösungen aufweisen.

Bei der Festlegung des Leistungskriteriums ist ein Prinzip zu beachten, das man als «Prinzip des optimalen Zielniveaus» charakterisieren könnte. Damit ist gemeint, daß es in vielen Fällen nicht nur aus Gründen der Ökonomie, sondern auch unter Rücksicht auf die affektiven und sozialen Lernziele, die im Unterricht verfolgt werden, unzumutbar sein kann, das intendierte Leistungsniveau allzu hoch anzusetzen. Die *Lernanstrengung* und der *Zeitaufwand*, die notwendig sind, um bei einem bestimmten Anteil von Schülern eine Leistung in der Silbentrennung von 95 und mehr Prozent richtigen Lösungen zu erreichen, führen vielleicht dazu, daß die Freude am Lernen (die Lernmotivation) oder die Selbststeuerung des Lernprozesses (die «intrinsische» Motivation) negativ beeinflußt werden¹.

Bei der Präzisierung der Lehrziele spitzt sich zugleich das Problem der inhaltlichen Relevanz von Lernzielen zu. Eine ungenaue Zielbeschreibung kann immerhin den Vorzug haben, daß «intuitiv» vom Lehrer solche Schülerleistungen angestrebt werden, die für die schulische Ausbildung «wesentlich» sind. Ein «operationalisiertes» Lehrziel hingegen legt eine *eindeutige Präferenz für spezifische Aspekte des Schülerverhaltens* fest. Die *Präzision* der Zielformulierung *allein* besagt noch nichts über die «Güte der Stellvertretung» (AEBLI), welche die ausgewählten Verhaltensweisen im Hinblick auf den weiten Bereich möglicher Verhaltens-

weisen, den sie repräsentieren, besitzen. Im Falle der Silbentrennung dürfte es z. B. wichtiger sein, daß die einzelnen Trennungen möglichst rasch und automatisiert vollzogen werden, während es keine besondere Rolle spielt, daß in jedem einzelnen Fall die Einsicht in die eben angewandte Trennungsregel aktiviert wird. Bei anderen Lernzielen dürfte die Operationalisierung nur dann als gelungen — oder «gültig», wie der Fachausdruck heißt — zu betrachten sein, wenn gesichert ist, daß ein bestimmtes Endverhalten als Ergebnis eines *Problemlöseprozesses*, nicht aber als automatisierte Reaktion auf einen Auslöser, gegeben werden kann. Um es an einem extremen Beispiel zu illustrieren: Die Lösung des Problems, welche Länge ein Zelt haben soll, das bei gegebener Breite und Höhe aus möglichst wenig Stoff gefertigt werden soll, kann vom Schüler A gegeben werden, weil er die Aufgabe verständig, d. h. unter situationsgerechter Anwendung der Regeln der Differentialrechnung löst; vom Schüler B aber deshalb, weil er den Rechenvorgang nach einem antrainierten Schema «mechanisch» durchführen konnte. Die beste Möglichkeit der Sicherung des einsichtsvollen Zustandekommens von Lösungen besteht darin, die situativen Bedingungen, auf die bestimmte Kenntnisse angewendet werden müssen, stark zu variieren, d. h. Anwendungsaufgaben zu stellen, die nicht schematisch gelöst werden können.

Zweite Empfehlung:

*Unabhängige Leistungsaspekte sollen isoliert erfaßt werden*¹.

Dieser Vorschlag, der mit der ersten Empfehlung in engem Zusammenhang steht, kann beispielsweise an der Prüfung der Fähigkeit, «einen Aufsatz zu schreiben», demonstriert werden. In der Regel wird zwar zwischen den Teilfertigkeiten

- fehlerfreie Schreibung
- Inhalt des Aufsatzes
- äußere Form

unterschieden. Tatsächlich sind noch eine Reihe weiterer Teilfunktionen unterscheidbar, vor allem hinsichtlich der Dimension «Inhalt»: Lebendigkeit der Schilderung, Bilderreichtum, Originalität der Darstellung, Flexibilität im Satzbau usw.

