

Zeitschrift: Schweizer Schule
Herausgeber: Christlicher Lehrer- und Erzieherverein der Schweiz
Band: 61 (1974)
Heft: 11: Erfolgskontrolle des Unterrichts : Prüfungen!

Artikel: Erfolgskontrolle des Unterrichts
Autor: Messner, Helmut
DOI: <https://doi.org/10.5169/seals-530262>

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. [Mehr erfahren](#)

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. [En savoir plus](#)

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. [Find out more](#)

Download PDF: 28.03.2026

ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>

Erfolgskontrolle des Unterrichts

Helmut Messner

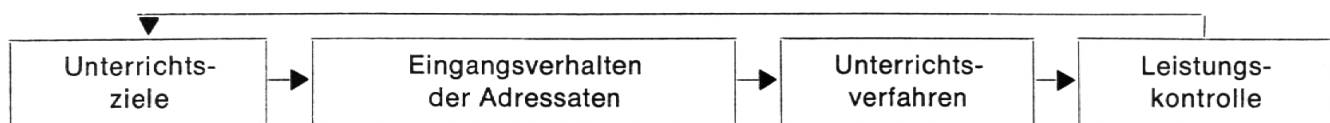
1. Erfolgskontrolle als wichtiger Bestandteil jeden Unterrichts

Viele Lehrer, die tagtäglich damit zu tun haben, die Leistungen ihrer Schüler zu kontrollieren und zu beurteilen, stellen sich die Frage nach dem Sinn und Zweck von Prüfungen und Noten. Handelt es sich hierbei um einen fragwürdigen institutionellen Zwang, den es zu boykottieren gilt, oder um eine didaktisch sinnvolle Massnahme? An dieser Frage scheiden sich die Geister nicht nur in der Lehrerschaft, sondern auch unter Erziehungswissenschaftlern. Die einen lehnen Prüfungen und Noten mit dem Hinweis auf ihre – z. T. empirisch erwiesenen – unerwünschten Nebenwirkungen im Bereich des sozialen und motivationalen Verhaltens der Schüler ab: Noten begünstigen ein Konkurrenzdenken unter den Schülern, das im Widerspruch zu Werten wie Solidarität und Kooperation steht, die die Schule zu pflegen habe. Noten bedingen ferner ein Lernverhalten, das nicht so sehr durch das Interesse am Lerngegenstand (intrinsische Motivation) als vielmehr durch die Hoffnung auf eine «gute» Note oder aus Angst vor einer «schlechten» Note (extrinsische Motivation) aktiviert und gesteuert wird. Prüfungsangst und andere Persönlichkeitsmerkmale können Prüfungsergebnisse völlig verzerren (vgl. SARASON 1971), so dass diese nur eine beschränkte Aussagekraft über die tatsächliche Leistungsfähigkeit haben. Leistungskontrollen leisten einer Aufspaltung der Klasse in «gute» und «schlechte» Vorschub und prägen so das Selbstbild der Schüler. Diese Nebenwirkungen der Leistungsbeurteilung stehen in der Tat im Widerspruch zu erklärten pädagogischen Zie-

len des Unterrichts, wie Förderung der Fähigkeit zur Kooperation und Solidarität, Weckung und Pflege sachbezogener Interessen, Aufbau eines positiven Selbstbildes, Abbau von Ängstlichkeit u. a. m.

Zu diesen pädagogisch motivierten Bedenken gegen das derzeitige System der Leistungsbeurteilung kommen methodische Einwände, die die geringe Objektivität und damit die fehlende Vergleichbarkeit von Noten sowie ihren beschränkten prognostischen Wert betreffen (vgl. INGENKAMP 1971). Die Ablehnung des gegenwärtigen Systems der Leistungsbeurteilung in der Schule aufgrund dieser Befunde ist zwar verständlich und nachvollziehbar – jedoch einseitig. Man mag das gegenwärtige System in vielen Punkten zu Recht in Zweifel ziehen; jede Form der Leistungskontrolle im Unterricht von vornherein abzulehnen, hiesse jedoch, «das Kind mit dem Bade ausschütten». Die Leistungskontrolle ist ein wichtiges und notwendiges Element eines jeden Unterrichtssystems, das seinem Wesen nach ein zweckrationales Gebilde ist. GLASER (1962) nennt vier grundlegende Variablen, die seiner Meinung nach für den Unterricht konstitutiv und bei dessen Planung und Analyse zu berücksichtigen sind: Unterrichtsziele, Eingangsverhalten der Adressaten, Unterrichtsverfahren und Leistungskontrolle. Diese Variablen stehen untereinander in wechselseitigen Abhängigkeitsverhältnissen und bilden zusammen ein System. Verändert man eine Variable, so verändern sich zwangsläufig auch die anderen. Neue Unterrichtsziele erfordern neue Unterrichtsverfahren und neue Leistungskontrollen.

Abb 1: Variablen eines Unterrichtssystems (nach GLASER 1962)



Die Leistungskontrolle erfüllt innerhalb dieses Systems vor allem zwei Funktionen: Sie dient erstens der Überprüfung der Wirksamkeit des Unterrichts (evaluative Funk-

tion) und zweitens der Rückmeldung an Lehrer und Schüler über erzielte Lernfortschritte und aufgetretene Lernschwierigkeiten im Unterricht. Diese beiden didaktischen

Funktionen der Leistungskontrolle sind zu unterscheiden von der selektiven Funktion, die Prüfungen auch erfüllen können. Bei Eignungs- und Befähigungsprüfungen (z. B. Aufnahmeprüfung, Lehrabschlussprüfung, Maturitätsprüfung) dient die Leistungskontrolle der Ausscheidung (Selektion) ungeeigneter Kandidaten. Bei solchen Prüfungen sollen Informationen über die relative Leistungsfähigkeit der einzelnen Schüler gewonnen werden, die Voraussagen über ihr künftiges Leistungsverhalten zulassen. In diesem Fall ist also die relative Rangordnung von Individuen in bezug auf ihre Leistungsfähigkeit leitendes Erkenntnisinteresse und nicht so sehr die Kontrolle der Wirksamkeit des Unterrichts.

1, Im Rahmen des Unterrichts hat die Leistungskontrolle indessen primär didaktische Funktionen. Sie dient einmal der Kontrolle des Wirkungsgrades des Unterrichts bezogen auf die angestrebten Lernziele. Die Bewertung der Unterrichtsqualität aufgrund äusserer Merkmale des Unterrichts oder seiner Wirkungen auf die Schüler bezeichnet man in der Erziehungswissenschaft als Evaluation. Man spricht deshalb auch von der evaluativen Funktion der Leistungskontrolle. Der Unterrichtserfolg wird am Grad der Übereinstimmung zwischen den intendierten Lernzielen und den tatsächlich erzielten Lernergebnissen abgelesen und gemessen. Bezugspunkt und Massstab der Leistungskontrolle sind hierbei ausschliesslich die im Unterricht angestrebten Lernziele. Im Hinblick auf die Evaluation des Unterrichts interessiert vor allem, ob und in welchem Grade die Schüler die im Unterricht angestrebten Lernziele erreicht haben. Die auf diese Weise gewonnenen Informationen sind für die weitere Unterrichtsplanung und die langfristige Verbesserung des eigenen Unterrichts wichtig. Wenn ein Teil der Schüler wesentliche Lernziele nicht erreicht hat, so hat der Lehrer die notwendigen und wirksamen Dispositionen zu treffen, um diese Lücke zu schliessen. Zahlreiche negative Resultate bei einer Leistungskontrolle über einen Unterrichtsabschnitt oder eine bestimmte Unterrichtseinheit sind ein Alarmzeichen. Sie deuten auf einen geringen Wirkungsgrad des Unterrichts und legen eine didaktische Neukonzeption nahe.

Eine weitere didaktische Funktion der Lei-

stungskontrolle ist die Rückmeldung an Schüler und Lehrer über erzielte Lernfortschritte und/oder aufgetretene Lernschwierigkeiten. Schüler verlangen und brauchen im Interesse ihrer Lernbemühungen eine Orientierung über ihre Lernfortschritte und Leistungsschwächen. Die Rückmeldung über den erzielten Lernerfolg ist eine lernpsychologisch notwendige Bedingung nicht nur für die Lernmotivation, sondern allgemein für ein gezieltes Lernverhalten (vgl. CORRELL 1966, 148 ff.). Die Rückmeldung wirkt als positive oder negative Verstärkung der erbrachten Leistung. Die differenzierte Kenntnis der eigenen Leistungsschwächen ist eine notwendige Voraussetzung für gezielte Bemühungen zur Leistungsverbesserung. So unterscheidet sich der «gute» Skilehrer vom weniger «guten» u. a. dadurch, dass ersterer in der Lage ist, seinen Schülern genaue Angaben über ihre Fehler zu machen (z. B. Haltungsfehler, Fehler in der Skiführung) und darauf abgestimmt geeignete Lernhilfen anzubieten vermag. Dies gilt in analoger Weise auch für den Schulunterricht. Schliesslich haben Leistungskontrollen auch für den Lehrer selbst eine wichtige Rückmeldefunktion. Sie geben ihm nicht nur Auskunft darüber, wie sein Unterricht «angekommen» ist und gewirkt hat, sondern liefern ihm gleichzeitig wichtige Aufschlüsse, wo der einzelne Schüler in bezug auf die Erreichung angestrebter Lernziele steht und welche individuelle Lernrückstände oder -ausfälle vorliegen. Die Kenntnis des individuellen Lernstandes der einzelnen Schüler befähigt den Lehrer, seinen Unterricht in didaktischer Hinsicht zu differenzieren bzw. zu individualisieren. Ohne Kenntnis der spezifischen Lernschwierigkeiten und Lernvoraussetzungen seiner Schüler ist eine Abstimmung des Unterrichtsangebots auf die individuellen Erfordernisse nicht möglich.

Zwischen der Funktion einer Leistungskontrolle und ihrer methodischen Gestaltung besteht ein enger Zusammenhang. Je nachdem, ob eine Prüfung der Erfassung der relativen Leistungsfähigkeit der einzelnen Schüler (selektive Funktion) oder der Kontrolle des Wirkungsgrades des Unterrichts (evaluative Funktion) dient, sind bei der Auswahl von Prüfungsaufgaben sowie bei der Anwendung eines Modells der Noten-

gebung andere Regeln zu beachten (vgl. GLASER 1973). Soll die relative Leistungshöhe der einzelnen Schüler ermittelt werden, so sind Aufgaben von steigendem Schwierigkeitsgrad und hoher Trennschärfe (siehe Punkt 2 dieses Beitrages) erforderlich. Soll dagegen in erster Linie die Wirksamkeit des Unterrichts kontrolliert werden, so sind solche Aufgaben auszuwählen, die die angestrebten Lernziele möglichst gut repräsentieren. Dazu ein Beispiel: Wer nach einer Unterrichtseinheit über die Dreisatz- oder Schlussrechnung primär an der Diagnose der relativen Leistungsfähigkeit seiner Schüler interessiert ist, wird angewandte Aufgaben unterschiedlichen Schwierigkeitsgrades stellen, die geeignet sind, die «guten» von den «schlechten» Schülern zu unterscheiden. Wer hingegen in erster Linie etwas über den Wirkungsgrad seines Unterrichts zu erfahren wünscht, wird vor allem solche Aufgaben stellen, die *verschiedene Aspekte der komplexen Fähigkeit* «Lösen von Dreisatzaufgaben» betreffen, die im Unterricht geübt worden sind: z. B. Schluss von einer Einheit auf eine Mehrheit; Schluss von einer Mehrheit auf eine Einheit; Schluss von einer Mehrheit auf eine Einheit und auf eine andere Mehrheit; Schluss von einer Mehrheit auf eine andere Mehrheit, die ein Vielfaches oder einen Teil der ersteren darstellt; Schlussrechnungen mit verschiedenen Massen; direkte und inverse Proportionen. Dieses Prinzip wird in der schulischen Prüfungspraxis häufig zu wenig beachtet. Eine Analyse der in der Schule verwendeten Modelle der Notengebung (vgl. 3.2. dieses Beitrages) und der gestellten Prüfungsaufgaben bei Klassenarbeiten zeigt, dass Leistungskontrollen im Unterricht offensichtlich vor allem eine selektive Funktion erfüllen und ihre evaluative Funktion nicht oder zu wenig beachtet wird. Mir scheint, dass auf diesen Umstand zu einem guten Teil jene schädlichen Nebenwirkungen auf die motivationale und soziale Entwicklung von Schülern zurückzuführen sind, die vor allem von jungen Lehrern zu Recht angeprangert werden. Wenn Prüfungen vorwiegend zum Zweck der Evaluation des eigenen Unterrichts und der Rückmeldung an Schüler und Lehrer durchgeführt werden, so ist zu erwarten, dass Prüfungen auch im Bewusstsein der Schüler eine andere Bedeutung

erhalten und jene unerwünschten Nebenwirkungen ausbleiben oder zumindest eingeschränkt werden, die ich eingangs erwähnt habe.

