

Sonderdruck aus

# **Abitur-Normen gefährden die Schule**

*Herausgegeben von Andreas Flitner  
und Dieter Lenzen*

R. Piper & Co. Verlag  
München 1977

Hans Brügelmann

## Einheitlichkeit durch Operationalisierung – Ein Phantom

*Zum Lernzieldilemma in den Normenbüchern*

### *I. »Vereinheitlichung« – wozu und wie?*

Im folgenden sollen – entsprechend der inhaltlichen Arbeitsteilung in diesem Band – die *tatsächliche Erreichbarkeit* einheitlicher Prüfungsanforderungen, ihre *theoretischen Voraussetzungen* und möglichen *Nebenwirkungen* erörtert werden. Daß ich mich dabei zunächst auf die *impliziten Ziele* und Annahmen der Normenbücher einlasse, darf also nicht als bildungspolitische und verfassungsrechtliche Unbedenklichkeitserklärung für das »Einheitlichkeits«-Postulat ausgelegt werden; ich verweise dazu auf die Beiträge von H. v. Hentig, W. Perschel und H. Rumpf.

Die KMK schlägt vor, Anforderungen und Bewertungsmaßstäbe in der Abiturprüfung zu vereinheitlichen, damit die begrenzten Studienplätze gerechter und auf durchsichtiger Weise verteilt werden können. Zu diesem Zweck sollen in den Normenbüchern länderübergreifend die *Prüfungsverfahren*, die *Prüfungsinhalte* und die *Bewertungskriterien* verbindlich festgelegt werden – u. a. »durch die Beschreibung der vom Schüler erwarteten Kenntnisse, Fähigkeiten und Fertigkeiten in einem Fach« (KMK 1975, S. 3).

Wieweit es *möglich* und in welchem Maße es überhaupt *sinnvoll* ist, Lernziele auf einer zentralen Ebene in spezifizierter Form festzulegen, ist in der Erziehungswissenschaft und insbesondere bei der Reform von Richtlinien und Curricula seit etwa 1970 ausführlich diskutiert worden. Die »Operationalisierung«, d. h. möglichst präzise Beschreibung von Lernzielen sollte zwei Funktionen erfüllen: Man hoffte, Mißverständnisse bei der Verbreitung von Richtlinien und Unterrichtsvorschlägen ausschalten, also die *Verständlichkeit* der Ziele erhöhen zu können; andererseits versprach eine detaillierte Festlegung von

Lernzielen eine verbesserte *Kontrollierbarkeit* des Unterrichts und seines Erfolgs (vgl. zusammenfassend: H. Meyer 1974).

Bei dem Versuch, die didaktischen Absichten durchgehend zu operationalisieren, geriet man jedoch bald in ein Dilemma, das zutreffend als »curriculare Unschärferelation« (H. Bauersfeld) bezeichnet wurde (s. ausführlicher unter III): Entweder kann man die inhaltliche Einbettung eines Lernziels, d. h. seine Bedeutung innerhalb eines Begründungsrahmens verdeutlichen *oder* einzelne seiner Elemente präzise erfassen – aber nie beides gleichzeitig. Zusätzlich stellt sich ein Problem, das Juristen unter dem Stichwort »Einzelfallgerechtigkeit *oder* Rechtssicherheit« seit Urzeiten in Konflikte bringt: Will man nämlich die Möglichkeit erhöhen, verschiedene Fälle entsprechend ihren jeweils besonderen Bedingungen zu beurteilen (*materielle* Gerechtigkeit), so benötigt man ein *flexibles* System *auslegungsfähiger* Kriterien; will man andererseits die Berechenbarkeit und Kontrollierbarkeit des Bewertungsverfahrens steigern (*formelle* Gerechtigkeit), muß man Kriterien *vorweg* möglichst *eindeutig* und *starr* festlegen.

In den veröffentlichten Beschlüssen der KMK wird zu diesem Problem nirgendwo Stellung bezogen, so daß unklar bleibt, was es mit der angezielten »Einheitlichkeit« denn nun genau auf sich hat. Es besteht insofern auch die Gefahr, daß sich die Auftraggeber bzw. die Autoren der Normenbücher unterschiedlich rechtfertigen können, je nachdem, ob man den *Anspruch* der Vereinheitlichung *oder* seine *Realisierung* in den vorliegenden Heften kritisiert. Nach dem Motto: Erstens habe man den Krug noch nie gesehen, zweitens sei er doch längst zurückgegeben worden und drittens könne man nun wirklich nichts dafür, daß er heruntergefallen sei, tauchen in den Normenbüchern z. B. folgende Argumente auf (vgl. KMK 1975, S. 3 und 5 und die fachspezifischen Einführungen zu den einzelnen Normenbüchern):

- (1) im Grunde seien die Maßstäbe doch recht *flexibel* gehalten und auf die Anforderungen der einzelnen Fächer abgestimmt; *oder* aber:
- (2) eine präzise Festlegung der Anforderungen sei unumgänglich, wenn mit der Chancengleichheit beim Hochschulzugang Ernst gemacht werden solle; *oder* schließlich:

(3) durch die Vorgaben werde natürlich weder die Lehrplanhoheit der Länder noch die Möglichkeit zu Curriculumreformen in Frage gestellt.