¹ Vgl. dazu das Prinzip der «optimalen Passung» bei HECKHAUSEN 1969, 212 f. Zum Thema «Lernmotivation» folgt ein ausführlicher Beitrag von Helmut MESSNER.

¹ Vgl. AEBLI 1967, 183 ff.

Am häufigsten dürfte gegen diese zweite Empfehlung bei Rechenaufgaben verstoßen werden. Hier vor allem empfiehlt es sich, die isolierte Prüfung abtrennbarer Leistungsaspekte dadurch zu ermöglichen, daß eine größere Anzahl wenig umfangreicher Aufgaben gestellt wird. Derartige Aufgaben vermitteln dem Lehrer Detailinformationen über die Lernfortschritte der einzelnen Schüler. Komplexe Aufgabenstellungen sind *nur dann zweckmäßig*, wenn man tatsächlich die Fähigkeit des Schülers, *eine umfangreiche Aufgabe zu lösen*, testen will.

Im Falle der «Silbentrennung» wurde der zweiten Empfehlung dadurch entsprochen, daß die Anwendung der Trennungsregeln an *vorgegebenen* Wörtern überprüft wurde, die lediglich durch senkrechte Striche abzutheilen waren. Auf diese Weise wird verhindert, daß bei der Prüfung der Fähigkeit «Silbentrennen» *zugleich* Rechtschreibkenntnisse eine Rolle spielen, wie dies etwa geschähe, wenn die zu trennenden Wörter diktiert würden.

Dritte Empfehlung:
Die Prüfungsbedingungen sollen objektiviert sein.

Auch dieser Vorschlag hängt mit den Empfehlungen 1 und 2 eng zusammen. Ihm wird entsprochen, wenn die Aufgaben schriftlich fixiert und nach vor der Prüfung festgelegten «Auswertungsregeln» (wann ist eine Lösung «teilrichtig»?) bewertet werden. Ziel dieser Bemühungen ist es, die Prüfungsmaßstäbe so präzise zu formulieren, daß die *beurteilenden Personen prinzipiell ausgetauscht* werden können, ohne daß sich die Beurteilung selbst ändert.

Um dies zu erreichen, wird bei der Konstruktion von Testaufgaben das Spektrum möglicher Antworten stark eingeschränkt bzw. überhaupt vorgegeben.

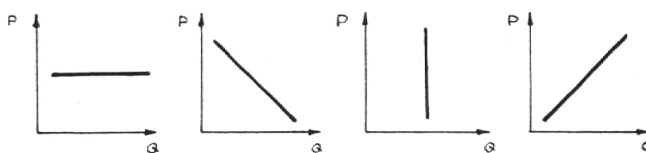
In der Testliteratur finden sich sehr anregende Aufgabenbeispiele:

1. Welche der folgenden Gleichungen ist *falsch*, wenn a und b zwei verschiedene (reelle) Zahlen sind?
 - A. $(a + b) + c = a + (b + c)$
 - B. $ab = ba$
 - C. $a + b = b + a$
 - D. $(ab)c = a(bc)$
 - E. $a - b = b - a$

(POSTLETHWAITE 1968, 122)

Erfaßt wird in diesem Fall die Fähigkeit, bestimmte Begriffe und Operationen zu verstehen (Gleichung, reelle Zahl, Symbole, Rechenoperationen) und zur Eliminierung falscher Gleichungen anzuwenden. Obwohl lediglich eine relativ einfache Handlung zu vollziehen ist — das Auswählen *eines* aus fünf Lösungsvorschlägen —, kann die richtige Wahl, falls die Lösung nicht schon bekannt ist, nur nach einem komplizierten Prozeß des Nachdenkens gefunden werden. Das bedeutet: die Kürze der gesuchten Antwort hat nichts mit der Komplexität des geforderten Lösungsprozesses zu tun.

2. Wenn die folgenden Begriffe in eine sinnvolle Reihenfolge gebracht werden, welcher steht dann in der Mitte?
 - a) Stamm, b) Baum, c) Wald, d) Zweig, e) Blatt.
3. Welche der folgenden Darstellungen stellt das Verhältnis zwischen Preis und Ware (Preis einer Ware, verfügbare Quantität der Ware) unter Wettbewerbsbedingungen am besten dar?