2. Prüfungen als Hilfsmittel zur Messung von Lernergebnissen

In der Diskussion zur Erfolgskontrolle im Unterricht wird manchmal zwischen Leistungsbeurteilung und Leistungsmessung unterschieden. Man spricht von *Leistungsmessung*, wenn Prüfungsergebnisse in der Form von Punktwerten oder Fehlerhäufigkeiten quantitativ ausgedrückt werden, ohne dass damit eine Bewertung der Leistung im Sinn von «gut» oder «schlecht» erfolgt. Bei der *Leistungsbeurteilung* dagegen wird die «Güte» der Leistung anhand einer Notenskala (Ziffernskala) oder einer Rangreihe von Prädikaten (ausgezeichnet, sehr gut, gut, befriedigend, genügend, ungenügend) eingestuft. Aus methodologischer Sicht besteht indessen kein Unterschied zwischen diesen beiden Vorgängen: Es handelt sich in beiden Fällen um eine Abbildung einer Leistung auf einer Skala, einen Vorgang, den wir in der Methodologie als *Messung* bezeichnen. Beide Messvorgänge unterscheiden sich lediglich durch die Art der Skalen oder Massstäbe, die der Messung zugrunde gelegt werden (Punkteskala oder Notenskala). Die Erfassung von Lernergebnissen in der Form von Punktwerten und ihre Umwandlung in eine Note stellen häufig zwei aufeinanderfolgende Schritte bei der Leistungsmessung im Unterricht dar.

2.1. Was wird in Prüfungen gemessen?

Gegenstand von Prüfungen sind *Lernergebnisse*, die im Unterricht angestrebt worden sind. Lernergebnisse, die keinen Bezug zu den *Lernzielen* des Unterrichts aufweisen, sind für die Kontrolle des Unterrichtserfolges nicht relevant, es sei denn, dass man durch die Prüfung unbeabsichtigte Nebenwirkungen des Unterrichts aufdecken möchte. Wenn wir von Lernergebnissen sprechen, so meinen wir in der Psychologie mehr oder weniger komplexe «innere Zustände» eines Schülers, die sein effektives Verhalten (z. B. Problemlöseverhalten) steuern und bestimmen. In unserer Umgangssprache kennzeichnen wir diese Zustände als Wissen, Verständnis oder Können. Es handelt sich

hierbei einerseits um erworbene Begriffe und Operationen, andererseits um erworbene Fertigkeiten (z. B. Lesen, Schreiben, Benutzen des Rechenschiebers, Hochsprung), Haltungen (Ordnungsliebe, Kameradschaft) und Einstellungen (Sympathie, Antipathie). Je nachdem, welche Lernergebnisse erfasst und erhoben werden sollen, sind andere Prüfungsmethoden erforderlich. Die in der Schule vorherrschenden mündlichen und schriftlichen Prüfungen dienen in erster Linie der Erfassung von Lernergebnissen im kognitiven und psychomotorischen Verhaltensbereich, also der Kontrolle erworbener Begriffe, Operationen und praktischer Fertigkeiten. Die Einschätzung von Haltungen, Einstellungen und Interessen der Schüler erfolgt meistens auf der Grundlage mehr oder weniger unsystematischer Verhaltensbeobachtungen. Ich befasse mich im folgenden insbesondere mit der Erhebung und Messung sogenannter «kognitiver Strukturen», womit ich zusammenfassend Begriffe, Operationen und Verhaltensschemata meine, über die ein Schüler verfügt und die er im Unterricht aufgebaut hat. *Kognitive Strukturen* sind in sich geschlossene und überdauernde Beziehungsgefüge, an denen Elemente und deren Relationen bzw. Verknüpfungen unterschieden werden können (vgl. AEBLI 1969).

Beispiele für Lernergebnisse, die kognitive Strukturen darstellen:

- Der Pythagoräische Lehrsatz
- Addition und Subtraktion von Brüchen
- Grundbegriffe der Mechanik (Arbeit, Leistung, Kraft, das Beharrungsprinzip, das Fallgesetz, das Kräfteparallelogramm)
- Geistesgeschichtliche Hintergründe der Französischen Revolution
- Faktoren der Standortwahl und Entwicklung von Siedlungen
- Erdumlauf-Beweise
- Verwendungsregeln des Dehnungs-h
- Das biologische Gleichgewicht
- Umweltfaktoren für Wachstum und Entwicklung von Pflanzen
- Der Symbolgehalt der Erzählung «Die Verwandlung» von Franz KAFKA

Die hier beispielhaft genannten Inhalte stellen kognitive Strukturen dar, die im einzelnen zu explizieren sind. Hier sind lediglich zusammenfassende Bezeichnungen für die individuelle Kenntnis von Zusammenhängen

und Beziehungen angeführt. So beinhaltet der Pythagoräische Lehrsatz eine flächenmässige Gleichheit zwischen den Quadraten über den Katheten mit dem Quadrat über der Hypothense bei rechtwinkligen Dreiecken. Diese Gleichheitsrelation konstituiert die kognitive Struktur des Pythagoräischen Lehrsatzes.

Erworbene Begriffe und Operationen lassen sich niemals direkt, sondern immer in der Form von daraus resultierenden Leistungen erfassen. Die Erfassung eines Lernergebnisses dieser Art gleicht in formaler Hinsicht der Messung der Temperatur, die ebenfalls nur indirekt über die Ausdehnung von Alkohol oder Quecksilber gemessen werden kann. Prüfungen haben also stets den Charakter von Indizienbeweisen: Die beobachtbare und damit messbare Leistung kann immer nur als Indiz der zugrunde liegenden überdauernden Verhaltensbereitschaft angesehen werden. Da bei Prüfungen immer nur Leistungen und niemals die überdauernden kognitiven Strukturen selbst unmittelbar erfasst werden können, ist es wichtig, dass man sich bei der Konstruktion von Probenarbeiten überlegt, welche Schülerleistungen das Vorhandensein der zu prüfenden Lernergebnisse anzeigen. Die Beschreibung schulischer Lernergebnisse in der Form von Leistungen, welche das Vorhandensein entsprechender Dispositionen anzeigen, bezeichnet man in der Didaktik als *Operationalisierung von Lernzielen*. Die Operationalisierung der angestrebten Lernziele ist eine notwendige Voraussetzung für die Entwicklung lernzielorientierter Prüfungen oder Tests. Mit einem operationalisierten Lernziel wird ausgedrückt, was der Lernende nach dem Unterricht können muss, wobei dieses Können als *beobachtbares Verhalten* formuliert wird (vgl. MAGER 1969³). Für viele komplexe Lernziele ist es schwierig, *gültige* operationale Kriterien anzugeben. Die Zuordnung von Leistungen zu Lernergebnissen erfordert fundierte psychologische und methodologische Kenntnisse.

2.2. Wie misst man Lernergebnisse?

Da Lernergebnisse immer nur indirekt über beobachtbare Leistungen erfasst und gemessen werden können, muss der Konstrukteur einer Prüfung oder eines Leistungstests

Beispiele für die Operationalisierung kognitiver Strukturen:

Kognitive Strukturen	Leistungsebene
Verständnis des Pythagoräischen Lehrsatzes	Bestimmen der Diagonale eines Quadrates bei gegebener Seitenlänge
Beherrschen der Regeln für die Addition und Subtraktion von Brüchen	Addieren und Subtrahieren von Brüchen an ausgewählten Beispielen
Verständnis des Beharrungsprinzips	Erklären der Schleudergefahr in Kurven
Verständnis des Kräfteparallelogramms	Zusammensetzen und Zerlegen von Kräften an einfachen Beispielen
Verständnis der Funktionsweise des Elektromotors	Angabe von Fehlerquellen für das Nichtfunktionieren eines Elektromotors
Verständnis der geistesgeschichtlichen Wurzeln der Französischen Revolution	Aufzeigen von politischen und gesellschaftlichen Theorien, die der Französischen Revolution zugrunde lagen
Theorie des Erdumlaufs	Aufzeigen von Beobachtungen, die nur durch die heliozentrische Theorie des Erdumlaufs erklärt werden können

geeignete Aufgaben und Fragestellungen erfinden, die Leistungen verlangen, welche das angestrebte Lernergebnis möglichst gut repräsentieren. Die Konstruktion von Prüfungsaufgaben ist eine schöpferische Tätigkeit, die ein vertieftes Verständnis des Lerninhaltes und der Psychologie des Problemlösens erfordert. Je nach Aufgabentyp werden in erster Linie reproduktive Leistungen als Antworten auf reine Wissensfragen, Transferleistungen (z. B. bei Verständnis- und Anwendungsaufgaben) oder schöpferisch-kreative Leistungen verlangt.

Beispiele für Wissensfragen:

Wo liegt...? Wann geschah...? Welches sind die Bestandteile von...? Was heisst...? Wie nennt man...? Wie lautet...?

Beispiele für Verständnisfragen:

Was geschieht, wenn... (Folgen voraussagen)?
Wie ist dieser oder jener Effekt zu erklären?
Welche Beziehung besteht zwischen...?
Welches sind die Gemeinsamkeiten, welches die Unterschiede zweier Gegebenheiten oder Phänomene?
Welches sind die Vorteile, welches die Nachteile einer Einrichtung oder Massnahme?

- Aufgabe: Überlegen Sie sich, ob folgende Fragen in erster Linie Reproduktionsleistungen oder Transferleistungen verlangen!
 - Was ist ein Siebdruck?
 - Welche Folgen zieht die chemische Schäd-

- lingsbekämpfung im Obst- und Gemüsebau nach sich?
- Warum kann eine Last mit einem Flaschenzug leichter gehoben werden als mit einer einfachen Rolle?
- Wie lautet das Ohmsche Gesetz?
- Worin unterscheiden sich Gotik und Renaissance in der Baukunst?
- Was wird mit der Erhöhung des Diskontsatzes bezweckt?
- Warum hat der Tag 24 Stunden?
- Wie beweist man die Gültigkeit des Pythagoräischen Lehrsatzes?

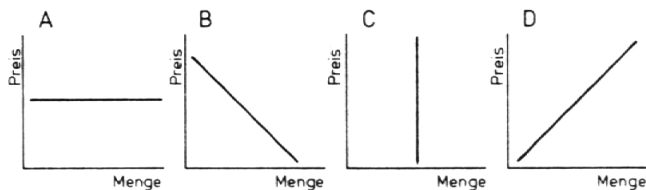
2.2.1. Aufgaben mit freier und gebundener Beantwortungsform

Die an unseren Schulen verbreiteten Klassenarbeiten und mündlichen Prüfungen bestehen in der Regel aus Aufgaben oder Fragen mit freier Antwortmöglichkeit. Beispiele für solche «offene» Aufgaben sind die üblichen Anwendungsaufgaben im Bereiche der Mathematik und der Naturlehre. Bei offenen Aufgaben formuliert der Schüler seine Antwort oder Lösung selbst. Auch der Aufsatz oder Kurzaufsatz stellt eine Aufgabe mit freier Beantwortungsform dar. Dagegen bevorzugt man in Leistungstests oder standardisierten Prüfungen (z. B. bei der theoretischen Autofahrprüfung) in der Regel Aufgaben mit gebundener Beantwortungsform. Bei solchen «geschlossenen» Aufgaben werden die Antwortmöglichkeiten vorgegeben und

der Prüfling hat unter ihnen diejenige Antwort oder Lösung auszuwählen, die er selbst für richtig hält. Das bloße Raten wird durch die Vorgabe von vier und mehr Antwortalternativen kontrolliert oder zumindest in Grenzen gehalten. Es gibt mehrere Typen von Aufgaben mit gebundener Beantwortungsform (vgl. WENDELER 1972⁴, 18 ff.). Die bekanntesten und verbreitetsten unter ihnen sind die Mehrfachwahlaufgabe und die Zuordnungsaufgabe. Die Mehrfachwahlaufgabe (multiple choice-item) besteht aus zwei Teilen: aus einem Aufgabenstamm (eine Frage oder eine unvollständige Feststellung) und mehreren Angaben in der Form von Antworten bzw. Lösungen oder Ergänzungen der unvollständigen Feststellung.

Beispiele für Mehrfachwahlaufgaben:

Welche der folgenden graphischen Darstellungen zeigt am besten die Nachfrage nach einem typischen Gebrauchsartikel unter Konkurrenzbedingungen?



(nach BLOOM 1972, 110)

«Alle Ideen stammen aus der Erfahrung oder aus der Reflektion über die Erfahrung. Sinneseindrücke, denen Bedeutung gegeben wird, sind Perzeptionen. Die Assoziation von Perzeptionen oder einfachen Ideen führt zu komplexen oder abstrakten Begriffen, deren ursprüngliche Quelle immer noch die Erfahrung ist.» Dieses Zitat veranschaulicht am besten die Ansicht des ...

- A – Positivismus
- B – Rationalismus
- C – Idealismus
- D – Empirismus
- E – Pragmatismus

(nach BLOOM 1972, 107)

Die Zuordnungsaufgabe ist eine Kombination mehrerer Wahlfachaufgaben.

Ein Beispiel für eine Zuordnungsaufgabe:

Sie haben nun einige Kenntnis über die Erde und ihre tatsächliche Bewegung erworben. In der folgenden Übung sollen Sie die Auswirkungen einiger völlig erfundener Bedingungen identifizieren. Bezeichnen Sie jede der folgenden Aufgaben nach dem hier angegebenen Schlüssel mit:

- A – wenn die Aussage richtig wäre, falls die Erdachse *nicht* zur Bahn geneigt wäre.
- B – wenn die Aussage richtig wäre, falls die Bahn der Erde ein Kreis statt einer Ellipse wäre.
- C – wenn die Aussage richtig wäre, falls sich die Erde nach Westen statt nach Osten drehen würde.
- D – wenn die Aussage richtig wäre, falls die Erde bei ihrem halben Durchmesser die gleiche Masse besäße.
- E – wenn die Aufgabe richtig wäre, falls die Erde keinen Mond hätte.

Nehmen Sie an, dass jeweils nur eine der fiktiven Bedingungen zutrifft.