Ich will damit nicht unterstellen, die *Autoren* der Normenbücher hätten sich ihrer Aufgabe leichtfertig *oder* in Unkenntnis der Probleme und möglicher Folgewirkungen entledigt (vgl. etwa die selbstkritischen Hinweise a. a. O., S. 7). Sie haben aus einer von Schule und Hochschule nicht verschuldeten Notlage das Beste zu machen versucht: Mehr Studienplätze können *sie* nicht schaffen, und das gegenwärtige Zulassungsverfahren ist in der Tat so bedrückend, daß jeder Versuch begrüßt werden muß, die Auswahlkriterien aussagekräftiger und überprüfbarer zu machen.

Dieses Verständnis für die Situation der Fachkommissionen darf aber nicht dazu führen, daß vordergründig einleuchtende Ideen wie die »Vergleichbarkeits«-Forderung pauschal übernommen und zur Rechtfertigung problematischer Maßnahmen herangezogen werden. Der begrenzte Raum und die bereits angedeutete Unterschiedlichkeit der einzelnen Normenbücher zwingen dabei zur Zuspitzung der Argumentation auf drei Grundthesen, die ich anschließend etwas ausführlicher erläutern möchte:

#### *These 1*

*Soweit* die Normenbücher beanspruchen, Lernziele als Kriterien für die Beurteilung des Lernerfolgs zu operationalisieren, werden sie ihrem *eigenen Anspruch nicht gerecht* (s. unten II).

#### *These 2*

Die inhaltliche *und* prozedurale Vergleichbarkeit von Abiturprüfungen läßt sich aus *grundsätzlichen Erwägungen* weder durch zentral operationalisierte Lernziele noch durch andere testorientierte Maßnahmen erreichen (s. unten III).

#### *These 3*

*Soweit* sich Lernziele sinnvoll operationalisieren lassen bzw. soweit ein genereller Einheitlichkeitsanspruch aufrechterhalten wird, führen detaillierte Lernzielkataloge zu *Nebenwirkungen im Unterricht*, die eine Realisierung des Vergleichbarkeits-Postulats zusätzlich in Frage stellen (s. unten IV).

## II. Wie einheitlich sind die Prüfungsanforderungen in den Normenbüchern tatsächlich?

In den Notendurchschnitt des Abiturzeugnisses gehen Urteile ein,

- a) die auf verschiedenen *Niveaus* vergeben werden (Grund- und Leistungskurse),
- b) die innerhalb der Schullaufbahn zu verschiedenen *Zeitpunkten* gefällt werden (Abschlußnoten der Klassen 11 und 13, Noten aus der Abiturprüfung im engeren Sinn),
- c) die sich auf verschiedene *Fächer* mit unterschiedlicher Gewichtung im Lehrplan beziehen (Deutsch, Biologie, Sport usw.),
- d) die von verschiedenen *Prüfern* vergeben werden (implizite persönliche, institutionelle und regionale Normen).

Macht man mit der Vereinheitlichung Ernst, wäre eine Abstimmung der Prüfungsanforderungen unter allen vier Gesichtspunkten erforderlich. Dies ist in den Normenbüchern nicht geschehen.

Die Verrechnung unterschiedlicher *Kursniveaus* wird in den Normenbüchern z. T. ausführlich und in *Beispielaufgaben* konkretisiert dargestellt (vgl. etwa Gemeinschaftskunde, S. 25 bis 63). Andere Fachkommissionen staffeln ihre *Lernzielkataloge* durch die Hinzufügung von zusätzlichen Anforderungen für das Leistungsfach (z. B. Physik, S. 10–13), während etwa im Normenbuch Englisch (S. 16f.) nur *Kriterien* für die Unterscheidung der Anforderungen genannt werden. Trotz der erheblichen formalen Unterschiede, die im fächerübergreifenden Vergleich sichtbar werden, bieten wenigstens die ersten beiden Modelle dem Prüfer konkrete Anhaltspunkte für die Einschätzung von Leistungen aus verschiedenen Kurstypen.

Verbindlich einbezogen werden in den Normenrahmen nur die Noten, die in der schriftlichen oder mündlichen Abiturprüfung vergeben werden. Endnoten der Oberstufe lassen sich ohne allzu großen Aufwand analog bestimmen, während für die »mitgenommenen« Noten der in Klasse 11 abgelegten Fächer eine parallele Formulierung zusätzlicher Anforderungsniveaus und Bewertungskriterien erforderlich wäre. Die *Zeitverschiebung* der Notenvergabe ist insofern nicht zureichend berück-

sichtigt worden.