(nach BLOOM 1956, 103)

Vierte Empfehlung:
Aufgaben ohne Trennschärfe sollen ausgeschieden oder geändert werden.

Von gut gewählten Prüfungsaufgaben wird erwartet, daß mit ihrer Hilfe «gute» und «schlechte» Schüler unterschieden werden können. Wenn eine Aufgabe dies leisten kann, spricht man davon, daß sie «*trennscharf*» sei. Als Maß für die Trennschärfe einer Aufgabe wird die Übereinstimmung genommen, die bei einer Gruppe von Testpersonen zwischen dem Punktwert besteht, den sie bei *dieser* Aufgabe, und jenem, den sie im Test insgesamt erreichen. Man spricht dann von *hoher* Trennschärfe einer Aufgabe, wenn Personen, die im Test sehr hohe Punktzahlen erreichen, auch *diese* Aufgabe im allgemeinen richtig lösen bzw. wenn Personen, die im Test nur sehr niedrige Werte erhalten, auch *diese* Aufgabe in der Regel nicht lösen konnten. Von *niedriger* Trennschärfe spricht man dann, wenn die richtigen oder falschen Lösungen der

einzelnen Aufgabe dem Lösungstrend im gesamten Test zuwiderlaufen. In der Testpsychologie werden zur Bestimmung der Trennschärfe von Testaufgaben differenzierte statistische Verfahren angewandt. Für die Schulpraxis sei im folgenden ein weit größeres Verfahren beschrieben, das jedoch immerhin ein ungefähres Urteil über die Trennschärfe einer Aufgabe ermöglicht und es erlaubt, solche Aufgaben auszuscheiden, bei denen sonst gute Schüler versagen bzw. sonst schlechte Schüler in auffällender Weise richtige Lösungen haben. Es muß angenommen werden, daß derartige Aufgaben Merkmale enthalten, die gute Schüler in die Irre führen.

Angenommen, eine Klasse zählt 30 Schüler¹. Man teilt in diesem Fall für die grobe Bestimmung der Trennschärfe die Klasse in eine Obergruppe (jene 15 Schüler, die in der Prüfung insgesamt am besten waren) und eine Untergruppe (die 15 schlechteren Schüler). Nun kann für jede einzelne Aufgabe untersucht werden, wieviele Schüler der Obergruppe bzw. wieviele Schüler der Untergruppe sie richtig gelöst haben, und davon die *Differenz* gebildet werden, z. B.

Aufgabe 1: O — U = 13 — 6 = 7

Aufgabe 2: O — U = 10 — 12 = —2

Als Faustregel kann gelten, daß die Trennschärfe hinreichend groß ist, wenn die Differenz eine positive Zahl ergibt, die höher ist, als 10 % der Schülerzahl, im genannten Fall also höher als 3².

Bei Aufgabe 1 wäre demnach eine sehr hohe Trennschärfe erreicht, während Aufgabe 2 aus irgendwelchen Gründen, über die man nun nachdenken kann, sonst gute Schüler benachteiligt (vgl. CHAUNCEY/DOBBIN 1968, 150 f. vgl. auch BIGLMAIER 1969, 28).

Fünfte Empfehlung:

Zur Relativierung der klasseninternen Bewertungsmaßstäbe sollen gelegentlich Schulleistungstests verwendet werden.

¹ Bei ungeraden Schülerzahlen, etwa 31, wird jener Schüler, der in der Leistungsreihe an 16. Stelle steht, weggelassen.

² In diesem Fall erreicht der Korrelationskoeffizient, durch den der beschriebene Zusammenhang statistisch ausgedrückt werden kann, einen Wert von ungefähr 0,3. Dieser Wert wird als Mindestwert für trennscharfe Aufgaben betrachtet.