1. Alle Tage wären gleich lang.
2. Gegenstände viermal so schwer, als sie jetzt sind.
3. Der Himmelsäquator und die Ekliptik wären identisch.
4. Die Sonne würde im Osten untergehen.
5. Man müsste einen anderen Nordstern auswählen.
6. Die Schwerkraft wäre viermal so gross.
7. Die Bahngeschwindigkeit der Erde würde sich während eines Jahres nicht verändern.
8. Wir wüssten sehr viel weniger über die Natur der Sonne.
9. Nacht und Tag wären auf allen Breiten während des Jahres gleich lang.

(nach BLOOM 1972, 146 f.)

Diese Zuordnungsaufgabe setzt sich aus neun Mehrfachwahlfragen mit je fünf Antwortalternativen (A, B, C, D, E) zusammen.

Geschlossene Aufgaben wie Mehrfachwahlaufgaben und Zuordnungsaufgaben haben gegenüber offenen Aufgaben den Vorteil, dass ihre Auswertung relativ einfach ist und u. U. auch maschinell (mit EDV) vorgenommen werden kann. Es gibt bei diesen Aufgaben eine oder mehrere eindeutig richtige oder falsche Lösungen, während es bei offenen Aufgaben häufig schwierig ist, zwischen richtigen, teilrichtigen und falschen Lösungen zu unterscheiden. Allerdings haben geschlossene Aufgaben gegenüber offenen auch einen gewichtigen Nachteil: Bei Aufgaben mit gebundener Antwortform wird nur mehr die *Entscheidung* des Prüflings für eine vorgegebene Lösungsalternative sichtbar. Die der Entscheidung zugrunde liegenden komplexen Denk- und Verarbeitungsprozesse bleiben dagegen verborgen. Bei Aufgaben mit freier Beantwortungsform werden hingegen auch die komplexen Denk-

und Verarbeitungsprozesse wenigstens bruchstückhaft in den produzierten Lösungswegen und/oder freien Antworten sichtbar. Für den Lehrer sind nun häufig gerade diese Denk- und geistigen Verarbeitungsprozesse von Interesse, um allfällige Leistungsschwächen, Denkfehler oder besonders originelle Lösungsansätze seiner Schüler kennen zu lernen. Das richtige oder falsche Ergebnis allein gibt darüber noch keine Auskunft. Bei Mehrfachwahlaufgaben wird das Ergebnis als richtig oder falsch eingestuft, gleichgültig welche Überlegungen oder Denkfehler dazu geführt haben. Der höhere Grad an Objektivität ist mit einem Verlust an didaktisch wertvoller Information verbunden.

2.2.2. Auswertung und Quantifizierung von Leistungsergebnissen

Bei der Auswertung von Antworten und Lösungen auf gestellte Fragen oder Aufgaben stellen sich jedem Lehrer zwei Probleme: Nach welchen Kriterien soll die Auswertung erfolgen? Wie verwandelt man einen Punktwert oder sonstigen Rohwert (z. B. Fehlerzahl bei einem Diktat oder bei einer Übersetzung) in eine Note?

Richtig-Falsch-Bewertung

Im Bereich der Mathematik, des Sprach- und Sachunterrichts lassen sich erbrachte Leistungen häufig danach beurteilen, ob sie *richtig* oder *falsch* sind. Die Summe der richtigen Lösungen ergibt den Punktwert eines Schülers. Auch die absolute oder relative Fehlerhäufigkeit kommt als Messwert für bestimmte Leistungen (z. B. Diktate, Übersetzungen) in Frage. Bei dieser Art der Auswertung werden alle Aufgaben gleich gewichtet, ohne den Schwierigkeitsgrad der Aufgabe oder die Schwere der Fehler zu berücksichtigen. Viele Lehrer gewichten ihre Aufgaben nach ihrem vermuteten Schwierigkeitsgrad, indem sie für die richtige Lösung der einzelnen Aufgaben unterschiedlich viele Punkte vergeben. Sowohl für die Gewichtung von Aufgaben nach ihrer Schwierigkeit oder von Fehlern nach ihrer Schwere als auch für ihre Gleichgewichtung lassen sich Gründe und Gegengründe anführen. In der Testpraxis ist es üblich, allen Aufgaben dasselbe Gewicht zu geben, weil

sich sonst bei der Aufgabenanalyse (die für die Erstellung standardisierter Leistungstests erforderlich ist) erhebliche Schwierigkeiten ergeben. Eine vorgängige Gewichtung der Aufgaben nach ihrem vermuteten Schwierigkeitsgrad ist u. a. auch deshalb problematisch, weil die meisten Lehrer und Testkonstruktoren über keine psychologisch fundierte Theorie des Schwierigkeitsgrades einer Aufgabe verfügen.

Bei der Richtig-Falsch-Bewertung ist es häufig schwierig zu entscheiden, ob eine Antwort bzw. Lösung vollständig richtig, zur Hälfte richtig oder vollständig falsch ist. Diese Entscheidung setzt Kriterien voraus, die je nach dem Inhalt der Leistung verschieden sein können.

Beispiel eines Kriterienkatalogs für die Richtig-Falsch-Bewertung von schriftlichen Aufgaben in der Berufskunde:

Eine Aufgabe ist *vollständig richtig* gelöst, wenn

- das Resultat stimmt
- wo nötig die richtige Einheit angegeben ist
- der Lösungsgang ersichtlich ist
- die Darstellung sauber ist
- verlangte Skizzen ausgeführt sind

Eine Aufgabe ist *vollständig falsch*, wenn

- der Rechnungsgang und das Schlussresultat nicht nur ungenau, sondern falsch sind
- nur ein richtiges Schlussresultat ohne Lösungsgang vorhanden ist
- zwei verschiedene Resultate unterstrichen sind, auch wenn eines davon stimmt

Eine Aufgabe ist zur *Hälfte richtig*, wenn

- das Resultat die richtigen Ziffern aufweist, aber ein Kommafehler, gleich welcher Grösse, vorhanden ist bei sauberer Darstellung
- die Dimension fehlt
- eine verlangte Skizze fehlt, die Aufgabe sonst aber richtig gelöst ist
- Rechnungsansatz und Formel vollständig richtig sind, der Rechnungsgang aber *einen* Rechnungsfehler aufweist

(Darstellung aus dem eidg. Expertenkurs vom 28./29. August 1969 in Zürich)

Im Bereich der geisteswissenschaftlichen Fächer (z. B. Geschichte, Literaturunterricht) sind selbstverständlich andere Kriterien für eine Richtig-Falsch-Bewertung erforderlich als in mathematisch-naturwissenschaftlichen Fächern.

Aufgabenanalyse

Durch Fehlerzählung kann nachträglich auf empirischem Wege der Schwierigkeitsgrad und die Trennschärfe der einzelnen Aufgaben bestimmt werden. Die Kenntnis des Schwierigkeitsgrades und der Trennschärfe einer Aufgabe ist für den Lehrer nützlich, wenn er sich eine Aufgabensammlung (Itembank) zu thematischen Unterrichtseinheiten oder Lernzielen aufbaut. Der Schwierigkeitsgrad und die Trennschärfe einer Aufgabe stellen Testeigenschaften dar, die für die diagnostische Qualität einer Aufgabe ausschlaggebend sind. Durch eine *Aufgabenanalyse* können diese Testeigenschaften von Aufgaben ermittelt werden. Eine gute Aufgabensammlung enthält Aufgaben, die einen mittleren Schwierigkeitsgrad haben und trennscharf sind, d. h. zwischen Schülern, die die Lernziele erreicht haben, und solchen, die sie nicht erreicht haben, deutlich unterscheiden. Die Aufgabenanalyse ergibt pro Aufgabe mindestens zwei Werte: die Schwierigkeit und die Trennschärfe der Aufgabe. Der *Schwierigkeitsgrad* einer Aufgabe ist relativ leicht zu bestimmen. Er entspricht der relativen Häufigkeit richtiger Lösungen in einer Gruppe (z. B. Klasse, Gruppe von Testpersonen). Die Berechnung des Schwierigkeitsgrades (P) erfolgt nach der Formel

$$P = \frac{x}{N} \cdot 100$$

wobei x die Anzahl der richtigen Lösungen und N die Gesamtzahl der an einer Prüfung beteiligten Schüler bezeichnet. Die Ermittlung des Schwierigkeitsgrades der einzelnen Aufgaben wird sehr einfach, wenn man bei der Korrektur einer Klassenarbeit die einzelnen Resultate (richtig=1, falsch=0) in eine vorbereitete Matrix einträgt, auf deren Achse die beteiligten Schüler und auf deren anderen die einzelnen Aufgaben aufgeführt sind.

Je höher der P-Wert, desto geringer ist der Schwierigkeitsgrad einer Aufgabe und umgekehrt. Eine Aufgabe ist umso leichter, je grösser die Wahrscheinlichkeit ist, dass sie richtig gelöst werden kann.

Viel schwieriger ist es, die *Trennschärfe* einer Aufgabe zu ermitteln. Aufgaben mit guter Trennschärfe werden von Schülern, die das Lernziel erreicht haben, häufiger gelöst als von Schülern, welche es nicht erreicht

haben. Aufgaben, die von beiden Schülergruppen gleich häufig gelöst werden, besitzen keine oder nur eine geringe Trennschärfe. Die Frage über die Merkmale verschiedener Autotypen im Rahmen einer Prüfungsarbeit aus dem Gebiet der Mechanik (Physik) hat wahrscheinlich eine geringe Trennschärfe, weil ihre Beantwortung weniger von der erfolgreichen Teilnahme am Unterricht als vielmehr von ausserschulischen Interessen und Erfahrungen abhängt. Eine genaue Bestimmung der Trennschärfe einer Aufgabe nach den Anweisungen eines testtheoretischen Lehrbuches (vgl. LIENERT 1967, 93 ff.) ist so zeitraubend, dass diese Verfahren für einen praktisch tätigen Lehrer kaum in Frage kommen. Es gibt jedoch ein Näherungsverfahren, das man ohne allzu grosse Umtriebe anwenden kann (vgl. WENDELER 1972⁴, 42 ff.): Man teilt die geprüfte Klasse nach dem erzielten Gesamtpunktwert der einzelnen Schüler in eine gute und schlechtere Hälfte. Dann zählt man für jede Gruppe getrennt, wie häufig die in Frage kommende Aufgabe richtig gelöst wurde. Beträgt die Differenz zwischen der Anzahl richtiger Lösungen bei beiden Gruppen wenigstens 10 % der insgesamt möglichen richtigen Lösungen (entspricht der Gesamtzahl beider Schülergruppen), so kann die Trennschärfe einer Aufgabe als hinreichend angesehen werden.

Verwendung von Schätzskalen

Für gewisse Leistungsresultate wie Aufsätze, Zeichnungen, Entwürfe für die Lösung technischer Probleme und ähnliche Werke sind nicht in erster Linie Richtig-Falsch-Kriterien, sondern andere Beurteilungsmassstäbe anzuwenden.

- *Aufgabe:* Versuchen Sie, sinnvolle Beurteilungsmassstäbe für die Bewertung eines Aufsatzes zu definieren!

- z. B. – Klarheit des Ausdrucks!
- Flüssigkeit des Stils
- Originelle Gedanken?
- Überzeugungskraft

– Schritt für Schritt erzählt
–

Um solche und ähnliche Kriterien zu Beurteilungszwecken verwenden zu können, ist ähnlich wie bei der Richtig-Falsch-Auswer-

Abb. 2: Auswertungsblatt für Aufgabenanalyse (nach WENDELER 1972, 40)

Aufgabe \ Schüler	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Punktwert des Schülers
A	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0	0	+	+	18
B	+	+	+	+	+	+	+	+		+	+	+	+	+	+	+	0	+	0	0	16
C	+	+	+	+	+	+	+	+	+	+	0	+	0	+	+	+	+	0	+	0	16
D	+	+	0	+	+	+	+	+	+	+	0	+	+	+	+	0	0	+	0	+	15
E	+	+	0	+	+	0	+	+	+	+	+	+	+	+	+	0	+	0	0	0	14
F	+	+	+	0	+	0	+	+	+	+	0		+	+	+	+	0	+	+		14
G	+	+	+	0	+	+	+	+	+	+	+	+	+	+	+	0	0	0		+	14
H	+	+	0	+	+	+	0	+	+	+	+	+	+	+	0	0	0	+		+	14
I	+	+	0	0	+	+	0	+	+	+	+	+	+	+	+	0	+	0	0	0	13
K	+	+	+	0	+	+	+	+	+	+	0	+	0	+	+	0	0	+	0	0	13
L	+	+	0	+	+	0	+	+	+	+	+	+	0	+		+	+	0	0	0	13
M	+		+	0	+	0	+	+	+		+	+		+		+	+	+			12
N	+	+			+		+		+		+	+	+	+	0	+	+	+			12
O	+	0	+	0	+	0	0	+	+	0	+	+	+	0	0	+	+	+	0	+	12
P	0	0	+	+	+	0	+	+	+		+	+	+	+	0	0	0	+	+		12
Q	+	+	+	+	+	+	+	+	+	0	+	0		+	+	0	0				12
R	+	+	0	0		+	+	+	+	+	+	0	0	+	0	+	+	0	+		12
S	0	+	+	0	0	+	+	0	+	+	0	0	+	+	+	+	+		0	0	11
Lösungen pro Aufgabe	16	15	11	9	16	11	15	16	17	13	13	14	12	17	10	10	9	9	5	5	

tung eine Operationalisierung derselben erforderlich. Wenn es auch schwierig sein dürfte, die einzelnen Ausprägungsgrade einer Beurteilungsdimension wie «Flüssigkeit des Stils» zu operationalisieren, so sollte man wenigstens für die positive und negative Ausprägung einer Beurteilungsdimension operationale, d. h. beobachtbare Merkmale angeben.

