

Der Bundesausschuß für Berufsbildung, der gemäß § 51 BBiG die Bundesregierung in grundsätzlichen Fragen der Berufsbildung berät, hat in der 13. Sitzung seiner zweiten Amtsperiode am 3. Mai 1974 die nachstehend abgedruckten Stellungnahmen beschlossen:

#### Stellungnahme zum „Gutachten für die Entwicklung eines Curriculums Berufswahlunterricht“

Der Bundesausschuß für Berufsbildung hat in seiner Empfehlung vom 26. Januar 1972 grundsätzliche Aussagen zu den Lernzielen und Inhalten der vorberuflichen Bildung gemacht und dabei u. a. gefordert, daß „die Arbeitsverwaltung sowohl personell als auch durch entsprechende Curricula in den Stand versetzt wird, . . . . Unterricht in Berufswahlkunde zu erteilen“<sup>1)</sup>.

In Übereinstimmung mit dieser Empfehlung hat die Bundesanstalt für Arbeit, die nach geltendem Recht für die Berufsberatung zuständig ist, ein wissenschaftliches Gutachten an die Professoren Dr. Harald Dibbern, Dr. Franz-Josef Kaiser und Dr. Adolf Kell in Auftrag gegeben.

Nach Auffassung des Bundesausschusses für Berufsbildung vermag das von diesen Wissenschaftlern am 30. April 1973 fertiggestellte Gutachten „Berufswahlunterricht in der vorberuflichen Bildung – Der didaktische Zusammenhang von Berufsberatung und Arbeitslehre“<sup>2)</sup> weiterführende Anregungen für die Realisierung der Empfehlung über die vorberufliche Bildung zu geben. Es wird auch begrüßt, daß durch die Veröffentlichung dieses Gutachtens eine breite Diskussion und Meinungsbildung möglich wird.

Der Bundesausschuß für Berufsbildung unterstützt – ohne sich mit Einzelaussagen des Gutachtens über inhaltliche, methodische, organisatorische und rechtliche Fragen zu identifizieren – grundsätzlich die Forderung nach einer von Schule und Berufsberatung gemeinsam zu leistenden kontinuierlichen Berufswahlvorbereitung.

Die Kultusminister/-senatoren der Länder und die Bundesanstalt für Arbeit werden gebeten, die curricularen Grundlagen für den Berufswahlunterricht in enger Zusammenarbeit von Schule und Berufsberatung baldmöglichst zu entwickeln und zu erproben.

#### Stellungnahme zur Anwendung des Muster-Berufsausbildungsvertrages im öffentlichen Dienst

Der Bundesausschuß für Berufsbildung hat am 9. Juni 1971 ein Muster eines Berufsausbildungsvertrages verabschiedet und allen zuständigen Stellen zur bundeseinheitlichen Beschlußfassung und Anwendung empfohlen. Trotz wiederholter Empfehlungen des Bundesausschusses haben der Bund, die Tarifgemeinschaft deutscher Länder und die Vereinigung der kommunalen Arbeitgeberverbände unter Ablehnung des vom Bundesausschuß für Berufsbildung beschlossenen Vertragsmusters ein eigenes Vertragsmuster für den öffentlichen Dienst empfohlen.

Dem öffentlichen Dienst kann im Bereich der Berufsbildung aus sachlichen und politischen Gründen keine Sonderstellung eingeräumt werden. Der Bundesausschuß für Berufsbildung fordert deshalb die Bundesregierung auf, dafür zu sorgen, daß auch im öffentlichen Dienst die Beschlüsse des Bundesausschusses beachtet werden.

<sup>1)</sup> „Empfehlung über vorberufliche Bildung und Beratungsdienste“ vom 26. Januar 1972, abgedruckt im Bundesarbeitsblatt Nr. 3/72.

<sup>2)</sup> Das Gutachten kann durch den Buchhandel oder unmittelbar beim Verlag Julius Klinkhardt, Bad Heilbrunn/Obb., zum Preis von 14,60 DM bezogen werden.

Wilfried Reisse

## Erläuterungen zu der Empfehlung für programmierte Prüfungen des Bundesausschusses für Berufsausbildung<sup>1)</sup>

In der „Empfehlung für programmierte Prüfungen“ werden zum Teil recht spezielle Vorschläge und Forderungen formuliert, deren Erläuterung den Rahmen einer solchen Empfehlung sprengen würde. Wegen der Schwierigkeit der hier behandelten Probleme erscheint jedoch eine Erläuterung und Interpretation dringend erforderlich, damit die Empfehlung recht bald in der Prüfungspraxis Berücksichtigung findet. Diese Lücke soll durch den vorliegenden Beitrag geschlossen werden, der sich daher im Aufbau und Inhalt eng an die Empfehlung anlehnt.