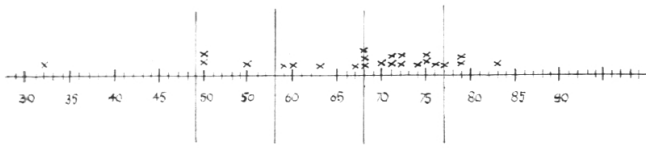
Die referierte Untersuchung von INGENKAMP hat eindrucksvoll gezeigt, daß der einzelne Lehrer wohl imstande sein dürfte, die Schüler ungefähr nach ihrer Leistung zu ordnen — mit allen durch subjektive Faktoren bzw. die Nichtstandardisierung der Prüfungsbedingungen bedingten Einschränkungen —, daß jedoch die für bestimmte Leistungen vergebenen Zensuren von Klasse zu Klasse stark schwanken. Eine Möglichkeit, dem abzuweichen, besteht darin, gelegentlich Schultests zur Revision der Benotungspraxis einzusetzen. Im folgenden wird von einem einfachen Verfahren berichtet, das von schwedischen Lehrern zu diesem Zweck angewandt wird.

Zweierlei ist jedoch zu bedenken. Erstens erfordert die Auswahl und Anwendung von Schultests Kenntnisse, die nur durch systematisches Studium erworben werden können. Zweitens kann sich das beschriebene Verfahren nur dann bewähren, wenn die Schultests vor allem solche Schülerverhaltensweisen prüfen, die *lernzielrelevant* sind. Dies ist jedoch gegenwärtig keineswegs immer der Fall. Überdies sind die Schultests nach der beschriebenen «Normalverteilungs-ideologie» konstruiert, d. h. so, daß von vornherein von einem Teil jeder Klasse schlechte Leistungen erwartet werden müssen. Diesem Trend entspricht auch die Streuung der Leistungsfähigkeit, die durch das Jahrgangsklassensystem bedingt ist. Der folgende Vorschlag stellt also lediglich eine partielle und sehr vorsichtig anzuwendende Korrekturmaßnahme dar.

In Schweden wird davon ausgegangen, daß in verschiedenen Klassen verschiedene Maßstäbe angewendet werden. Zur Korrektur der individuellen Notengebung werden standardisierte nationale Tests verwendet (vgl. HENRYSSON 1968, 77 ff.).

Ein Lehrer, der in einer Klasse mit 24 Schülern unterrichtet, gibt im Fach Deutsch beispielsweise nach Verrechnung seiner sämtlichen Informationen in einem Punktsystem folgende Zensuren:

Punkte	Zensur	Verteilung	
77—	1	4	
68—76	2	12	
59—67	3	4	
50—58	4	3	AM = 2,38
—49	5	1	
		N = 24	



Der standardisierte Deutsch-Test ergibt folgende Resultate:

Zensur	Rohwerte	Verteilung	
1	84—	4	
2	67—83	5	
3	46—66	9	AM = 2,75
4	30—45	5	
5	0—29	1	
		N = 24	

Auf Grund dieser Verteilung werden auf der Punktskala, die zur Ausgangsbewertung führte, die Notengrenzen neu festgelegt, und zwar so, daß sie einerseits die Häufungen auf der Punktskala, andererseits die Verteilung im standardisierten Test berücksichtigen. Es können etwa die Häufungen von Punkt 79—83 der Note 1, jene von 74—77 der Note 2 zugeordnet werden. Die weiteren Grenzen können bei 67 sowie bei 49 gezogen bleiben.

Es gibt somit folgende korrigierte Verteilung der Zensuren:

Zensur	Verteilung	
1	3	
2	5	Der neue Mittelwert von 2,88
3	9	zeigt die Annäherung an die
4	6	Verteilung im objektiven Test
5	1	an.

Es sei abschließend nochmals betont, daß die fünf Empfehlungen zur Verbesserung der schulischen Prüfungspraktiken nicht als Rezepte, sondern als gezielte Hinweise gedacht sind, über bestimmte Aspekte der schulischen Kontrollpraxis nachzudenken. Den auf Reformen bedachten Lehrer vermögen sie vielleicht dazu anregen, weitere Informationen aus der Fachliteratur zu entnehmen.