● **Aufgabe:** Überlegen Sie sich beobachtbare Merkmale für die positive und negative Ausprägung der Beurteilungsdimension «Flüssigkeit des Stils»!

Positive Ausprägung:

- gut lesbar

-

Negative Ausprägung:

- umständliche Satzkonstruktionen

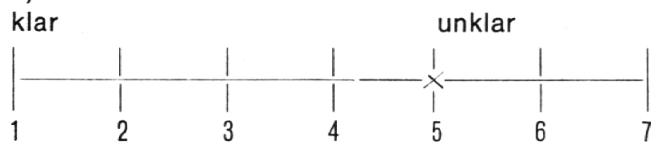
-

Die Verwendung von Schätzskalen ohne weitere Angabe operationaler Kriterien für die positive und negative Ausprägung der

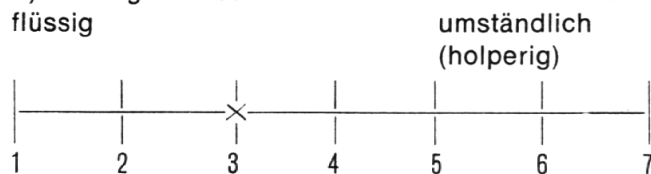
interessierenden Beurteilungsdimensionen führt zu Bewertungen und Einschätzungen, die in starkem Masse von der Person des Beurteilers und von deren unausgesprochenen Beurteilungskriterien abhängig sind.

Beispiele für Schätzskalen zur Beurteilung eines Aufsatzes:

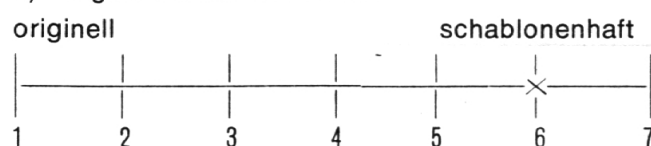
a) Klarheit des Ausdrucks



b) Flüssigkeit des Stils



c) Originalität der Gedanken



Die auf diese Weise ermittelten Werte lassen sich zu einem Durchschnittswert verarbeiten, der als quantitatives Mass der Leistung angesehen werden kann. Rohwerte oder qualitative Beschreibungen einer Leistung genügen vielfach, um den eigenen Unterricht zu evaluieren und den Schülern die notwendige Rückmeldung geben zu können. Ein Vergleich der Leistungen eines Schülers in verschiedenen Fächern zum Zwecke der Erstellung eines Leistungsprofils oder ein Vergleich der Leistungen einer Klasse mit denen einer anderen ist indessen auf der Basis von Rohwerten nicht möglich. Dazu ist ein gemeinsamer Massstab erforderlich, mit dem wir die verschiedenartigen Leistungen messen können. Die an unseren Schulen übliche *Notenskala* ist als solcher *gemeinsamer Massstab* gedacht. In der Testpsychologie wird als gemeinsamer Massstab häufig die *Prozentrangreihe* angewandt. Ein Vergleich der Leistungen erfolgt nicht notwendigerweise in selektiver Absicht, sondern kann auch für evaluative Zwecke wichtig sein. Durch Leistungsvergleiche sind Aussagen über den relativen Wirkungsgrad verschiedener Lehrmethoden oder anderer Unterrichtsvariablen möglich; Leistungsvergleiche sind aus diesem Grunde ein wichtiges Instrument der Unterrichtsforschung. Der gemeinsame Massstab allein reicht noch nicht aus, um vergleichbare Messwerte zu erhalten. Wir benötigen zudem Regeln für die Umwandlung von Rohwerten zu Werten der Einheitsskala (z. B. Notenskala). Die Umwandlung eines Rohwertes in eine Note setzt ein gemeinsames Modell voraus, um vergleichbare Werte zu erhalten. Verschiedene Modelle können, von demselben Punktwert ausgehend, zu unterschiedlichen Zensuren führen. Erst Punktwert und Beurteilungsmodell zusammen definieren eine Note.

2.2.3. Gruppenbezogene und lernzielbezogene Beurteilung

Das Urteil über eine Schülerleistung fällt unterschiedlich aus, je nachdem, welchen Bezugspunkt oder Massstab man wählt. Wählt man die Leistungen der andern Schüler als Bezugspunkt, so kann man eine Leistung als durchschnittlich, überdurchschnittlich oder unter dem Durchschnitt liegend beurteilen. Wird eine Leistung eines Schü-

lers in Relation zu den Leistungen einer Bezugsgruppe (Klassenverband, Teilnehmer an einer Prüfung) beurteilt, so spricht man in der Testpsychologie von einer gruppen- oder normbezogenen Beurteilung. Man kann eine Leistung jedoch auch danach beurteilen, ob und in welchem Grade sie sich mit einem angestrebten Lernziel deckt.

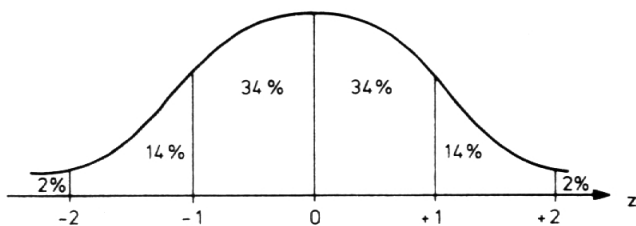
Misst man eine Leistung an einem vorgegebenen Lernziel und nicht am Klassendurchschnitt, so spricht man von *lernzielbezogener Beurteilung.*

In der Schule wie in der Testpsychologie ist das Urteil über eine Leistung in der Regel von Gruppennormen abhängig. Dies gilt beispielsweise für die *Prozentrangreihe*, eine in der Psychologie und bei Leistungstests häufig verwendete Relativskala. Durch die Angabe eines Prozentranges für eine erbrachte Leistung erfährt der Teilnehmer an einer Prüfung, wie er im Vergleich zu den anderen Teilnehmern «abgeschnitten» hat. Die Angabe «82. Prozentrang (Perzentil)» sagt aus, dass nur 18 % der Vergleichsgruppe einen höheren Testwert erzielt haben als der Proband im 82. Prozentrang. Der Prozentrang ist ein genaueres Mass für die relative Position einer Leistung als eine Notenstufe, weil die zugrunde liegende Skala hundert und nicht nur sieben oder fünf Stufen enthält. Die Anwendung der Prozentrangskala empfiehlt sich vor allem für die Beurteilung von Prüfungsergebnissen bei einer grossen Teilnehmerzahl (z. B. Beurteilung der Leistung im Schlusstest bei Funkkollegs). Für die Umwandlung eines Rohwertes in einen Prozentrang sei auf die üblichen statistischen Lehrbücher verwiesen (z. B. HELLER/ROSEMANN 1974, 177 ff.).

Auch die Notenskala wird in der Regel als Relativskala behandelt, die auf dem Klassendurchschnitt und der Normalverteilung von Leistungen gründet. Die *klasseninterne Normierung* schulischer Noten zeigt sich u. a. darin, dass für dieselbe Leistung in verschiedenen Klassen unterschiedliche Noten vergeben werden je nachdem, ob der Klassendurchschnitt hoch oder niedrig ist. Der einzelne Lehrer orientiert sich bei der Beurteilung an einem klasseninternen Massstab (vgl. INGENKAMP 1971, 156 ff.). Dabei dient der Leistungsdurchschnitt einer Klasse als Bezugspunkt für die Zuteilung der einzelnen Notenstufen. Ein weiteres Merkmal

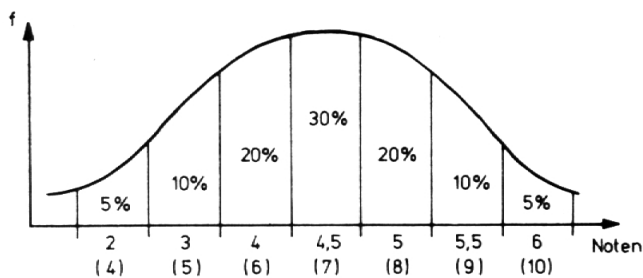
ist auffallend: Die mittlere Notenstufe wird am häufigsten vergeben, höhere und tiefere Noten werden weniger häufig und die Extremwerte eher selten vergeben. Statistisch ausgedrückt: Die Häufigkeiten verteilen sich näherungsweise *normal* auf die einzelnen Notenstufen. Die Normalverteilung stellt eine standardisierte Verteilungsform dar, bei der die Häufigkeiten um den Mittelwert symmetrisch abnehmen, so dass sich im Idealfall eine Glockenform ergibt (GAUSSsche Glockenkurve).

Abb. 3: Standardisierte Normalverteilung



Dieses mehr oder weniger intuitive Vorgehen bei der Notenvergabe kann objektiviert werden, indem man für die Umwandlung von Rohwerten in Noten feste Regeln angibt. Im Falle der gruppenbezogenen Beurteilung werden die relativen Notenanteile festgelegt, die den einzelnen Notenstufen in Anlehnung an die Normalverteilung zukommen. Nach dieser prozentualen Verteilung kann dann die Zuweisung von Rohproduktwerten zu Noten erfolgen. Abbildung 4 zeigt die relativen Notenteile, die sich aufgrund der Normalverteilung bei einer siebenstufigen Notenskala ergeben.

Abb. 4: Relative Notenanteile bei einer siebenstufigen Notenskala



(Die in Klammern stehenden Ziffern repräsentieren die Stufen der italienischen Notenskala.)
Ausgehend von dieser relativen Notenverteilung bringt man im Falle der gruppenbezogenen Beurteilung die vorliegenden Roh-

werte in eine Rangreihe und teilt diese Rangreihe entsprechend der Anzahl der verfügbaren Notenstufen in ebenso viele Intervalle ein, die die jeweils definierten Prozentanteile umfassen: Die obersten 5 % der Rangreihe erhalten die Note 6, die nächsten 10 % die Note 5^{1/2} usw. Bei dieser Art der Notengebung erhält man aus Symmetriegründen den gleichen Anteil der Noten 2 und 6. Die mittlere Notenstufe 4^{1/2} ist am häufigsten vertreten.

Dieses Modell der Notenzuweisung ist in pädagogischer wie methodologischer Hinsicht fragwürdig . Die Orientierung an der Normalverteilung hat bei der Vergabe von Noten die fragwürdige Konsequenz, dass bei jeder Beurteilung unabhängig von der absoluten Leistungshöhe nach einem strengen mathematischen Kalkül ein gewisser Prozentsatz an Versagern definiert wird. Dieses Beurteilungsmodell beruht ferner auf der unzulässigen Annahme, dass die Leistungen (Rohwerte) der Schüler einer Klasse wirklich annähernd normal verteilt sind. Dies gilt zwar für Leistungen bei Intelligenztests sowie für die Häufigkeitsverteilung von Körpergrößen und Gewichtsmassen (z. B. bei Rekruten), nicht aber notwendigerweise für Schulleistungen. Schulleistungen sind das Ergebnis einer gezielten Einflussnahme durch Unterricht und folglich in der Regel nicht normal verteilt, es sei denn, dass in erster Linie die Intelligenz der Schüler für ihren Schulerfolg ausschlaggebend wäre, was nur bei einem «schlechten» Unterricht zutrifft. Die auf diese Weise gewonnenen Noten sind zudem kaum vergleichbar, weil sie auf unterschiedliche Bezugsgruppen abgestimmt sind und somit je etwas anderes bedeuten.

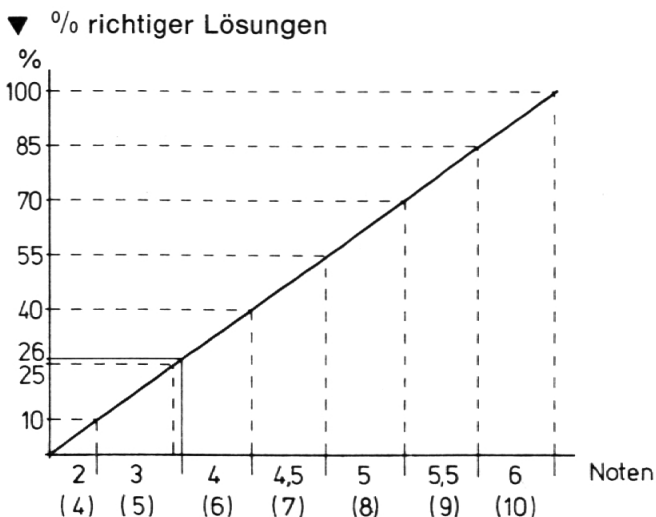
Bei der lernzielbezogenen Beurteilung wird die Schülerleistung nicht am Klassendurchschnitt, sondern an einem vorgegebenen Lernziel gemessen. Urteile wie «gut», «befriedigend» und «ausreichend» sind Ausdruck einer lernzielbezogenen Beurteilung, während Urteile wie «durchschnittlich», «überdurchschnittlich» usw. Ausdruck einer gruppenbezogenen Beurteilung sind. Bei der lernzielbezogenen Beurteilung spielt es keine Rolle, welche Leistungen die anderen Schüler erbracht haben: Ein Schüler wird für eine bestimmte Leistung immer dieselbe Note erhalten, ob er nun insgesamt zu einer

«guten» oder zu einer «schlechten» Klasse gehört (nach WENDELER 1972⁴, 72).