In *inhaltlicher* Hinsicht ist eine Abstimmung zwischen den Anforderungen in verschiedenen Fächern erst gar nicht versucht worden. Dies mag grundsätzlich unmöglich sein, bedeutet allerdings eine erhebliche Einbuße schon für die Glaubwürdigkeit des Vergleichbarkeitsanspruchs. Aber auch in *formaler* Hinsicht hält die KMK in ihrem Beschluß vom 6. 2. 1975 »fachspezifische Unterschiede für vertretbar, wenn die Anwendung der Lernzielstufen und Bewertungskriterien in den Fächern (Hervorhebung von mir, H. B.) als überzeugend und verständlich gelten kann« (KMK 1975, S. 3). S. Reichert u. a. (1976, S. 4) stellen als Ergebnis ihrer synoptischen Analyse der Normenbücher denn auch fest, »daß die Fachkommissionen die gleiche Aufgabenstellung sehr unterschiedlich behandelt haben«. Im wesentlichen beschränkt sich die Übereinstimmung auf so globale Kategorien wie »Wissen«, »Anwenden« und »Urteilen« (KMK 1975, S. 6) oder »Reproduktion«, »Reorganisation«, »Transfer« und »Problemlösendes Denken« (Musik, Biologie u. a., allerdings z. T. schon mit Variationen auf dieser Ebene). Dies ist schon deshalb unbefriedigend, weil z. B. einfache Transferleistungen durchaus niedriger eingestuft werden könnten als anspruchsvolle Reproduktionsleistungen und weil dieselbe Aufgabe – je nach individueller Vorerfahrung – unterschiedliche Leistungsstufen dieser Typologie anfordert. Mehr noch: Weniger die Schaffung einheitlicher Etiketten scheint das Problem zu sein als vielmehr ihre *gleichartige Verwendung* in unterschiedlichen inhaltlichen Zusammenhängen. Gerade die Verknüpfung von *Verhaltens-* und *Inhaltsaspekt* in der Lernzielformulierung wird aber sehr unterschiedlich gehandhabt, denn »manchmal liegt der Akzent bei den Formulierungen auf den formellen Verhaltensbeschreibungen; manchmal wird darauf weniger Wert gelegt, und die Auflistung präziser Inhaltsangaben scheint der Hauptzweck eines Zielkatalogs zu sein. Zuweilen wird mit dem Titel eines Katalogs Fachbezogenheit angekündigt, die Definition der Ziele beinhaltet dann aber kaum fachliche Inhaltsangaben (z. B. Chemie)« (Reichert u. a. 1976, S. 28). Auch in der Darstellung werden Inhalts- und Verhaltensbeschreibung z. T. getrennt (Biologie, Chemie, Evangelische Religion), z. T. verknüpft (z. B. Mathematik s.

den Überblick a. a. O., S. 12). Es zeigen sich mithin auch in der Beschreibungsform erhebliche Unterschiede, die eine fächerübergreifende Vereinheitlichung in Frage stellen.

Schließlich zur *fachinternen* Vereinheitlichung, die auf eine Einengung des Bewertungsspielraums verschiedener Prüfer zielt. Eine Operationalisierung der Lernziele setzt selbst in einem weiteren Verständnis eine möglichst eindeutige Beschreibung des angestrebten Schülerverhaltens voraus (nach dem engeren sozialwissenschaftlichen Gebrauch müßte auch eine Meßvorschrift für die Feststellung des Lernerfolgs vorgegeben werden). Gemeinhin wird eine *inhaltlich* gefüllte Beschreibung eines *beobachtbaren Schülerverhaltens* unter bestimmten *Bedingungen* verlangt, das sich nach einem *vorgegebenen Maßstab* bewerten läßt (vgl. ausführlicher: H. Meyer 1974, S. 55ff.). Je weiter diese Bedingungen eingehalten werden, um so vergleichbarer die Noten, ließe sich zunächst annehmen (s. aber unten III). Wie sehen die Normenbücher nun in dieser Hinsicht aus?

Einige Fachkommissionen (z. B. Englisch und Französisch) haben sich auf die Beschreibung allgemeiner Gesichtspunkte zurückgezogen, da fachdidaktische Kontroversen einer Festlegung von Details entgegenstünden (jeweils S. 9; die differenzierten Zahlenbeispiele [Englisch, S. 20ff.] dürfen nicht darüber hinwegtäuschen, daß es sich nur um die Verrechnung globaler Leistungsbereiche handelt). In ähnlicher Weise verweigern die Normenbücher für Griechisch und Latein eine fachspezifische Auslegung von Anforderungen, die in der jetzigen Form allenfalls Perspektiven für einen fremdsprachlichen Unterricht gleich welcher Art darstellen. Demgegenüber legen sich die Normenbücher für Deutsch (ziemlich einseitig, vgl. D. Lenzen 1976, S. 302) und für Physik bzw. Mathematik auch in inhaltlicher Hinsicht fest. Mit z. T. sehr detaillierten Ziel- und/oder Inhaltskatalogen wird im Grunde die Lehrplanhoheit der Bundesländer außer Kraft gesetzt – oder die Vergleichbarkeit der Prüfungsanforderungen insofern zur Fiktion, als die Fassung zwischen Normenbuch und geltendem Lehrplan von Land zu Land unterschiedlich ausfällt. Auch die Beschwörung, man wolle »trotzdem genügend Freiheit . . . gewährleisten« (Physik, S. 10), wird kaum verhindern können, daß sich der Unter-

richt de facto auf die Kernanforderungen beschränkt.

Selbst diese *didaktisch* schon sehr einengenden Vorgaben sind aber für eine *diagnostische* Verwendung noch nicht zureichend operationalisiert, um verhindern zu können, daß verschiedene Prüfer zu unterschiedlichen Urteilen über dieselbe Schülerleistung kommen. Wenn etwa in Physik »Vertrautheit mit der Beschreibung einer geradlinigen Bewegung (Weg-Zeit-Funktion; Geschwindigkeit und Beschleunigung und ihre Ableitungen)« (S. 10) gefordert wird, wenn darüber hinaus vier Lernzielstufen (S. 13ff.) und zwei Schwierigkeitsgrade (S. 16) angeboten werden, so beginnt der Streit doch gerade an dem Punkt, wo nun eine bestimmte *Lösung* bezogen auf die Formulierung der *Aufgabe* (tatsächliches Experiment, Versuchsskizze, Entwicklung oder Anwendung einer Formel usw.) und bezogen auf den vorhergegangenen *Unterricht* (und evtl. andere Vorerfahrungen) des Schülers einzustufen ist.