Fragen, die im Zusammenhang mit Prüfungen und Lernkontrollen und deren Funktion in der Berufsbildung allgemein zu klären sind und Überlegungen, die zu der vorliegenden Empfehlung führten, treten dagegen weitgehend in den Hintergrund.

**Die vorliegende Empfehlung soll dazu beitragen, daß Materialien und Verfahren für Prüfungen in der Berufsausbildung sachgemäß entwickelt und verwendet werden.**

„Prüfungen in der Berufsausbildung“ (als Anwendungsbereich der Empfehlung) umfassen sicher nicht nur die Zwischen- und Abschluß- bzw. Facharbeiter- und Gesellenprüfungen in der beruflichen Erstausbildung, sondern alle Varianten von Lernerfolgskontrollen, die mit unterschiedlichen Zielsetzungen bei beruflichen Bildungsgängen realisiert werden, beispielsweise

- Prüfungen bzw. Lernerfolgskontrollen im Verlauf oder als Abschluß von Maßnahmen der beruflichen Fortbildung und Umschulung;
- zusätzliche innerbetriebliche und evtl. auch überbetriebliche Prüfungen bzw. Lernerfolgskontrollen, wenn entsprechend der in der jeweiligen Ausbildungsordnung festgelegten sachlichen und zeitlichen Gliederung Ausbildungsabschnitte abgeschlossen worden sind;
- Lernerfolgskontrollen als Teil von Lehrmaterialien und Mehrmediensystemen;

<sup>1)</sup> „Empfehlung für programmierte Prüfungen“, Beschluß des Bundesausschusses für Berufsbildung vom 1. März 1973. Der Verfasser hat als Sachverständiger des BBF bei der Erarbeitung dieser Empfehlung mitgewirkt.

- Untersuchung des Lernerfolgs bei der Auswertung von Modellversuchen zur beruflichen Bildung.

Die Empfehlung kann darüber hinausgehend auch die Diskussion über Prüfungen in beruflichen Schulen anregen, zum Beispiel über

- die „Klassenarbeiten“ als mehrfache Kurzprüfungen bzw. Kurzkontrollen in beruflichen Schulen (Teilzeit- und Vollzeitform) und die
- schriftlichen Abschlußprüfungen bei vollschulischen beruflichen Ausbildungsgängen (zum Beispiel Fachschulen).

Eine „sachgemäße“ Entwicklung und Anwendung von Prüfungsmaterialien und Verfahren muß sich am Kenntnisstand in diesem Bereich orientieren. Da Probleme und Verfahren von Leistungsmessungen – wie sie bei Prüfungen erfolgen müssen – seit Jahrzehnten bevorzugter Gegenstand der psychologischen Leistungsdiagnostik und deren Anwendungen im Bildungsbereich darstellen, muß der umfangreiche Bestand an gesicherten technologischen Kenntnissen aus diesem Problemkreis einbezogen werden. Es besteht sonst die Gefahr, daß mit erheblichen Kosten und – was wesentlich schwerer wiegt – mit beträchtlichen Nachteilen für die Auszubildenden längst bekannte Ergebnisse ein zweites Mal entdeckt und damit Fehlentwicklungen der bisherigen Testpsychologie wiederholt werden.

„Programmierte Prüfungen“ sind zur Zeit noch Prüfungen mit informellen (nichtstandardisierten) objektiven Tests, um hier die Bezeichnungen einzuführen, die in der testpsychologischen Literatur für solche Verfahren verwendet werden. „Informelle“ Tests sind Instrumente, die nicht den Güte Merkmalen – den „Standards“ – entsprechen, die von den „standardisierten“ Tests als wissenschaftlich begründete Routineverfahren gefordert werden. „Objektiv“ heißt in diesem Zusammenhang, daß die Form der Aufgabenbeantwortung bei programmierten Prüfungsaufgaben (die „gebundene“ Form, bei der die Beantwortung nicht frei erfolgt, sondern an vorgegebene Antwortmöglichkeiten „gebunden“ ist) eine objektive Auswertung ermöglicht.