Für die Anwendung eines lernzielbezogenen Beurteilungsmodells gelten folgende Verfahrensregeln:

- In einem ersten Schritt wird derjenige Punktwert oder diejenige Leistungshöhe definiert, die ein Schüler mindestens erreichen muss, um eine positive Note zu erhalten (Erfolgskriterium).
- Die oberhalb und unterhalb dieser Marke liegenden prinzipiell möglichen Rohwerte oder Leistungsniveaus werden entsprechend der Anzahl der verfügbaren Notenstufen in Intervalle unterteilt. Die Intervalle können gleich gross (lineare Skala) oder zunehmend kleiner (degressive Skala) gewählt werden.
- Jedem Intervall wird entsprechend der Rangfolge eine Note zugeordnet.
- Durch den Vergleich des durch einen Schüler erreichten Punktwertes (bzw. der erreichten Leistungshöhe) mit der vorliegenden Punkte-Noten-Skala kann die gesuchte Note leicht ermittelt werden.

Abb. 5: Beispiel eines linearen lernzielbezogenen Beurteilungsmodells (Nomogramm)



► Erfolgskriterium

(Die in Klammern stehenden Ziffer repräsentieren die Stufen der italienischen Notenskala.)

Das Verfahren der lernzielorientierten Beurteilung ist im Rahmen der üblichen Schulprüfungen geeigneter als gruppenorientierte Beurteilungsmodelle (Prozentrangskala, Normalverteilung). Bei diesem Beurteilungsmodell ist es nicht mehr vom Klassendurch-

schnitt abhängig, welche Note man als Schüler erhält, sondern davon, ob und in welchem Grade man ein angestrebtes Lernziel erreicht hat. Eine so gewonnene Note ist Ausdruck des Lernstandes des Schülers und nicht seiner relativen Leistungsfähigkeit. Sie zeigt, wie erfolgreich sich der Schüler am Unterricht beteiligt hat bzw. wie wirksam der erteilte Unterricht für ihn war.

3. Testtheoretische Anforderungen an Prüfungen

Aus der Sicht der Testtheorie muss *eine methodisch einwandfreie Prüfung objektiv, zuverlässig und gültig sein*. Bei «hausgemachten» Prüfungen kann man im Gegensatz zu standardisierten Leistungstests diese Gütekriterien nur *näherungsweise* erreichen. Zu diesem Zwecke sind bei den einzelnen Schritten der Leistungsbeurteilung ganz bestimmte Bedingungen einzuhalten, die im folgenden erläutert werden.

3.1. Objektivität einer Prüfung

«Unter Objektivität eines Tests verstehen wir den *Grad, in dem die Ergebnisse eines Tests unabhängig vom Beurteiler sind*. Ein Test wäre demnach völlig objektiv, wenn verschiedene Untersucher bei demselben Probanden zu gleichen Ergebnissen gelangten. Man spricht deshalb auch von ‚interpersoneller Übereinstimmung‘ der Untersucher» (LIENERT 1961, 13).

Mögliche Ursachen für allfällige Nicht-Übereinstimmungen der Prüfungsergebnisse können sowohl in der Art der Durchführung einer Prüfung als auch in der Auswertung der Leistungsresultate liegen. Die *Durchführungsobjektivität* kann durch zufällige oder systematische Verhaltensvariationen des Prüfers (z. B. Art der Verstärkung, Hilfestellungen) im Falle von mündlichen Prüfungen oder durch ungleiche Prüfungsbedingungen (z. B. erlaubte und unerlaubte Hilfsmittel) beeinträchtigt werden. Um eine möglichst hohe Durchführungsobjektivität zu gewährleisten, ist auf seiten des Lehrers ein hohes Mass an Selbstkontrolle und eine eindeutige Regelung der Prüfungsbedingungen erforderlich.

Die *Auswertungsobjektivität* ist bei Schulprüfungen insbesondere dann beeinträchtigt, wenn klare Beurteilungsmassstäbe fehlen.

Dies führt dazu, dass jeder Prüfer oder Korrektor seine individuellen Gütekriterien anwendet, die bekanntlich mehr oder weniger stark divergieren. Das Fehlen eindeutiger Beurteilungsmassstäbe begünstigt das Auftreten verzerrender sozialpsychologischer Effekte (Halo-Effekt, Stereotype, Sympathie bzw. Antipathie ...) (vgl. WEISS 1965).

Eine möglichst genaue Definition der Prüfungsanforderungen, der Beurteilungskriterien und der Regeln für die Umwandlung eines Rohwertes in ein Prädikat oder in eine Note vor der Durchführung und Auswertung der Prüfung kann ihre Objektivität erheblich steigern. Besondere Schwierigkeiten bereitet die Formulierung klarer Beurteilungskriterien. Schon bei der Beurteilung von Lösungen mathematischer Aufgaben ist es häufig schwierig, zu entscheiden, ob eine Lösung als richtig, teilrichtig oder falsch einzustufen ist (vgl. dazu Abschnitt 2.2.2. dieses Beitrages). Eine gerechte, d. h. objektive Beurteilung eines Aufsatzes oder einer Zeichnung bereitet noch viel grössere Schwierigkeiten. Trotzdem lässt sich auch bei derartigen komplexen Leistungen die Beurteilung objektivieren, indem man die Beurteilungsmassstäbe festlegt und diese operationalisiert.

Durch die Verwendung von Aufgaben mit gebundener Beantwortungsform (Mehrfachwahlaufgaben, Zuordnungsaufgaben) anstelle von Aufgaben mit freier Beantwortungsform kann die Auswertungsobjektivität erheblich gesteigert werden. Auf die damit verbundenen Nachteile (Informationsverlust) wurde bereits unter 2.2.1. hingewiesen.

3.2. Zuverlässigkeit (Reliabilität) einer Prüfung

«Unter Zuverlässigkeit eines Tests verstehen man den *Grad der Genauigkeit, mit dem er ein bestimmtes Persönlichkeits- oder Verhaltensmerkmal misst*.... Ein Test wäre demnach vollkommen reliabel, wenn die mit seiner Hilfe erzielten Ergebnisse den Probanden genau, d. h. fehlerfrei beschreiben bzw. auf der Testskala lokalisieren» (LIE-NERT 1961, 14 f.).

Die Zuverlässigkeit einer Prüfung oder eines Tests bezieht sich auf das *Wie* der Messung, unabhängig vom *Was* oder Gegenstand der Messung. Ein Beispiel: Ein Gummiband

wäre ein wenig zuverlässiger Massstab für die Länge eines Gegenstandes. Ebenso wäre eine Prüfung wenig zuverlässig, wenn der Prüfling bei gleichen Prüfungsanforderungen zu einem anderen Zeitpunkt andere Ergebnisse erzielte. Schüleraussagen wie «Ich habe Pech bzw. Glück gehabt» sind ein Ausdruck dafür, dass Prüfungsergebnisse häufig recht zufällig zustande kommen und also unzuverlässig sind.

Wegen der geringen Zuverlässigkeiten der meisten Prüfungen genügt es nicht, einfach die Punktwerte oder Noten anzusehen, um daraus seine Schlüsse über die Leistungsfähigkeit oder den Lernstand eines Prüflings abzuleiten. Wenn der eine Schüler 15 und der andere 10 Punkte bei der gleichen Prüfung erzielt hat, so kann man daraus noch nicht ohne weiteres folgern, dass der eine besser als der andere ist. Um dies tun zu können, müsste die Messung völlig zuverlässig bzw. reliabel sein. Ein in einer Prüfung erzieltetes Testergebnis kann im Einzelfall durch viele Störfaktoren beeinflusst und deshalb in seiner Zuverlässigkeit beeinträchtigt sein.

Mögliche Ursachen, die die Zuverlässigkeit einer Prüfung beeinträchtigen, sind:

- Faktoren, die im *Schüler* selbst liegen (Ablenkung, Ermüdung, schlechter Gesundheitszustand ...)
- Faktoren, die in der *äusseren Situation* liegen (schlechte Lüftung, Hitze, Materialdefekt z. B. beim Rechenschieber ...)
- Faktoren, die mit der *Formulierung oder Auswahl der Aufgaben* zusammenhängen
- Faktoren, die im *Beurteiler* liegen (z. B. Müdigkeit, Zerstreuung bei der Auswertung).

Eine Prüfung gilt dann als zuverlässig bzw. reliabel, wenn bei einer Wiederholung der Prüfung in etwa gleiche Leistungsergebnisse erzielt würden (Wiederholungsreliabilität).

Dies wäre ein Beweis dafür, dass sie von individuellen und situativen Störfaktoren weitgehend unabhängig ist. Besondere Aufmerksamkeit ist der Formulierung und Auswahl der Aufgaben zu widmen, Unklar oder mehrdeutig formulierte Aufgaben bzw. Fragen sowie einseitig ausgewählte Prüfungsaufgaben können die Zuverlässigkeit einer Prüfung erheblich beeinträchtigen. Durch eine möglichst sorgfältige Formulierung der Prüfungsfragen bzw. -aufgaben sowie durch

eine möglichst *repräsentative Auswahl* derselben kann die erwünschte *Fassungsunabhängigkeit* einer Prüfung erreicht werden. Die *Beurteilerunabhängigkeit* einer Prüfung wurde bereits unter dem Stichwort Objektivität erläutert.

Eine in jeder Hinsicht gewährleistete Zuverlässigkeit einer Prüfung ist schwer zu erreichen. Trotzdem ist es gut, dass man sich als Lehrer immer wieder Rechenschaft ablegt, welche Faktoren Prüfungsergebnisse verzerren können bzw. bei einer konkreten Prüfung wirksam gewesen sein mögen. In jedem Fall ist der erzielte Punktwert eines Prüflings mit Vorsicht zu interpretieren. Für gewichtige Entscheidungen (Versetzung) empfiehlt es sich, aus den eben genannten Gründen, die Ergebnisse mehrerer Prüfungen als Grundlage für ein Leistungsurteil heranzuziehen. Vom Standpunkt der Testtheorie aus sind Prüfungsformen zu bevorzugen, die eine mehrmalige Leistung und eine breite Erfahrungsgrundlage für ein Leistungsurteil ermöglichen. Angewandt auf Abschlussprüfungen (z. B. Maturität) bedeutet dies, dass die im Laufe der Schulzeit erbrachten Leistungen eine viel zuverlässigere Beurteilungsbasis darstellen als eine einzelne punktuelle Prüfung. Die Anforderungen der Testtheorie sprechen für die sogenannte «Erfahrungsnote».

3.3. Gültigkeit (Validität) einer Prüfung

«Die Validität eines Tests gibt den *Grad der Genauigkeit an, mit dem dieser Test dasjenige Persönlichkeitsmerkmal oder diejenige Verhaltensweise, das (die) er messen soll oder zu messen vorgibt, tatsächlich misst.* Ein Test ist demnach vollkommen valide, wenn seine Ergebnisse einen unmittelbaren und fehlerfreien Rückschluss auf den Ausprägungsgrad des zu erfassenden Persönlichkeits- oder Verhaltensmerkmals zulassen . . . » (LIENERT 1961, 16 f.).

Während die Zuverlässigkeit eines Tests etwas über das *Wie* der Messung aussagt, handelt es sich bei der Gültigkeit um das *Was*. Dazu ein Beispiel aus dem Bereich der Technik: Ein Manometer misst in erster Linie den Druck und nicht die Temperatur eines Reifens. Im Hinblick auf die Messung der Temperatur hat das Manometer nur geringe Validität. Unter besonderen Umstän-

den kann man zwar vom gemessenen Druck auf die herrschende Temperatur schliessen, in der Regel aber genügt diese Angabe nicht. Druckdifferenzen können Ausdruck von Temperaturdifferenzen sein, müssen es aber nicht notwendigerweise sein. Deshalb vermag ein Manometer keinen hinreichenden Aufschluss über die Temperatur zu geben; seine Validität für die Temperaturmessung ist beschränkt. Ähnlich verhält es sich manchmal bei Prüfungen: Es kann vorkommen, dass das Prüfungsergebnis viel eher Auskunft über die psychische Belastbarkeit eines Schülers gibt als über den Stand seiner fachlichen Kenntnisse. Manchmal sind Noten eher ein Mass für die beredete Wendigkeit oder die Originalität beim «Schwindeln» und nicht so sehr für die fachliche Leistungsfähigkeit. In diesen Fällen wäre das Prüfungsergebnis von geringer Gültigkeit hinsichtlich der fachlichen Kompetenz oder des tatsächlichen Lernstandes des Geprüften. Es ist wichtig zu wissen, dass es recht verschiedene Arten von Gültigkeit gibt, die je nach Prüfungstyp von unterschiedlicher Bedeutung sind.

3.3.1. Inhaltliche Validität einer Prüfung

Im Falle von Klassenprüfungen bemisst sich die inhaltliche Validität daran, ob die gewählten Aufgaben mit den stofflichen Zielen des Unterrichts übereinstimmen. Werden Fähigkeiten und Kenntnisse mitgeprüft, welche nicht Gegenstand des Unterrichts waren, so ist die inhaltliche Validität einer Prüfung beeinträchtigt. Je besser eine Prüfung oder die einzelnen Aufgaben sich mit den inhaltlichen Zielen des Unterrichts decken, desto besser ist ihre inhaltliche Validität. Die Operationalisierung der angestrebten Lernziele erleichtert das Formulieren inhaltlich valider Prüfungsaufgaben. Auf der anderen Seite wird die inhaltliche Validität einer Prüfung beeinträchtigt, wenn in den verwendeten Aufgaben oder Fragestellungen Begriffe vorkommen, welche die Schüler noch nicht kennengelernt haben oder wenn ihre Lösung bzw. Beantwortung Vorkenntnisse voraussetzt, deren Einführung im Unterricht sehr lange zurückliegt und die nicht mehr wiederholt worden sind.