Um nicht mißverstanden zu werden: Mir scheint eine solche (wenn nicht sogar weitergehende) Flexibilität der Aufgabenstellung und -bewertung sinnvoll, und ich halte auch viele der angegebenen Kriterien für eine praktische Hilfe; nur sollte man eine solche Orientierungs- und Ordnungshilfe nicht mit dem Anspruch einer verbindlichen Vereinheitlichung der Prüfungsanforderungen und -maßstäbe belasten.

Selbst für die *fachbezogene* Vereinheitlichung ist also vorläufig festzuhalten, daß entweder dieser Anspruch stark reduziert wird oder daß der zweite Anspruch, »die Prüfungsanforderungen zu vereinheitlichen, ohne gleichzeitig curriculare Angaben in Form eines Lehrplans festzuschreiben« (KMK 1975, S. 3), zumindest faktisch unterlaufen wurde. *Systemimmanent* gedacht muß man also entweder auf eine Festlegung detaillierter Lernziele verzichten und auslegungsfähige Kriterien höherer Ordnung entwickeln – deren Auslegung dann im Rahmen unterschiedlicher Richtlinien *nachträglich institutionell moderiert* werden müßte (dieses System erfreut sich im englischen »Certificate of Secondary Education [Mode 3]« einer gewissen Beliebtheit); oder man gibt genaue Lernziele vor, die zwar nicht rechtlich, aber de facto den Unterricht steuern – womit die Lehrplanhoheit der Länder nurmehr auf dem Papier steht (dieser Effekt läßt sich – ebenfalls in England – am Beispiel des

»General Certificate of Education« beobachten).

Nach dieser Zwischenbilanz stellt sich aber die grundsätzlichere Frage, unter welchen Voraussetzungen bzw. in welchem Maße eine zentrale Operationalisierung tatsächlich zu vergleichbaren Prüfungsanforderungen führen kann.

### III. Wie einheitlich können zentrale Prüfungsanforderungen grundsätzlich sein?

Bevor man untersucht, wie sich die Bewertung von Prüfungsleistung vergleichbarer gestalten läßt, ist zu fragen, was die Prüfungsergebnisse überhaupt aussagen sollen. Zwei Begründungsstränge sind denkbar:

(1) Wenn Abiturnoten eine *Vorhersage* des voraussichtlichen Berufs- oder wenigstens Studienerfolgs erlauben, könnten sie unter dem Gesichtspunkt einer möglichst ökonomischen Verwendung der öffentlichen Haushaltsmittel die Auswahl oder Zurückweisung von Studienbewerbern rechtfertigen. Drei Klippen müßten jedoch umschifft werden, sollte diese Rechtfertigung standhalten:

– Wie kann man herausbekommen, vor welchen *tatsächlichen Anforderungen* die heutigen Abiturienten in 10, 20 oder 30 Jahren stehen werden, wie also die angestrebten Berufe in Zukunft aussehen werden?

– Was macht – ein einigermaßen zutreffendes Berufsbild einmal vorausgesetzt – eigentlich einen »guten« Lehrer, Kaufmann, Techniker aus, d. h. wie ist Einigung über die *Kriterien* für den Berufs- oder wenigstens Studienerfolg zu erzielen?

– Einigkeit über das Erfolgskriterium vorausgesetzt – läßt sich überhaupt ein stichhaltiger Zusammenhang zwischen Abiturnoten und dem Erfolgskriterium herstellen, d. h. aufgrund welcher Erfahrungen oder theoretischer Konstrukte können Abiturnoten als *aussagekräftige Indikatoren* für die Berufs- oder Studienprognose gelten?

Nicht nur für herkömmliche Benotungsverfahren, sondern auch für neu zu konstruierende Tests werfen diese Fragen so erhebliche Probleme auf (vgl. D. Hopf 1974, O. Hopf 1976, K. Ingenkamp 1971 und J. Ziegenspeck 1975), daß diese Begrün-

dungsmöglichkeit m. E. ausfällt (vgl. ausführlicher zu diesem Problem den Beitrag von D. Schmied).

(2) Bleibt also der zweite Weg über die Behauptung, Abiturnoten dienen zur *Feststellung des bisherigen Lernerfolgs*, so daß sich mit ihrer Hilfe die Vergabe von Studienplätzen wenigstens als eine Art individueller Belohnung rechtfertigen ließe (implizit hofft man dabei meist wohl doch, daß gute Prüfungsleistungen nicht nur auf eine erfolgreiche Schullaufbahn, sondern auch auf grundlegendere Fähigkeiten und Einstellungen schließen ließen, die im Leben allgemein und im Studium speziell Erfolg versprechen).

Die Aussagekraft der Abiturnoten in diesem Begründungsmodell hängt zunächst davon ab, wie weit die Prüfungsmaßstäbe mit den Unterrichtsabsichten und -inhalten übereinstimmen. Operationalisierte Lernziele bieten sich in dieser Hinsicht als Brückenglieder an, weil sie sowohl auf die *Steuerung* didaktischen Handelns zielen als auch Maßstäbe zur *Überprüfung* seiner Wirksamkeit anbieten. Diese Attraktivität des Instruments leidet aber unter folgenden Tücken:

(a) Lernziele stellen immer nur Kürzel für umfangreichere Begründungszusammenhänge dar. Der vermeintliche Gewinn an Eindeutigkeit durch eine ausdifferenzierende Operationalisierung in Form einzelner Schülerverhaltensweisen wird erkaufte mit einem Verlust an inhaltlicher Einbettung. D. h. ohne Einigkeit über den fachwissenschaftlichen, stufendidaktischen und bildungspolitischen Bezugsrahmen ist eine Einigung auf atomisierte Verhaltensbeschreibungen wertlos, da inhaltlich mehrdeutig. Die neusprachlichen Normenbücher haben aus der Einsicht in die Umstrittenheit fachdidaktischer Konzeptionen zwischen den Richtlinien und in der Unterrichtspraxis die Konsequenz gezogen und sich auf die Beschreibung allgemeiner Gesichtspunkte und formaler Kriterien beschränkt; die Fachkommission Deutsch ist demgegenüber expansiv geworden und hat mit ihren Vorgaben die Richtlinien einiger Bundesländer in Frage gestellt. In anderen Normenbüchern muß der inhaltliche Kontext im Interpretationsrahmen der durchaus unterschiedlichen Richtlinien *nachträglich* wiederhergestellt werden. Dies gilt auch in den Fällen, wo Inhalte zwar benannt werden, aber allenfalls in Stichwortform oder auf der Ebene

allgemeiner Konzepte (vgl. etwa Gemeinschaftskunde). Fazit: *Prüfungshoheit* und *Lehrplanhoheit* können zwar juristisch getrennt werden (s. KMK 1975, S. 5), in der Praxis werden aber Konkurrenzen auftreten, die entweder die regionale Einheitlichkeit der *Prüfungsanforderungen* oder die *Passung* von Prüfungsmaßstab und Unterricht oder die faktische Geltung der *Richtlinien* in einzelnen Ländern aufheben.

(b) Vorausgesetzt, daß für dieses Problem eine politische Lösung gefunden werden kann, bleibt zu klären, ob durch operationalisierte Lernziele Aufschluß über Lernerfolg und Lernfähigkeit von Schülern gewonnen werden kann. Meine Ausgangsthese dazu (s. S. 2) lautete: Ein Plus an Berechenbarkeit und Kontrollierbarkeit von Prüfungsmaßstäben wird erkaufte mit einem Verlust an Angemessenheit im Einzelfall und an Aussagekraft über die Lösung der jeweiligen Prüfungsaufgabe hinaus. Zur Begründung (vgl. ausführlicher H. Brügelmann 1972):

(ba) Jede Bewertung des Lernerfolgs eines Schülers ist auch eine Bewertung des jeweiligen Lernangebots, der Fähigkeit des Lehrers, des Niveaus der Lerngruppe und der Lernbedingungen. Je einheitlicher Prüfungsanforderungen formuliert und je weiter sie vom spezifischen *Unterrichtskontext* abgehoben werden, um so geringer ist die Aussagekraft der Abiturnoten für die tatsächliche Leistungsfähigkeit des einzelnen Schülers. Je mehr andererseits Rücksicht darauf genommen werden soll, daß dieselbe Aufgabe-auf-dem-Papier de facto für verschiedene Schüler eine unterschiedliche Anforderung bedeutet, desto auslegungsfähiger müssen die Prüfungskriterien gehalten werden (mit einem entsprechenden Verlust an Kontrollierbarkeit). In diesem Dilemma werden die Grenzen einer *zentralen* Operationalisierung von Prüfungsmaßstäben offenkundig.

(bb) Stellt man weniger auf den Vergleich zwischen verschiedenen Schülern, sondern zunächst allein auf die Einheitlichkeit des Urteils zwischen verschiedenen Prüfern ab, so stellen sich zusätzliche Probleme. Noten werden um so einheitlicher ausfallen, je mehr sich die Prüfungskriterien auf *Oberflächenbeschreibungen* des Schülerverhaltens beschränken – die aber ihrerseits wiederum mehrdeutig sind, wenn man aus einer festgestellten Leistung (Aufgabenlösung) auf die zugrunde liegen-

de Fähigkeit schließen will. Salopp formuliert: Wenn zwei dasselbe tun, ist es nicht immer dasselbe; kreuzen z. B. mehrere Schüler in einem Test dieselbe Auswahlantwort an, so kann es sich beim ersten Schüler um die bloße Wiederholung einer auswendig gelernten Lösung handeln; beim zweiten um einen Transfer von einer ähnlichen, früher bearbeiteten Aufgabe her; beim dritten um die Aktualisierung und selbständige Verknüpfung allgemeiner Kenntnisse und Fähigkeiten, während der vierte mehr oder weniger intelligent rät. D. h. aber, daß die Eindeutigkeit der Aufgaben-auf-dem-Papier die faktische Mehrdeutigkeit der Lösung nur verdeckt: Die für das Urteil über die *Fähigkeit* des Prüflings, dieselbe Aufgabe auch zu einem anderen Zeitpunkt oder ähnliche Aufgaben zum jetzigen Zeitpunkt lösen zu können, die für eine solche Verallgemeinerung eigentlich interessante *Tiefenstruktur* muß immer interpretativ erschlossen werden. In ein solches Urteil aber gehen Annahmen über den vorangegangenen Lernprozeß und über schwer objektivierbare Aspekte der Prüfungssituation bzw. des Schülerverhaltens ein, die einen *Deutungsspielraum* voraussetzen, der notwendigerweise die Kontrollierbarkeit von Urteilen wieder einschränkt.