Eine für die Weiterentwicklung der programmierten Prüfungen sehr wichtige Unterscheidung ist die zwischen (mehr) „lernzielorientierten“ und (mehr) „normorientierten“ Leistungstests. Lernzielorientierte (lehrzielorientierte, kriteriumsbezogene) Tests zielen ab auf einen Vergleich der individuellen Leistung mit vorgegebenen Ausbildungszielen, bei normorientierten Messungen liegt der Akzent mehr auf einem Vergleich des individuellen Ergebnisses mit einem Gruppendurchschnitt. Obwohl Zwischen- und Abschlußprüfungen nach ihrer Intention als auch nach der teilweise geübten Praxis eher den mehr lernzielorientierten Kontrollen zuzuordnen sind, besteht gegenwärtig die Gefahr, die für normorientierte Tests gedachten Konstruktions- und Analyseverfahren auf programmierte Prüfungen unkritisch zu übertragen.

Zur Leistungsmessung und Lernerfolgskontrolle mit Testverfahren gibt es eine Vielzahl von Veröffentlichungen. Einige deutschsprachige Arbeiten mit Einführungs- und Übersichtscharakter sollen hier herausgegriffen werden: SÜLLWOLD 1964 beschreibt u. a. allgemeine Merkmale und Grundtypen von Schultests oder pädagogischen Tests sowie Güte Merkmale dieser Verfahren, wobei besonders beeindruckt, welche umfangreichen Erfahrungen auf diesem Gebiet bereits in den USA vorliegen. Wichtige Hinweise für die Praxis der Formulierung, Zusammenstellung, Erprobung und Revision von Aufgaben bei informellen Tests wie programmierten Prüfungen werden u. a. in Veröffentlichungen von BIGLMAIER 1969, SEELIG 1970, WENDELER 1970 und GRONLUND 1974 behandelt. Als anspruchsvolles Standardwerk der „klassischen“ Testkonstruktion gilt immer noch das Buch von LIENERT 1969, das jedoch im Hinblick auf die darin behandelten teststatistischen Verfahren eher für Spezialisten geeignet sein dürfte. Von den zahlreichen Veröffentlichungen INGENKAMPs zu den hier behandelten Fragen erscheint als eine praxisorientierte Einführung besonders der Sammelband „Tests in der Schulpraxis“ (INGENKAMP 1971)

empfehlenswert, in dem eine Reihe wichtiger Themen behandelt werden (Aufgaben der Testanwendung, Testentwicklung, Vergleich zwischen Test und traditioneller Prüfung, Anforderungen an Tests usw.) Die genannten Veröffentlichungen beziehen sich jedoch eher auf die seit mehr als einem halben Jahrhundert gebräuchlichen normorientierten Testverfahren. Der Aspekt der Lernzielorientierung als ein relativ neuer Ansatz wird bei INGENKAMP 1973, BÜSCHER 1974 und – allerdings mit eher theoretischer Orientierung – bei KLAUER 1972 behandelt. Schließlich ist noch auf die Veröffentlichung von HUBBARD 1974 über Prüfungen in der medizinischen Ausbildung in den USA aufmerksam zu machen. In dieser Arbeit wird beispielhaft gezeigt, wie vorbildliche Testprogramme für Prüfungen in der Berufsausbildung entwickelt und eingesetzt werden.

## 1. Anforderungen an die programmierten Prüfungsaufgaben

### 1.1 Die Aufgaben für programmierte Prüfungen müssen inhaltlich gültig sein (Validität), d. h. sich auf die Lernziele bzw. Inhalte beziehen, die z. B. in Ausbildungsordnungen und Rahmenlehrplänen festgelegt sind.

Bei jeder Aufgabe sollten sich die Aufgabenautoren, die pädagogischen Mitarbeiter eines Verlags, der eine Aufgabensammlung anbietet, die Mitglieder einer Aufgabenauswahlkommission bzw. eines Prüfungsausschusses die Frage stellen, ob mit der Aufgabe tatsächlich ein wichtiges Ausbildungs- bzw. Unterrichtsziel repräsentiert wird und wie sich ihre Auswahl begründen läßt. Oder sie sollten umgekehrt prüfen, ob die jeweilige Aufgabe

- sich auf Rand- und Detailwissen bezieht, das für die Berufspraxis relevant ist,
- ein „Steckenpferd“ des Aufgabenautors darstellt,
- nur formuliert wurde, weil der entsprechende Inhalt leicht in eine programmierte Aufgabe umzuwandeln war! (Lernstoff mit „Aufzählcharakter“: Länder der Bundesrepublik, Bestandteile des Pinsels, Rechte des Käufers usw.).

Der Kritiker fragt also, was der Auszubildende zu leisten vermag, wenn er die in Frage stehende Aufgabe richtig beantwortet und ob diese Leistung überhaupt wichtig ist.