3.3.2. Konstruktvalidität

Wir haben weiter oben Lernergebnisse als

überdauernde Verhaltensbereitschaften oder kognitive Strukturen charakterisiert, die nur auf indirektem Wege über spezifische Leistungen erschlossen werden können. Die in einer Prüfung geforderten Leistungen sind also Kennzeichen (Indikatoren) überdauernder Verhaltensbereitschaften (Dispositionen) oder kognitiver Strukturen, die im Unterricht aufgebaut wurden. Begriffe, die nicht direkt beobachtbar sind, jedoch zur Erklärung beobachtbaren Verhaltens (z. B. Leistungen) herangezogen werden (z. B. Motivation, Intelligenz, Kognitive Struktur, Intra-Extraversion) bezeichnet man in der Psychologie als *hypothetische Konstrukte*. Auch andere Wissenschaften kennen hypothetische Konstrukte, so z. B. die Physik und Chemie, deren Begriffssysteme zahlreiche hypothetische Konstrukte aufweisen (z. B. Kraft, Gravitation, Molekulargewicht, Atom, Elektron, Magnetisches Feld).

Von *Konstruktvalidität* einer Prüfung spricht man dann, wenn durch sie jene kognitiven Strukturen und Fähigkeiten gemessen werden, die man zu messen beabsichtigt. Dazu ein Beispiel: Die bloße Wiedergabe (Reproduktion) eines auswendig gelernten Lehrsatzes oder einer memorierten Formel ist kein valider Indikator für das Verständnis und die Transparenz der den Lehrsätzen oder Formeln zugrunde liegenden Bedeutungen. Reproduktive Leistungen sind wohl valide Indikatoren für Wissen, nicht aber notwendigerweise für Verständnis. Welche Leistungen sind valide Indikatoren für Verständnis? Um solche Leistungen bestimmen zu können, benötigen wir eine Theorie des Verständnisses. Die Gestaltpsychologie charakterisiert Verständnis mit dem Begriff der *Einsicht*. Einsicht ist dann vorhanden, wenn ein Sachzusammenhang (eine Sachstruktur) – bildlich gesprochen – dem Schüler transparent oder durchsichtig geworden ist. AEBLI (1963, 62 ff.) betont unter Bezugnahme auf PIAGET den «operativen» Charakter kognitiver Strukturen, die sich darin vor allem von sogenannten *Automatismen des Verhaltens* (z. B. auswendig gelernte Formeln und mechanisierte Lösungsverfahren) unterscheiden. Automatismen sind stereotype und starre Verhaltensweisen, deren richtiger Ablauf von spezifischen äusserlichen Bedingungen abhängt (auslösende Reize). Operationen sind dagegen vor al-

lem durch ihre *Beweglichkeit* und ihre vielseitige *Anwendbarkeit* gekennzeichnet. Die Beweglichkeit der Operation besteht in ihrer *Umkehrbarkeit* (Reversibilität) und *Variierbarkeit* (Assoziativität). Automatismen (Gewohnheiten) bilden verhältnismässig isolierte Verhaltensweisen, während Operationen Gesamtsysteme bilden. Nach dieser Theorie kann die Verfügbarkeit von verstandenen Begriffen und Operationen durch Aufgaben überprüft werden, welche die Umkehrung von Operationen und die Anwendung von Begriffen und Operationen in neuen Sachzusammenhängen verlangen. Für die Prüfung des Verständnisses eignen sich nach AEBLI (1971⁷, 185 ff.) folgende Leistungen:

- *Wiedergabe* memorierter Texte (z. B. Lehrsätze, Formeln, Regeln) *mit eigenen Worten*
- *Wiedererkennen* bekannter Strukturen in neuen Zusammenhängen
z. B.: Erklären können, warum eine Last mit dem Flaschenzug mit viel weniger Kraftaufwand gehoben werden kann als mit einer einfachen Rolle ($\text{Arbeit} = \text{Kraft} \times \text{Weg}$).
- *Erklären von Zusammenhängen* und nicht Wiedergabe von Einzelwissen
z. B.: Gründe und Folgen eines historischen Ereignisses angeben; Gegenüberstellung bzw. Vergleich zweier Phänomene; Beweggründe und Absichten menschlichen Handelns aufzeigen.
- *Umkehrleistungen* (Transformation einer gelernten Struktur)
z. B.: Durchdenken eines Kausalzusammenhanges von den Folgen zu den Ursachen aufsteigend, anstelle des Durchdenkens dieses Zusammenhanges von den Ursachen zu den Folgen absteigend.
Berechnen der Seite eines Rechteckes bei gegebener Fläche und der anderen Seite (Umkehroperation zur üblichen Flächenberechnung des Rechteckes).
- *Anwendungsleistungen*: Das Anwenden gelernter Begriffe und Operationen auf neue Sachverhalte stellt den besten Beweis für das Vorhandensein von Verständnis dar. Echte Anwendungsaufgaben dürfen jedoch keine signalartigen Ausdrücke enthalten (z. B. Berechne die Fläche!), die es dem Schüler gestatten, die Aufgabe auch aufgrund blinder Automatismen zu lösen.

Beispiele für Anwendungsleistungen:

Übersetzen eines Textes in eine andere Sprache (Anwenden grammatikalischer und lexikalischer Kenntnisse)

Lösen von Problemen mit Hilfe erworbener Verhaltensschemata (z. B. im Bereich der Arithmetik, der angewandten Geometrie, der Naturwissenschaften)

Nennen von Beispielen für eine bekannte begriffliche Struktur (z. B. Geräte des Alltags, die nach dem Hebelprinzip funktionieren)

- **Aufgabe:** Entwickeln Sie eine geeignete Aufgabe oder Frage, um das Verständnis eines eingeführten Begriffs oder Prinzips (Hebelprinzip, Goldene Regel der Mechanik, Nahrungskette in der Biologie, Elektroanalyse in der Chemie, Begriff des Mehrwertes bei MARX u. ä.) zu überprüfen!

3.3.3. Voraussagevalidität

Bei Eignungs- und beruflichen Befähigungsprüfungen hat die Prüfungsleistung die Funktion eines *Stellvertreters* für die Leistungsfähigkeit im nachfolgenden Lernen oder in der beruflichen *Ernstsituation*. Grundsätzlich sollte die Prüfung, um die wünschenswerte Voraussagevalidität zu erhalten, möglichst identisch mit der Ernstleistung sein (z. B. Auto-Fahrprüfung im normalen Verkehr; Lösen von Problemen, wie sie auch in der Ernstsituation vorkommen). In vielen Fällen ist die Identität aus technischen und zeitlichen Gründen jedoch schwer herzustellen.

- **Aufgabe:** Überlegen Sie sich, ob folgende Prüfungsleistungen valide Stellvertreter der Ernstsituation sind!

Prüfungsleistung

- Probelektion eines Lehrerstudenten
- Abschneiden in der Lehrabschlussprüfung
- Leistung in der Aufnahmeprüfung

Ernstleistung

- Schulführung (Unterricht)
- Tüchtigkeit im Beruf
- Bewährung im nachfolgenden Lernen

3.4. Eichung (Normierung) einer Prüfung

Ein Test wird *geeicht*, indem dieser einer grossen Anzahl von Personen vorgelegt wird, die für eine bestimmte Bezugsgruppe (Altersgruppe, Klassenstufe, Geschlecht) repräsentativ ist. Aufgrund der in der repräsentativen Stichprobe gewonnenen Ergebnisse wird eine Prozentrangskala entwickelt, an der die Einzelresultate, die mit diesem Test in der späteren Praxis gewonnen werden, gemessen werden können. Auf diese

Weise kann jeder gewonnene Testwert hinsichtlich seiner Position innerhalb der Gesamtheit der möglichen Testwerte der relevanten Bezugsgruppe eingestuft werden (auf der Prozentrangskala). Die auf dem Markt erhältlichen Schulleistungstests (vgl. Testprogramm des Beltz- und Klett-Verlages) sind in der Regel für eine bestimmte Schulstufe und Region geeicht.

Für normale schulische Prüfungen ist eine Normierung im strengen Sinne des Wortes in den seltensten Fällen möglich und sinnvoll, wohl aber für regionale oder überregionale Prüfungen (z. B. Lehrabschlussprüfungen). Im Rahmen des schulischen Alltags ist es sinnvoll und nützlich, entwickelte Prüfungen mit Kollegen der gleichen Schulstufe und des gleichen Faches zu besprechen und auszutauschen, um den klasseninternen Massstab zu relativieren bzw. auszugleichen. Ein Vergleich der Leistungen der eigenen Klasse mit denen der Klasse des Kollegen oder einer noch grösseren Bezugsgruppe gibt Aufschluss darüber, wo die eigene Klasse hinsichtlich der fachlichen Leistungsfähigkeit innerhalb einer grösseren Bezugsgruppe steht.

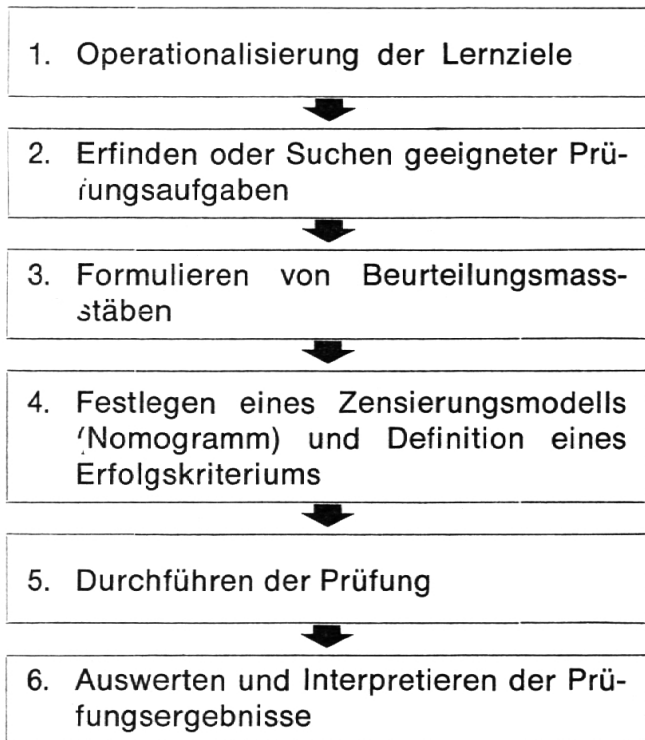
4. Arbeitsschritte bei der Entwicklung und Durchführung von Erfolgskontrollen des Unterrichts

Erfolgskontrollen im Unterricht können sich auf in sich geschlossene thematische *Unterrichtseinheiten* innerhalb eines Faches oder auf den Lehrstoff eines längeren Zeitraumes beziehen. Bei der Entwicklung und Durchführung solcher Erfolgskontrollen lassen sich sechs *Arbeitsschritte* unterscheiden, die allesamt für die Aussagekraft einer Prüfung wichtig sind (vgl. Abbildung 6).

4.1. Operationalisierung der Lernziele

Da Lernergebnisse stets nur indirekt über beobachtbare Leistungen bzw. Verhaltensweisen erfasst und gemessen werden können, sind in einem ersten Schritt jene beobachtbaren Verhaltensweisen bzw. Leistungen festzuhalten, die für die angestrebten Lernergebnisse (kognitiven Strukturen) repräsentativ sind. Aus operationalen Lernzielbeschreibungen ist zu entnehmen, was der Lernende *tun* oder *ausführen* können

Abb. 6: Abfolge der Arbeitsschritte bei der Entwicklung und Durchführung von Erfolgskontrollen des Unterrichts



muss, um zu zeigen, dass er das Ziel erreicht hat (nach MAGER 1969³, 13 ff.). Die Beschreibung der Lernziele in Begriffen des *Endverhaltens* ist eine notwendige Voraussetzung für die Konstruktion geeigneter Prüfungsaufgaben. Nicht jede Verhaltensweise oder Leistung ist jedoch geeignet, die Verfügbarkeit einer kognitiven Struktur anzuzeigen (vgl. Abschnitt 3.3.2.). Verhaltensweisen wie «aufzählen», «nennen» u. ä. m. sind recht unzuverlässige Indikatoren für die Einsicht in Sachzusammenhänge. Von einem Schüler, der die grundlegenden Zusammenhänge des biologischen Gleichgewichts verstanden hat, erwarten wir, dass er die Folgen der Vergasung von Füchsen im Zusammenhang mit der Tollwutbekämpfung aufzeigen kann. Diese Leistung ist eine mögliche Operationalisierung der kognitiven Struktur «Das biologische Gleichgewicht».

- **Aufgabe:** Suchen Sie für Lerninhalte aus Ihrem Unterrichtsfach Leistungen, die eine Operationalisierung dieser Inhalte darstellen.