(bc) Leistungen lassen sich um so leichter abprüfen und in eine Rangfolge bringen, je weiter man sie in einzelne *Komponenten* aufschlüsselt, denen dann jeweils genau umschriebene Aufgaben zugeordnet werden können. Auf diese Weise – so die Hoffnung – kann man dann auch komplexere Fähigkeiten prüfungstechnisch in den Griff bekommen. Daß dabei die Verknüpfung von Teilfähigkeiten, ihre jeweilige Veränderung durch das Zusammenspiel mit anderen Komponenten und erst recht die übergreifenden Koordinationsleistungen nicht erfaßt werden können, dürfte nur für ein statisches Modell kompakter Qualifikationsbausteine belanglos sein (daß selbst elementare Fähigkeiten wie »Lesen« überaus komplex und dynamisch sind, habe ich an anderer Stelle gezeigt; (vgl. H. Brügelmann/D. Fischer 1977). Der *qualitative* Gehalt von Kenntnissen, Einstellungen und Fähigkeiten, seine Veränderung durch neue Erfahrungen oder in der Ausrichtung auf unterschiedliche Aufgaben kann – soweit überhaupt – nur in *komplexen* Aufgaben und über einen *längeren Zeitraum* mit entsprechend *vielfälti-*

ger und intensiver Beobachtungsmöglichkeit erfaßt werden. Für solche Einschätzungen lassen sich zwar Beispiele und die Aufmerksamkeit fokussierende Kriterien, aber keine *verbindlichen* Auswertungsschlüssel vorgeben. Dies unterstreicht, daß Prüfungsvorschläge in hohem Maße nur *Anregungscharakter* als Hilfe zur Organisation der diagnostischen Wahrnehmung beanspruchen können.

(bd) Nicht allein die Komplexität, sondern erst recht die *Kreativität* von Leistungen wird zum Problem für ein Prüfungssystem, das um seiner Berechenbarkeit willen Beurteilungskriterien *vorweg* möglichst *eng* festlegen muß. Denn qua Definition schließen Leistungen wie Selbständigkeit, Kritikfähigkeit, Originalität u. ä. die *Normabweichung* als Gütemerkmal mit ein. Auch in dieser Hinsicht muß ein Kriteriensatz also offen und *entwicklungsfähig* bleiben, wenn nicht gerade diejenigen Leistungen aus der Bewertung ausgeschlossen werden sollen, die in den Richtlinien der Länder, in der pädagogischen Programmatik, in psychologischen Taxonomien und in den Anforderungen der Hochschule einen hohen Rang genießen. Insofern würde die generelle Forderung nach *Operationalisierung* das Spektrum notenrelevanter Leistungen unzumutbar reduzieren.

(3) Faßt man die skizzierten Argumente zusammen, so wird deutlich, daß die Ausblendung von individuellen Erfahrungshintergründen bei Prüflingen wie bei Prüfern zwar vordergründig mehr »Einheitlichkeit« schafft, daß eine solche Objektivierung aber die *prinzipielle Interpretationsbedürftigkeit* menschlichen Handelns unter den Tisch kehrt. Die Übereinstimmung von Prüfern ist genausowenig ein Kriterium für die Güte ihres Urteils, wie sich in den Wissenschaften eine historisch zufällige Mehrheit als Beweis für die Richtigkeit einer Erklärung auslegen läßt.

Es sollte damit deutlich geworden sein, daß die Realisierung der »Einheitlichkeits«-Forderung nicht nur methodische Probleme aufwirft, daß also die vorliegenden Normenbücher nicht lediglich unzulängliche Beispiele für ein an sich vernünftiges und erreichbares Konzept sind. Da »Vergleichbarkeit« zwei sich wechselseitig aufhebende Forderungen einschließt, nämlich die generelle Berechenbarkeit *und* die individuelle

Angemessenheit einer Note, sind Brüche und Spielräume in der Praxis unvermeidlich, wenn die Prüfungsergebnisse nicht trivial sein sollen. Damit deutet sich eine Verschiebung des Problems an: Wenn die *technische Präzision* von Prüfungen, also der Aufgaben, Verfahren und Kriterien der Beurteilung nur beschränkt gesteigert werden kann, weil sie an die eben skizzierten prinzipiellen Grenzen stößt, gewinnt die *soziale Kontrolle* von Urteilen an Bedeutung (grundsätzlicher dazu: H. Brügelmann 1974).

Die zentrale Problematik der politischen Verantwortung von Prüfungsentscheidungen bricht also in den Normenbüchern immer wieder durch, wird aber nicht thematisiert. Die Brisanz steckt in so alltäglichen Fragen wie: Wer soll (und wer kann) das letzte Wort haben, wenn beim Hochschulzugang Konflikte auftreten zwischen Schülern und Lehrern; zwischen Lehrern und Verwaltung; zwischen Schüler A und Schüler B; zwischen Bundesländern; zwischen Schule und Hochschule? Prüfungskriterien sind eine nützliche Orientierungshilfe. Die rechtlichen Zuständigkeiten und fachlichen Qualifikationen der an solchen Bewertungen Beteiligten werden aber demgegenüber in der Praxis nachdrücklich durchschlagen und bedürften von daher besonders dringend einer Regelung, will man nicht unter dem Mantel säuberlicher »Normen« einen Kampf ums Überleben entfachen, in dem der jeweils Stärkere obsiegt – sei es ein Schüler wegen der besseren Ausstattung seiner Schule oder eines besseren Unterrichts; sei es ein Pädagoge wegen seines höheren Prestiges im Kollegium oder der einflußreicheren Position in der Verwaltungshierarchie; sei es ein Bundesland wegen der zahlenmäßigen Überlegenheit seiner bildungspolitischen oder fachdidaktischen Position in einem Fachausschuß der KMK. Diese *latenten* Bedingungen und Auswirkungen der Einführung der Normenbücher scheinen mir das gegenwärtig heikelste Problem zu sein, zu dem ich wenigstens ein paar Stichworte anfügen möchte (vgl. zu einigen Aspekten ausführlicher die Beiträge in: A. Flintner 1976).