### 1.2 Anerkannte Formulierungsregeln sind zu beachten.

Das Schreiben guter Testaufgaben gilt mit Recht als hochqualifizierte Tätigkeit (EBEL 1951). Eine Voraussetzung dafür ist, daß man die einzelnen Aufgabentypen und deren Vor- und Nachteile nennt. Man sollte dabei von der allgemein bei der Testkonstruktion üblichen Einteilung ausgehen und die folgenden Aufgabentypen unterscheiden:

- Zweifachwahlaufgaben (Richtig-Falsch-Aufgaben): Es werden eine Reihe von Aussagen oder Antworten angeboten, bei jeder einzelnen muß der Prüfungsteilnehmer angeben, ob sie – meist unter einer vorgegebenen Fragestellung – zutrifft oder nicht.
- Mehrfachwahlaufgaben: Von mehreren vorgegebenen Antwortmöglichkeiten ist die richtige oder beste zu kennzeichnen.
- Zuordnungsaufgaben: Jede Aufgabengruppe besteht aus einer Liste von Fragen, Problemen oder allgemein „Prämissen“ und einer Liste von Antworten oder Lösungen. Die einzelnen Elemente dieser Listen sind nach einer Anweisung einander zuzuordnen.

Innerhalb dieser für die Konstruktionspraxis wichtigsten Aufgabentypen gibt es – besonders bei den Mehrfachwahlaufgaben – zahlreiche Varianten. Bei der Wahl eines Aufgabentypus sind dessen Vor- und Nachteile zu beachten. Unterschiede bestehen u. a. hinsichtlich der Verständlichkeit, der Auswertbarkeit, des Komplexitätsgrades, der Häufigkeit von Zufallslösungen und der Abhängigkeit von Gedächtnisleistun-

gen. Beispielsweise sind Richtig-Falsch-Aufgaben relativ leicht zu formulieren, aber durch viele Zufallslösungen und ihre Abhängigkeit von Gedächtnisleistungen gekennzeichnet. Mit Zuordnungsaufgaben kann man mit relativ geringem Aufwand einen Lernbereich „abtasten“, muß aber einige Mühe verwenden, um bei jeder Aufgabe zu verdeutlichen, wie die Zuordnung erfolgen soll. Gute Mehrfachwahlaufgaben sind wohl am schwierigsten zu formulieren, sie gelten aber wegen ihrer umfassenden Anwendbarkeit und ihrer günstigen teststatistischen Eigenschaften als bevorzugter Aufgabentyp.

Für das Aufgabenschreiben allgemein werden in der Literatur zahlreiche Regeln vorgeschlagen. Darüber hinaus gibt es noch Regeln, die für spezielle Aufgabentypen wichtig sind. Im Rahmen dieses Beitrags ist es jedoch nicht möglich, alle Hinweise mit Hilfe von Beispielen zu erläutern<sup>1)</sup>. Statt dessen werden die hier in die Empfehlung aufgenommenen generellen Formulierungsregeln kurz kommentiert wiedergegeben:

#### Insbesondere müssen

- Aufgaben so einfach und verständlich wie möglich formuliert werden.**

Beispielsweise sind Doppelfragen und Negationen ungeeignet, ebenso unbekannte Fachausdrücke, soweit sie nicht selbst Gegenstand der Aufgabe sind.

- Aufgaben in der Regel so formuliert und gestaltet sein, daß die Aufgabe der Anzahl zutreffender Antwortmöglichkeiten bei den einzelnen Aufgaben nicht notwendig ist.**

Diese Empfehlung zielt darauf ab, die Aufgaben so zu formulieren, anzuordnen und einzuleiten, daß weder Unklarheiten über den Beantwortungsmodus (die „Lösungstechnik“) auftauchen, noch ständige Hinweise dazu die Bearbeitung stören.

- Aufgaben alle Informationen enthalten, die für die richtige Lösung notwendig sind.**

Bei einer Verwendung nicht erprobter Aufgaben zeigt sich häufig, daß die Prüfungsteilnehmer zu manchen Aufgaben Fragen haben. Es stellt sich dann heraus, daß wichtige Angaben im Aufgabentext fehlen. Andererseits sind aber auch zusätzliche Informationen zu vermeiden, die sich als unbeabsichtigte Lösungshinweise auswirken (beispielsweise fällt die richtige Alternative heraus, wenn sie hinsichtlich der Länge oder des Präzisionsgrades von den anderen Antwortmöglichkeiten deutlich abweicht).