Das folgende *Literaturverzeichnis* nennt wichtige Fachliteratur zum Problem schulischer Erfolgskontrolle. Titel, die auf Grund ihrer *praktischen Bedeutung* und ihrer *Verständlichkeit* besonders zur Lektüre empfohlen werden, sind durch ein * gekennzeichnet.

Hans AEBLI (1966 a): Der Beitrag der Psychologie zur Gestaltung der Lehrpläne. In: Zeitschrift für Pädagogik. 6. Beiheft, 217—226.

Hans AEBLI (1966 b): Psychologische Didaktik. Stuttgart. 2. Auflage.

* Hans AEBLI (1967): Grundformen des Lehrens. Stuttgart. 4. Auflage.

Salomon E. ASCH (1951): Änderung und Verzerrung von Urteilen durch Gruppen-Druck. In: Martin IELE u. a. (Hrsg.): Texte aus der experimentellen Sozialpsychologie. Neuwied 1969, 57—73.

* Franz BIGLMAIER (1969): Leistungsmessung durch informelle Lehrtests (II. Teil). In: betriefft: erziehung 4 (1969), 26—28.

Benjamin BLOOM (Ed.) (1956): Taxonomy of Educational Objectives. Handbook I: Cognitive Domain. New York.

Wolfgang BREZINKA (1969): Über Absicht und Erfolg der Erziehung. Konstanzer Universitätsreden (22). Konstanz.

* Henry CHAUNCEY und John E. DOBBIN (1968): Der Test im modernen Bildungswesen. Stuttgart.
Henry DYER (1966): Die Entdeckung und Entwicklung von Bildungszielen. In: Proceedings of the 1966 Invitational Conference on Testing Problems, 12—24.

Karl-Heinz FLECHSIG (1969): Didaktik als Beruf. In: Konstanzer Blätter für Hochschulfragen. 25 (1969), 5—23.

Robert GLASER (1963): Instructional Technology and the Measurement of Learning Outcomes. In: American Psychologist, 519—521.

Heinz HECKHAUSEN (1969): Förderung der Lernmotivation und der intellektuellen Tüchtigkeiten. In: Heinrich ROTH (Hrsg.): Begabung und Lernen. Stuttgart, 193—228.

Sten HENRYSSON (1968): Pädagogische Testanwendung in Schweden. In: INGENKAMP und MARSOLEK 1968, 75—82.

Johann F. HERBART (1960): Allgemeine Pädagogik. Weinheim. 3. Auflage.

Elfriede HÖHN (1967): Der schlechte Schüler. München.

F. HOLZINGER (1955): Leistungssteigerung durch Leistungsmessung. Wien.

Karlheinz INGENKAMP (1968 a): Untersuchungen zur Übergangsauslese. Weinheim.

Karlheinz INGENKAMP (1968 a): Zur Problematik der Jahrgangsklasse. Weinheim.

* Karlheinz INGENKAMP (1969): Möglichkeiten und Grenzen des Lehrerurteils und der Schultests. In: ROTH a. a. O., 407—431.

Karlheinz INGENKAMP und Theresia MARSOLEK (Hrsg.) (1968): Möglichkeiten und Grenzen der Testanwendung in der Schule. Weinheim.

* Robert F. MAGER (1969): Lernziele und Programmierter Unterricht. Weinheim. 3. Auflage.

* Robert F. MAGER (1970): Motivation und Lernerfolg. In: betriefft: erziehung 3 (1970), 13—18.

Renate MAYNTZ u. a. (1969): Einführung in die Methoden der empirischen Soziologie. Köln.
 Rudolf MESSNER und Peter POSCH (1969): Lehrplan von morgen. In: IBB 1 (1969), 31—52.
 T. N. POSTLETHWAITE: IEA. Leistungsmessung in der Schule. Frankfurt 1968.
 Heinrich ROTH (Hrsg.): Begabung und Lernen. Stuttgart 1969.
 Helmut SCHELSKY (1957): Schule und Erziehung in der industriellen Gesellschaft. Würzburg.
 Emil SCHMALOHR (1961): Psychologie des Erstlese- und Schreibunterrichts. München.
 Richard STROHAL (1961): Bemerkungen zum Begriff der psychischen Disposition und seiner Bedeutung für die Pädagogik. In: Wolfgang BRE-

ZINKA (Hrsg.): Weltweite Erziehung. Freiburg, 251—263.