4.2. Konstruktion geeigneter Prüfungsaufgaben

In einem zweiten Schritt sind solche Prüfungsaufgaben zu erfinden oder (aus Lehr-

büchern) zu suchen, deren Lösung die in den operationalen Lernzielbeschreibungen enthaltenen Leistungen *in möglichst direkter Weise* verlangt. Der Grundsatz der *direkten Erfassung von Leistungen* (AEBLI 1971⁷, 180) ist insbesondere bei der Prüfung praktischer Fertigkeiten schwer zu verwirklichen, weil dafür in der Regel umfangreiche Vorbereitungen notwendig sind. So gibt es beispielsweise für die *Lehrerbildung* keine bessere und direktere Prüfung der Lehrbefähigung als den Unterricht in einer normalen Klasse. Die praktische Durchführung von Prüfungslektionen erfordert jedoch so viel Vorbereitungsarbeit, dass viele Institutionen der Lehrerbildung sich darauf beschränken, die Lehrbefähigung indirekt anhand verbaler Fragen über den Unterricht zu erfassen. Ein weiteres Beispiel für die indirekte Erfassung einer Leistung sind die im traditionellen Fremdsprachenunterricht üblichen Übersetzungen muttersprachlicher in fremdsprachliche Texte und umgekehrt. Hier wird der muttersprachliche Ausdruck als Bedeutungsträger zwischen den fremdsprachlichen Ausdruck und der Sachvorstellung eingefügt, was dem Ziel des direkten Fremdsprachenunterrichts widerspricht. Ziel des Fremdsprachenunterrichts ist es, den Schüler zu befähigen, aus der Situation oder Vorstellung heraus direkt (ohne den Umweg über die Muttersprache) den fremdsprachlichen Ausdruck zu finden. Aus diesem Grunde sollten anstelle von Übersetzungen viel eher Bildbeschreibungen, Beschreibungen von vorgestellten Szenen und kleine Aufsätze mit scharf umrissenen Themen geschrieben werden (nach AEBLI a. a. O., 181 f.).

Im Bereich der natur- und geisteswissenschaftlichen Fächer ist der Grundsatz der direkten Erfassung von Leistungen leichter zu verwirklichen, weil sich diese in der Regel im symbolischen Medium der Sprache repräsentieren. Wenn wir die Schüler die Folgen einer bestimmten Massnahme (z. B. Vergasen der Füchse *voraussagen* oder das Auftreten eines natürlichen oder gesellschaftlichen Phänomens (z. B. Landflucht) *erklären* oder zwei Werke *vergleichen* lassen, so geschieht das immer im Medium der Sprache. Dies gilt auch für das *Zuordnen* eines Werkes oder einer These zu ih-

rem Autor oder zu einer geistesgeschichtlichen Richtung.

Ein anderer Grundsatz besagt, dass Prüfungen eher eine Vielzahl kurzer, voneinander unabhängiger Aufgaben enthalten sollen als Aufgaben, die lange, zusammenhängende Ketten von Handlungen und Operationen erfordern (nach AEBLI a. a. O., 184). Solche komplexen Prüfungsaufgaben vermitteln leicht ein falsches Bild von der Leistungsfähigkeit eines Prüflings, da Fehler in der ersten Phase der Arbeit den Fortgang der Lösung bzw. Antwort belasten und den Schüler verwirren können. Dies gilt insbesondere für mathematische Aufgaben, deren Ergebnis über eine Kette von Zwischenresultaten erreicht wird.

Wenn das Wissen und die Denkfähigkeit eines Schülers in einem bestimmten Sachgebiet (Naturwissenschaften, Geographie, Geschichte, Sozialkunde) geprüft werden soll, so ist es vorteilhafter, eine Reihe eng gefasster Probleme zu stellen, die kurze Antworten (bis zu 100 Worten) erfordern als weite Themenstellungen, die einen langen Aufsatz verlangen. Die auf diese Weise gewonnene Stichprobe von Leistungen ist repräsentativer für den zu prüfenden geistigen Bereich als ein einziger «Sondenstich» in der Form eines Aufsatzes oder Referates. Gut konstruierte Leistungstests enthalten aus diesem Grunde eine Vielzahl kurzer, voneinander unabhängiger Aufgaben.

Bezüglich des *Schwierigkeitsgrades* von Aufgaben habe ich bereits darauf hingewiesen, dass eine *Vielzahl* voneinander unabhängiger Aufgaben *mittleren Schwierigkeitsgrades* eine sichere Aussage über die fachspezifische Leistungsfähigkeit eines Schülers erlauben als wenige Aufgaben *gestaffelter Schwierigkeit*.

4.3. Formulieren von Beurteilungsmassstäben

Die Beurteilung der in Prüfungen gezeigten Leistungen beruht auf Kriterien, die im Interesse der Transparenz und Objektivität einer Prüfung offenzulegen sind. Am häufigsten werden erbrachte Leistungen danach beurteilt, ob sie *richtig oder falsch* sind. Wo es nicht völlig eindeutig ist, welche Lösung oder Antwort als richtig, teilrichtig oder falsch einzustufen ist, sollte im Falle von

schriftlichen Prüfungen eine operationale Definition der Merkmale richtiger und falscher Lösungen angegeben werden (vgl. Abschnitt 2.2.2.). Werden bei Aufgaben mit freier Beantwortungsform Punkte vergeben, so ist von vorneherin festzulegen, wieviele Punkte für die einzelnen Aufgaben zugeteilt werden können und nach welchen Kriterien die Zuteilung erfolgt. Bei Sachfragen richtet sich die Zuteilung von Punkten häufig nach der Anzahl richtiger Elemente in der Antwort; in anderen Fällen bildet die Fehlerzahl die Grundlage für die Ermittlung eines Punktwertes (z. B. bei Diktaten, Übersetzungen). Da sich bei der Vergabe von Punkten besonders leicht Verzerrungen ergeben, hervorgerufen durch Halo-Effekte oder Vorurteile, spielen klare Beurteilungskriterien für die Objektivität der Beurteilung eine wichtige Rolle. Ohne klare Beurteilungskriterien kann es sich auch bei quantifizierenden Beschreibungen von Prüfungsergebnissen um scheinobjektive Beurteilungen handeln.

Nicht jede Leistung kann nach dem Richtig-Falsch-Kriterium beurteilt werden. Dies gilt beispielsweise für praktische Fertigkeiten sowie kreativ-schöpferische Leistungen. Im Falle des Maschinenschreibens ist die Zahl der richtigen Anschläge pro Zeiteinheit das entscheidende Beurteilungskriterium. Daneben spielen auch das Schriftbild und die Gestaltung einer Textseite eine gewisse Rolle. Komplexe Leistungen oder Produkte wie der Unterricht eines Lehramtskandidaten, der Entwurf eines Zweckbaues durch einen Architekten oder der Entwurf eines Titelblattes durch einen Setzerlehrling werden nach anderen Kriterien beurteilt als Antworten auf Sachfragen oder Lösungen mathematischer Aufgaben. Die Verwendung von sogenannten *Schätzskalen* stellt gegenüber einer intuitiven Beurteilung bereits einen gewissen Fortschritt dar, zeitigt häufig jedoch noch recht unzuverlässige Resultate. Der Wert von Schätzskalen kann wesentlich verbessert werden, wenn wenigstens die extremen Ausprägungen einer Beurteilungsdimension durch Beispiele beschrieben werden.

Komplexe Leistungen können unter verschiedenen Gesichtspunkten beurteilt werden, die voneinander unabhängig sind. So kann ein Aufsatz bezüglich des Reichtums des Vokabulars, mit dessen Hilfe ein Schüler seine Gedanken ausdrückt, bezüglich

der Originalität und Folgerichtigkeit der Gedankenführung sowie bezüglich der grammatikalischen Richtigkeit des Satzbaus und der verwendeten Formen beurteilt werden. Wenn Prüfungsaufgaben komplexe Leistungen verlangen, sollen unabhängige Leistungsaspekte einzeln beurteilt werden. Dieses *Prinzip der differenzierten Beurteilung (Notengebung)* (nach AEBLI a. a. O., 182) führt zu einer Erhöhung des Informationsgehaltes der Rückmeldung an den Schüler und ist daher in didaktischer Hinsicht wichtig.

4.4. Definition eines Erfolgskriteriums bzw. eines Zensierungsmodells

Wir haben im vorherigen Abschnitt über Kriterien zur Beurteilung einzelner Leistungen gesprochen. Zur gesamthaften Beurteilung des Prüfungsergebnisses benötigen wir vielfach auch ein *Erfolgskriterium*, das die *untere Grenze* für das als erfolgreich geltende Prüfungsergebnis markiert. Wir haben damit ein Leistungsmass – in der Testtheorie als «cut-off-score» bekannt –, um zu beurteilen, ob ein Prüfling die Lernziele erreicht oder nicht erreicht bzw. ob er die Prüfung bestanden oder nicht bestanden hat.

Es gibt mehrere Möglichkeiten, um die untere Grenze des als ausreichend geltenden Prüfungsergebnisses zu beschreiben. In eher selteneren Fällen geben wir eine *zeitliche Begrenzung* an, wie beispielsweise, wenn 100 m in einer begrenzten Zeit gelaufen werden müssen, um in den Sportclub aufgenommen zu werden. Viel häufiger wird das Erfolgskriterium durch die Mindestzahl richtiger Antworten oder erzielter Punkte (manchmal auch durch die maximale Fehlerhäufigkeit) definiert, die für das Bestehen der Prüfung erforderlich ist. Wo man im Einzelfall die Grenze zieht, hängt weitgehend von den Lernzielen ab, deren Realisierung man überprüft. Die untere Grenze sollte auf jeden Fall das Erreichen der Lernziele zum Ausdruck bringen.

Für evaluative Zwecke genügt es, ein Erfolgskriterium zu definieren und die Prüfungsergebnisse daran zu messen. Um Noten vergeben zu können, reicht die Kenntnis der unteren noch zulässigen Leistungsgrenze jedoch nicht aus. Die Vergabe von Noten aufgrund vorliegender Rohwerte (Fehlerhäufigkeit, Anzahl richtiger Lösungen, Punkte)

setzt voraus, dass wir entsprechend der Zahl der positiven Noten die *Intervalle definieren*, die oberhalb und unterhalb des Erfolgskriteriums für die einzelnen Noten gelten. Es erleichtert die Auswertung von Prüfungsergebnissen, wenn wir die Zuordnung von Noten zu Rohwertintervallen in der Form eines Nomogramms festhalten. Es sei hier nochmals auf die verschiedenen Möglichkeiten von Zensierungsmodellen verwiesen, die unter Abschnitt 2.2.3. behandelt wurden.

4.5. Durchführen der Prüfung

Wer einen Endtest oder eine Prüfung durchführt, sollte *vorher die Bedingungen festlegen*, unter denen die Erfolgskontrolle erfolgt, und diese den Prüflingen rechtzeitig mitteilen. Was wird den Prüflingen an Hilfsmitteln (Rechenschieber, Tabellenbuch, Formelsammlung, Wörterbücher) erlaubt? Was wird ihnen verweigert? Wieviel *Zeit* steht zur Verfügung? Ob und welche *Hilfsmittel* bei Prüfungen benutzt werden dürfen oder nicht, stellt für manche Lehrer ein Problem dar. Es gibt hierfür keine allgemeine Verhaltensregel; die Entscheidung über die zu verwendenden Hilfsmittel ist von den Lernzielen abhängig, deren Realisierung man überprüfen möchte. Wenn man beispielsweise in einer Mathematikprüfung feststellen möchte, ob die Schüler eine bestimmte mathematische Struktur (z. B. Lehrsatz des Pythagoras) verstanden haben und diese anwenden können, so wird die Aussagekraft der Prüfung durch die Verwendung von Hilfsmitteln wie Rechenschieber, Tabellenbuch und Formelsammlung in keiner Weise beeinträchtigt. Wenn man dagegen in erster Linie die Rechenfertigkeit überprüfen möchte, so sind solche Hilfsmittel unzweckmässig. Es ist wichtig, dass für alle Teilnehmer die gleichen Bedingungen gelten. Deshalb sind nur solche Hilfsmittel zu erlauben, die für die ganze Klasse zur Verfügung stehen.

Noch ein Wort zum «Prüfungsklima», das die Prüfungsergebnisse je nach der Persönlichkeitsstruktur der Prüflinge beeinflussen kann. Eine aufgeregte, nervöse Stimmung – hervorgerufen durch Zeitdruck oder andere Faktoren – kann die Leistungsfähigkeit einzelner Schüler vermindern. Es ist deshalb wichtig, dass bei Prüfungen ein möglichst

entspanntes und vor allem angstfreies Klima vorherrscht.

4.6. Auswerten und Interpretieren der Prüfungsergebnisse

Sind die Masstäbe der Leistungsbeurteilung festgelegt, so ist eine relativ objektive Auswertung der Prüfungsergebnisse möglich. Allerdings sind damit noch lange nicht alle Fehlerquellen in der Beurteilung ausgeschaltet. Das Urteil des Lehrers kann trotz objektiver Leistungsmaßstäbe durch seine Erwartungen hinsichtlich der Leistungsfähigkeit der einzelnen Schüler verzerrt werden. Dies ist beispielsweise der Fall, wenn ein sogenannter «guter» Schüler für die gleiche Leistung besser bewertet wird als ein «schlechter» Schüler. Die Verzerrung der Beurteilung einzelner Leistungsaspekte durch andere Eindrucksmerkmale (z. B. Schriftbild, Gliederung) oder durch eine allgemeine Vorstellung von der zu beurteilenden Person bezeichnet man in der Psychologie als «Halo-Effekt». Die Wahrscheinlichkeit von Halo-Effekten wird vermindert, wenn eine Prüfung durch zwei oder mehrere Experten ausgewertet und beurteilt wird.