#### IV. *Wie einheitlich und verbindlich dürfen Prüfungsanforderungen sein?*

Meine bisherigen Anmerkungen könnten zu dem Fehlschluß verleiten, die Normenbücher würden zwar ihrem eigenen Anspruch nicht gerecht, dafür drohe aber auch nicht das befürchtete Einheitscurriculum mit Zentralabitur – und immerhin sei eine gewisse Vereinheitlichung besser als gar keine. Eine solche Argumentation verkennt die psychologische Wirkung und die Eigendynamik, die ein Instrument wie die Normenbücher entfalten kann (vgl. etwa die Andeutungen in: KMK 1975, S. 7).

(1) Solange der Vergleichbarkeitsanspruch aufrechterhalten bleibt, werden nämlich auch unscharfe oder von den Fachkommissionen bewußt auslegungsfähig gehaltene Lernzielformulierungen *unter den Prüfern Unsicherheit erzeugen*. Der Sache nach unvermeidliche Interpretationsunterschiede können als persönliche Unzulänglichkeit angekreidet werden. Die auf dem Papier vielleicht schon aufgegebene Einheitlichkeit wird dann institutionell erzeugt, d. h. die Interpretation von Autoritäten wird verbindlich und schafft eine zweite Normebene, die nicht mehr durchsichtig und kontrollierbar ist, sondern sich aus den jeweiligen Machtverhältnissen ergibt. Daß diese Auslegungspraxis wiederum von Schule zu Schule und von Region zu Region unterschiedlich aussehen dürfte, könnte sich produktiv auswirken, wenn sie gewollt wäre und sich jeweils für die besonderen Bedingungen des Einzelfalls nutzen ließe. Dem steht aber bisher der papierne Einheitlichkeitsanspruch entgegen.

(2) Ein zweites Problem entsteht für die *Lehrer* bei der *didaktischen Planung* ihres Unterrichts. Gleichgültig, ob die Lernziele in den Normenbüchern verbindlich konkretisiert wurden oder ob nur der entsprechende Anspruch aufrechterhalten wird – die Auswahl der Unterrichtsinhalte in den Schulen wird sich nach den (vermeintlich) zu erwartenden Prüfungsaufgaben richten. Diese ihrerseits müssen um des Einheitlichkeitsanspruchs willen prüfungstechnischen Gesichtspunkten genügen, so daß sich die ursprüngliche Begründungsfolge: (1) Was soll eigentlich in der Schule gelernt werden? (2) Wie kann man die beabsichtigten Lerneffekte feststellen?, leicht umkehrt in die Abfolge: (1) Welche Lerneffekte sind überhaupt prüf-

bar? (2) Welcher Unterricht kann den Schülern in dieser Perspektive zum Erfolg verhelfen?

Es geht hier nicht um die Unterstellung, solche Konsequenzen seien *gewollt*; allerdings ist zu fragen, ob eine erhebliche *Reduktion des Unterrichtsangebots tatsächlich* zu vermeiden ist, wenn die Einheitlichkeitsforderung ein solches Gewicht gewinnt, daß sie nur auf Kosten der *inhaltlichen Bedeutsamkeit*, des Erfahrungsbezugs und des örtlichen, gruppenspezifischen oder individuellen Interesses realisiert werden kann. Auch in den *Prüfungsanforderungen* selbst dürfte es zu einer *quantitativen* Ausweitung in den leichter prüfbaren Bereichen kommen, weil eine *qualitative* Steigerung der Anforderungen einen zunehmenden Verzicht auf die Berechenbarkeit und Kontrollierbarkeit von Noten einschließt (s. oben III.2). Man sollte sich nämlich nicht der Illusion hingeben, der prüfungsorientierte Curriculumanteil könnte derart begrenzt werden, daß *daneben* Lernziele wie selbständiges Arbeiten, Kooperationsfähigkeit, Sachinteresse, Umweltverständnis, Handlungsfähigkeit verfolgt werden könnten. Zum einen setzt der Erwerb solcher Grundqualifikationen einen Unterrichtsstil voraus, der für die Schule insgesamt bestimmend werden muß. Zum anderen fallen etwa schon heute in vielen 1. (!) Schuljahren der von den Richtlinien verbindlich (!) geforderte Sachunterricht und musische Angebote aus, weil anders nach Einschätzung von Lehrern oder Eltern die Anforderungen im Lesen, Schreiben und Rechnen nicht zu erfüllen sind; daß dabei oft nicht mehr »das Gymnasium«, sondern zunehmend der Numerus clausus selbst als Argument ins Feld geführt wird, läßt für die Sekundarstufe II Schlimmes erwarten.

(3) Schließlich zu den *Schülern* selbst. Je eindeutiger die Unterrichtsnormen festgelegt sind, desto weniger Raum bleibt für die individuell unterschiedlichen Lebensperspektiven und biographischen Erfahrungen, für selbständiges Lernen und abweichende Meinungen, für die Bereitschaft zum Risiko im Lernprozeß. Denn solange der Numerus clausus über Notendurchschnitte abgewickelt wird, werden Schüler für Prüfungsnoten lernen. Ist es schon deprimierend, wenn die von der Schule angezielte Auseinandersetzung des Schülers mit Erfahrungen und Problemen aus seiner Umwelt zum Kampf gegen

den Lernerfolg des Mitschülers wird, so führt das Lernzieldilemma das System der Normenbücher selbst ad absurdum: Je *genauer* für den Schüler vorhersehbar ist, was und wie er geprüft werden wird, desto leichter ist es für ihn, bloß prüfungstaktisch zu lernen – und die Validität der Bewertungsmaßstäbe endgültig außer Kraft zu setzen; je *offener* das Bewertungssystem, desto größer die Anpassung an den Lehrer aufgrund zunehmender Unsicherheit über die Kriterien und tatsächlicher Abhängigkeit von seinem Urteil.