Ralph W. TYLER (1949): Achievement Testing and Curriculum Construction. In: G. WILLIAMSON (Ed.): Trends in Student Personnel Work. Minneapolis.

Erich WEINGARDT (1969): Der Voraussagewert des Reifezeugnisses für wissenschaftliche Prüfungen. In: ROTH a. a. O., 433—447.

Rudolf WEISS (1965): Zensur und Zeugnis. Linz.
 * Rudolf WEISS (1968): Leistungsbeurteilung durch Ziffernnoten. Linz.

Maria ZILLIG (1928): Einstellung und Aussage. In: Zeitschrift für Psychologie. 106 (1928), 78 ff.

Anwendung von Testverfahren in der Schule

Hinweise auf Publikationen von Karlheinz INGENKAMP

Psychologische Tests für die Hand des Lehrers?¹

«Häufig kann man unter Lehrern eine eigenartig zwiespältige Haltung gegenüber Tests beobachten. Wenn der Psychologe seine Beurteilungen und Prognosen auf Testergebnisse stützt oder die Heranziehung von Tests empfiehlt, dann begegnet er oft einer verbreiteten Ablehnung dieser ‚mechanistischen Stichproben‘, dieser ‚bloß reaktiven und vom Zufall abhängigen Seelendurchleuchtungen‘, dieses ‚der Eigenart des Seelischen nicht angemessenen Versuchs einer zahlenmäßigen Erfassung‘. Die in dieser Einseitigkeit unsachliche Einstellung schlägt dann häufig ins Gegenteil um, sobald die Kollegen Tests in die Hand bekommen und damit Versuche anstellen. Vor allem Persönlichkeitstests erfreuen sich dann großer Beliebtheit, man ist erstaunt, was ‚alles herauskommt‘, findet die Handhabung nicht schwierig, man hat vielleicht auch das Gefühl, Verborgenes ergründen zu können.» (S. 21)

Die Haltung, die Karlheinz INGENKAMP, ehemaliger Volks- und Mittelschullehrer, Diplom-Psychologe und derzeitiger Dozent für Erziehungswissenschaft in Berlin, bei seinen deutschen Kollegen bemerkte, kann man im Zusammenhang mit der Testanwendung in der Schule auch in unseren Lehrerkreisen feststellen: *ablehnende Skepsis* auf der

einen, *überspannte Hoffnung* auf der andern Seite.

Ist der psychologische Test lediglich ein Instrument für den *Fachpsychologen*, oder kann er ein brauchbares diagnostisches Hilfsmittel auch für den *Lehrer* sein?

Der Autor beantwortet die Frage differenziert: Es gibt Testverfahren (z. B. projektive Tests), die gehören *nicht* in die Hand des Lehrers; mit andern Tests dagegen sollte auch der Lehrer soweit vertraut sein, daß er sie, wo es die pädagogische Situation erfordert, richtig anwenden kann. Damit ist die Frage zunächst nur verschoben, denn es muß nun gesagt werden, in welcher Situation welcher Test unter was für Umständen und Rücksichten eingesetzt werden kann. INGENKAMP gibt darauf präzise Antworten. Nach einer *Einführung in die methodischen Kriterien zur Beurteilung* von Tests, eröffnet er einen *Überblick über alle Verfahren*, die heute dem Lehrer als *Tests* oder testähnliche Prüfungen angeboten werden. Ferner bezieht er Stellung zur Frage der *Eignung oder Nichteignung* dieser Verfahren für die Anwendung durch den Lehrer in der Schule. Hinweise erläutern die *sachgerechte Anwendung* der Testverfahren, und schließlich enthält die Publikation noch *Informationen über Bezugsquellen und Preise des Testmaterials*. Also doch eine vorbehaltlose Propaganda für den «Großeinsatz» psychologischer Tests in