Bei der Interpretation eines Prüfungsergebnisses genügt es nicht, einfach die Punktwerte anzusehen und daraus seine Schlüsse zu ziehen. Schüler A hat 15 Punkte erreicht, Schüler B nur 10; also ist A besser als B. Dieser Schluss ist nur bedingt gültig. Diese Punktwerte sind bei einer ganz bestimmten Stichprobe von Aufgaben erreicht worden; bei einer gleich grossen Stichprobe ähnlicher Aufgaben hätten die beiden Schüler vermutlich nicht die gleichen Werte erzielt (Fassungsabhängigkeit einer Prüfung). Mit einer begrenzten Stichprobe von Aufgaben kann man nicht die «wahre» fachspezifische Leistungsfähigkeit eines Schülers feststellen. Man erhält stets ein Leistungsbild, das von Zufallsfaktoren abhängig und also ungenau ist (vgl. Abschnitt 3.3.). Wenn wir zu einem Stoffgebiet (z. B. Bruchrechnen) eine grosse Anzahl von Tests mit je 20 Aufgaben durchführten, so würde der Gesamtdurchschnitt die «wahre» Leistungsfähigkeit eines Prüflings in bezug auf das Bruchrechnen ausdrücken. Die Mittelwerte der Einzeltests würden annähernd normal um den Gesamtmittelwert streuen. In der Praxis ist es

jedoch nicht möglich und auch nicht sinnvoll, zum gleichen Stoffgebiet eine Reihe von Tests durchzuführen, um den «wahren» Wert zu ermitteln. Wir können der Ungenauigkeit einer Einzelmessung dadurch Rechnung tragen, dass wir bei der Interpretation eines Prüfungsergebnisses eher von einem Punktebereich (Vertrauensintervall des «wahren» Wertes) als von einem einzelnen Punktwert ausgehen. Der Vertrauensbereich des «wahren» Wertes lässt sich bei der Auswertung von Tests mit statistischen Mitteln ziemlich genau bestimmen. Damit haben wir bereits ein sehr spezifisches testtheoretisches Problem «angeschnitten», das eher für den Testpsychologen als für den praktisch tätigen Lehrer von Bedeutung ist. Wer sich damit weiter auseinandersetzen will, sei auf die Auswahlbibliographie verwiesen (insbesondere auf WENDELER 1972⁴, 52 ff.).

Schlussbemerkungen

Die in diesem Beitrag behandelten technischen und methodologischen Aspekte des Prüfens dürfen nicht isoliert gesehen werden. Die angestrebte Objektivierung der Erfolgskontrolle erfüllt keinen Selbstzweck. Sie steht immer im Dienste der Verbesserung des Unterrichts und der Förderung des einzelnen Schülers, indem sie erst gezielte Massnahmen zur Individualisierung des Unterrichts ermöglicht.

Bei aller Objektivierung der Leistungsmessung darf indessen auch der pädagogische Aspekt der Leistungsbeurteilung nicht aus dem Auge verloren werden. Die Beurteilungsmasstäbe beruhen zum Teil auf gesellschaftlichen Wertvorstellungen, die mit dem pädagogischen Anliegen der individuellen Förderung der einzelnen Schüler im Widerspruch stehen. Die unserem Beurteilungssystem eigene Tendenz, alle Schüler an den gleichen Masstäben zu messen, ist in pädagogischer Hinsicht fragwürdig. Dieses System trägt einerseits der individuellen Lebens- und Lernbiographie der Schüler zu wenig Rechnung und fördert andererseits eine Uniformität auf Kosten individueller Vielfalt. Der verantwortungsbewusste Lehrer wird bei der Leistungsbeurteilung auch diese Zusammenhänge zu sehen und zu berücksichtigen versuchen.

Literaturverzeichnis

- AEBLI, H. (1963). *Psychologische Didaktik*. Stuttgart: Klett.
- (1969). Über den Aufbau kognitiver Strukturen. In: *Bericht über den 26. Kongress der Deutschen Gesellschaft für Psychologie*. Göttingen: Hogrefe 1969, S. 72–78.
- (1971⁷). *Grundformen des Lehrens*. Stuttgart: Klett.
- BLOOM, B. S. (1972). *Taxonomie von Lernzielen im kognitiven Bereich*. Weinheim: Beltz.
- CORRELL, W. (1966). *Einführung in die Pädagogische Psychologie*. Donauwörth: Auer.
- GLASER, R. (1962). Psychology and Instructional Technology. In: R. GLASER (ed.), *Training Research and Education*. Pittsburg 1972, Seiten 1–30.
- (1973). Unterrichtstechnologie und die Messung von Lernergebnissen. In: HOFER/WEINERT (Hrsg.), *Pädagogische Psychologie, Grundlagentexte 2: Lernen und Instruktion*. Fischer Taschenbuch 1973, S. 313–319.
- HELLER, K. / ROSEMAN, B. (1974). *Planung und Auswertung empirischer Untersuchungen*. Stuttgart: Klett.
- INGENKAMP, K. (1971). *Die Fragwürdigkeit der Zensurengebung*. Weinheim: Beltz.
- LIENERT, G. A. (1967²). *Testaufbau und Testanalyse*. Weinheim: Beltz.
- MAGER, R. F. (1969³). *Lernziele und Programmierter Unterricht*. Weinheim: Beltz Bibliothek.
- POSTLETHWAITE, T. N. (1968). *IEA-Leistungsmessung in der Schule*. Frankfurt: Diesterweg.
- SARASON, S. B. et al. (1971). *Angst bei Schulkindern*. Stuttgart: Klett.
- WEISS, R. (1965). *Zensur und Zeugnis*. Linz.
- WENDELER, J. (1972⁴). *Standardarbeiten*. Weinheim: Beltz Bibliothek.

Annotierte Auswahlbibliographie zum Thema Erfolgskontrolle des Unterrichts für die Hand des Lehrers

A. Allgemeine Literatur

- AEBLI, H. (1971). Die Kontrolle des Unterrichtserfolgs. In: AEBLI, H. (1971⁷). *Grundformen des Lehrens*. Stuttgart: Klett, S. 179–195.
Dieser Beitrag enthält allgemeine Grundsätze zur Planung und Durchführung von Prüfungen mit einem wichtigen Abschnitt zur Prüfung des Verständnisses.
- GÖLLER, A. (1968²). *Zensuren und Zeugnisse*. Stuttgart: Klett.
Eine systematische Einführung in Probleme der Leistungsmessung und -beurteilung unter besonderer Berücksichtigung der Notengebung, wobei testtheoretische Aspekte eher vernachlässigt werden.

- LICHTENSTEIN-ROTHER, Ilse (Hrsg.) (1971). *Schulleistung und Leistungsschule*. Bad Heilbrunn/Obb.: Klinkhardt.
Ein Sammelband mit einer Reihe interessanter Beiträge von namhaften Autoren zu den Themen: Leistungsschule – Leistungsgesellschaft, Beurteilen und Messen von Schülerleistungen, Bedingungen von Schulerfolg und Leistungsniveau, Änderung der Lernbedingungen und neues Verständnis der Schulleistung.
- MESSNER, Rudolf (1971). Die Messung des schulischen Erfolgs. In: RUMPF / MESSNER (Hrsg.) 1971). *Didaktische Impulse*. Wien: ÖBV, S. 145–174 (Vorabdruck in «schweizer schule» 16/1970).
Eine Einführung in die Probleme der Kontrolle des Unterrichtserfolgs durch den Lehrer mit praktischen Anregungen zu einer partiellen Verbesserung gegenwärtiger Prüfungsformen.
- INGENKAMP, K. (Hrsg.) (1971). *Die Fragwürdigkeit der Zensurengebung*. Weinheim: Beltz Studienbuch.
Ein Sammelband mit einer Reihe von wissenschaftlichen Beiträgen über Fehlerquellen und Mängel des gegenwärtigen Systems der Notengebung.

B. Spezielle Literatur: Leistungsmessung durch Informelle Tests

- CHAUNCEY, H. / DOBBIN, J. E. (1968). *Der Test im modernen Bildungswesen*. Stuttgart: Klett.
Dieses Buch ist keine Anleitung zum Testen, sondern vermittelt einen Überblick über die Verwendung von Tests im Bildungswesen zum Zwecke der Erfolgskontrolle und der Bildungsberatung. Im Anhang enthält es mehrere Beispiele von Mehrfachwahlfragen zu verschiedenen Fächern und ein Kapitel über «Statistische Kurzverfahren zur Analyse Informeller Tests».
- GAUDE, P. / TESCHNER, W.-P. (1971²). *Objektivisierte Leistungsmessung in der Schule*. Frankfurt: Diesterweg.
Bericht über den Einsatz Informeller Tests im leistungsdifferenzierten Unterricht der Gesamtschule Berlin mit einem Kapitel zur Planung, Entwicklung und Auswertung Informeller Tests sowie mit zahlreichen Beispielen solcher Tests für verschiedene Schulfächer.
- HOPF, Dieter (1970). Möglichkeiten und Grenzen der Anwendung von Tests. In: HOFER/WEINERT (Hrsg.) (1973). *Funk-Kolleg Grundlagentexte. Pädagogische Psychologie, Bd. 2: Lernen und Instruktion*. Fischer Taschenbuch, Seiten 302–313.
Eine kurze Darstellung der Eigenschaften psychologischer Tests, ihrer Anwendungsmöglichkeiten und ihrer Gefahren beim Einsatz zu Selektionszwecken.

- INGENKAMP, K. (1971). *Tests in der Schulpraxis*. Weinheim: Beltz Bibliothek.
Ein Sammelband mit einer Reihe praxisrelevanter Beiträge über die Entwicklung und Beurteilung von Leistungstests für die Schule und einem kurzen Überblick über wichtige standardisierte Schultests (Schulreifetests, Schulleistungstests, Eignungstests ...).
- WENDELER, J. (1972⁴). *Standardarbeiten*. Weinheim: Beltz Bibliothek.
Eine kurze Anleitung zur Entwicklung und Auswertung Informeller Leistungstests (Standardarbeiten).

Hinweis

Die vorliegende Nummer der «schweizer schule» dürfte auch insbesondere für angehende Lehrer von Interesse sein. Seminarlehrern machen wir daher das Angebot, für ihre Klasse eine entsprechende Anzahl dieses Heftes **gratis** bei der Administration, Gotthardstrasse 27, 6300 Zug, zu beziehen (solange Vorrat!). CH

Aktuelle Kurzmeldungen der «schweizer schule»

CH: 23 Berufsmittelschulen

Ende 1973 bestanden 23 Berufsmittelschulen mit insgesamt 3307 Schülern in der Schweiz. Im Sinne eines Versuches wurde den kaufmännischen Berufsschulen Bern und Zürich die Führung einer kaufmännischen Berufsmittelschule gestattet. Eine Arbeitsgruppe ist vom Eidgenössischen Volkswirtschaftsdepartement beauftragt, die provisorische Wegleitung über die Errichtung und Organisation von Berufsmittelschulen auf Grund der bisherigen Erfahrungen zu überprüfen und Vorschläge für eine definitive Gestaltung dieser Institution auszuarbeiten.

CH: Die EDK ein zweites Mal vor der Bewährung

Ein Minimum von drei wöchentlichen Stunden in der zweiten Landessprache fordert die Expertenkommission für Fremdsprachenunterricht, die kürzlich in Zürich der Konferenz der kantonalen Erziehungsdirektoren unter dem Vorsitz von Regierungsrat Dr. Alfred Gilgen, Zürich, ihren Schlussbericht unterbreitet hat. Ihr wichtigster Antrag sieht vor, den Beginn des Unterrichts vor die Pubertät zu legen. Dabei empfehlen die Bildungsexperten für die Westschweiz, Italienisch- und Romanischbünden Deutsch als erste Fremdsprache, deren Stelle in der Deutschschweiz und dem Tessin durch den Französischunterricht eingenommen werden soll.

Hinsichtlich des Beginns des Unterrichts scheint die Kommission den Westschweizer Schülern mehr zuzumuten. In diesem Landesteil soll der Deutschunterricht bereits ab dem dritten Schuljahr eingeführt werden, während in den Deutsch-

schweizer Schulen Französisch erst ein Jahr später im Lehrplan figurieren soll. Für die Kantone Graubünden und Tessin sind spezielle Regelungen vorgesehen. Der Unterricht in der ersten Fremdsprache soll allen Schülern – mit Ausnahme derjenigen an Sonderschulen – offenstehen.

CH: Volksschul-Oberstufe als Orientierungsstufe?

Die interkantonale Studiengruppe «Ausbau der Oberstufe in der Volksschule» hat vor drei Wochen im Gottlieb-Duttweiler-Institut in Rüslikon eine Tagung über Probleme der Schulreform durchgeführt. In vier Referaten wurde die Frage behandelt, inwiefern die heutige Oberstufe der Volksschule zu einer Orientierungsstufe für den ganzen Schülerjahrgang ausgebaut werden könnte, wobei pädagogische, psychologische und regionale Gesichtspunkte zu berücksichtigen wären. Im Rahmen der Bildungsreform sei eine moderne Schule zu schaffen. Nur durch gründliche und langfristige Orientierung werde es möglich sein, den Jugendlichen den Einstieg in jene Bildungswege zu ebnet, die ihren Begabungen und Neigungen entsprechen.

BE: Bieler befürworten den Bau eines neuen Gymnasiums

Mit 7596 gegen 7157 Stimmen hat der Bieler Souverän einen Kredit in der Höhe von 39,5 Millionen Franken zum Bau eines neuen Gymnasiums knapp gutgeheissen.

LU: Spende für Luzerner Schüler

Die Stadt Luzern hat eine Vergabung von 100 000