Meine Argumentation darf insofern nicht als Plädoyer für ein unberechenbares oder gar willkürliches Prüfungssystem mißdeutet werden. Konsequenterweise müßte vielmehr die *Koppelung* einer existenzentscheidenden Frage wie des Hochschulzugangs an einen Teufelskreis taktischer Winkelzüge von Prüfungsbehörden und Prüflingen aufgehoben werden, in dessen Fortgang sowohl der letzte Rest an Validität auf der Strecke bleiben muß, den zentrale Prüfungsmaßstäbe überhaupt erreichen können, als auch alles das, was Schule für eine Jugend vielleicht noch bedeutsam machen könnte.

Faßt man die Begründung der drei Eingangsthesen zusammen, muß man also *sowohl* den Wert zentraler Normen *als auch* den Beitrag einer individuellen Benotung für eine inhaltlich gerechte und prozedural nachprüfbar Entscheidung über die Studienberechtigung grundsätzlich in Frage stellen. Das Rationalitätsversprechen der Normenbücher und der zu seiner Rechtfertigung betriebene technische Aufwand verschleiern die Problematik der Willkür und der Zufälligkeit des Auswahlresultates eher, als daß sie sie lösen.

Sicher ist es dringend notwendig, intensiv über Möglichkeiten der Feststellung und Bewertung von Lernerfolgen, vielleicht mehr noch von *Lernschwierigkeiten* nachzudenken. Sicherlich kann und muß auch manches getan werden, um die Form und das technische Instrumentarium solcher Einschätzungen zu verbessern – auch im Sinne eines länderübergreifenden Austauschs von Vorschlägen und Erfahrungen. Eine solche Diskussion dürfte aber nicht von aufgezwungenen politischen Ansprüchen ausgehen, sondern müßte sich der oben skizzierten Problematik des Gegenstandes stellen können. Perspektiven für eine solche experimentelle d. h. erfahrungsoffene

und die Vielfältigkeit bzw. Differenziertheit des pädagogischen Sachverhalts nicht verkürzende Entwicklung lassen sich an dieser Stelle nur schlagwortartig bezeichnen: *Orientierungshilfen* statt Vorschriften; *Rahmenbestimmungen* statt Detailverordnungen; pädagogisch orientierte *Diagnose* statt bloßer Selektion; Förderung der diagnostischen *Fähigkeit von Lehrern* statt Perfektionierung von Instrumenten; Regelung der *sozialen Kontrolle* statt Konzentration auf technische Präzision.

Die Kultusminister stehen vor einer schwierigen Entscheidung. Der bildungspolitische Druck ist stark und das Versprechen der Einheitlichkeit populär. Aber wer übernimmt die Verantwortung für die Konsequenzen?

## Literatur

- BRÜGELMANN, H.: Lernziele im offenen Curriculum. In: Thema Curriculum (1972) 2, S. 16–45.
- BRÜGELMANN, H.: Towards Checks and Balances in Educational Evaluation – on the use of social control in research design. In: Safari (Hrsg.), Innovation, Evaluation, Research and the Problem of Control. Workshop Curriculum Bd. I (Narr), Tübingen 1974, S. 25–26.
- BRÜGELMANN, H./FISCHER, D.: Lesefertigkeit oder Spracherfahrung – eine falsche Alternative. Teil B: Taktiken des Lesens – Skizze eines kognitiven Modells des Leseprozesses. In: Arbeitskreis Grundschule (Hrsg.): Fabeln und Erstlesewerke II, Frankfurt 1977.
- FLITNER, A. (Hrsg.): Der Numerus clausus und seine Folgen, Stuttgart 1976.
- HOPF, D.: Das Numerus-clausus-Verfahren. Möglichkeiten, Grenzen, Folgewirkungen. In: Neue Sammlung 14 (1974) 2, S. 180–188.
- HOPF, D.: Zensuren und Tests zur Hochschulauswahl. In: A. Flitner (Hrsg.), Der Numerus clausus und seine Folgen, Stuttgart 1976, S. 100–107.
- INGENKAMP, K. (Hrsg.): Die Fragwürdigkeit der Zensurenggebung, Weinheim 1971.
- KMK: Einheitliche Prüfungsanforderungen in der Abiturprüfung (14 Hefte zu den einzelnen Fächern, denen jeweils die KMK-Beschlüsse vom 6. 2. 1975 und 23. 5. 1975 vorangestellt sind), Neuwied 1975.
- LENZEN, D.: Die Illusion der Vereinheitlichung – Normenbücher zwischen Testpsychologie und Verfassungsrecht. In: Neue Sammlung 16 (1976) 4, S. 295–308; auch in: A. FLITNER 1976. MEYER, H. L.: Trainingsprogramm zur Lernzielanalyse, Frankfurt 1974.
- REICHERT, S., u. a.: Vereinheitlichung der Normenbücher? Eine analytisch-synoptische Studie. Staatsinstitut für Schulpädagogik Arbeitsberichte, Nr. 16, München 1976.
- ZIEGENSPECK, J.: Zensur und Zeugnis in der Schule, Hannover <sup>2</sup>1975.