- Antwortmöglichkeiten eindeutig als falsch oder richtig einzuordnen sein.**

Das heißt, Experten für den jeweiligen Fachbereich sollten darin übereinstimmen, welche Antwortmöglichkeiten als richtig oder falsch gelten. Diese Fachleute müssen jedoch berücksichtigen, daß die für den Auszubildenden „richtigen“ Lösungen zwar die beste und am ehesten zutreffende Möglichkeit darstellt, aber nicht unbedingt der Antwort entsprechen müssen, die sie als Experten selbst auf die gestellte Frage geben würden.

Diese Regeln könnte man noch ergänzen. So wird in Anweisungen zum Aufgabenschreiben etwa gefordert, daß die Aufgaben unabhängig voneinander sein sollen, d.h. die Fragestellung oder Lösung einer Aufgabe darf keine Hinweise für die Lösung anderer Aufgaben enthalten.

Neben diesen allgemeinen Prinzipien gibt es noch spezielle Regeln für einzelne Aufgabentypen: Beispielsweise sollen bei Mehrfachwahlaufgaben die Antwortmöglichkeiten grammatikalisch auf den „Aufgabenstamm“ abgestimmt sein und

sich nicht gegenseitig voraussetzen oder einschließen. Für Zuordnungsaufgaben wird u. a. empfohlen, vollständige (Ein- zu eins-)Zuordnungen (gleichviel Fragen und Antworten) zu vermeiden.

In die Empfehlung ist eine Regelung aufgenommen worden, die sich auf einen Fragenkomplex bezieht, der für die derzeitige Praxis von erheblicher Bedeutung sein dürfte. Danach sollen

- Mehrfachwahlaufgaben in der Regel vier bis sechs Antwortmöglichkeiten vorgeben, wobei die Anzahl der richtigen Lösungen weniger als 50% der Zahl der Antwortmöglichkeiten betragen soll.**

Zunächst werden damit Unter- und Obergrenzen für die Zahl der Antwortmöglichkeiten festgelegt. Mehrfachwahlaufgaben mit drei vorgegebenen Antworten sind wegen der hohen Ratewahrscheinlichkeit nicht sinnvoll. Es ist üblich, bei weniger anspruchsvollen Verfahren mit vier Antwortmöglichkeiten auszukommen, bei besser konstruierten Materialien (standardisierten Tests) werden meistens fünf Antwortmöglichkeiten angeboten. Im übrigen sollte die Zahl der Antwortmöglichkeiten in einem Aufgabensatz konstant gehalten werden. Sechs Antwortmöglichkeiten könnten dann sinnvoll sein, wenn mehrere Antwortmöglichkeiten als richtig gelten sollen. Weiter wird in diesem Abschnitt eine Regelung für die Zahl der richtigen Lösungen empfohlen: Bei vier Antwortmöglichkeiten sollte nur eine richtig sein, von fünf und sechs Antwortmöglichkeiten können zwei zutreffen.

Damit wird eine Frage angeschnitten, die bei programmierten Prüfungen offensichtlich umstritten ist, bei standardisierten Schulleistungstests dagegen praktisch kaum noch eine Rolle spielt: Wie Aufgaben mit mehreren richtigen Antworten konstruiert und ausgewertet werden sollen. Die Einordnung dieser bei programmierten Prüfungen noch häufig verwendeten Aufgaben in die oben skizzierten Aufgabentypen ist schwierig: Man kann sie als Gruppe von Zweifachwahlaufgaben (Richtig-Falsch-Aufgaben) ansehen, da bei jeder „Antwortmöglichkeit“ entschieden werden muß, ob sie hinsichtlich der Fragestellung richtig oder falsch ist. Manchmal werden Aufgaben mit mehreren richtigen Antworten („Mehrfachantwortaufgaben“) als Variante der Typs „Mehrfachwahlaufgabe“ angesehen. Diese Zuordnung ist wichtig, da die mit der Aufgabe zu erfassenden Lernbereiche und -inhalte, die Formulierungsregeln und ganz besonders der Bewertungsmodus in beiden Fällen verschieden sind. Dazu kommt noch, daß Aufgaben dieser Art bei der teststatistischen Verarbeitung (z. B. bei der Ermittlung von Kennwerten) beträchtliche Schwierigkeiten bereiten.

Der Verfasser schlägt folgende Wege vor, um allmählich zu einer Lösung dieser Fragen zu kommen:

- Zunächst ist zu prüfen, ob der Inhalt bzw. das Lernziel es zwingend notwendig machen, mehrere richtige Lösungen anzubieten oder ob nicht deswegen dieser Aufgabentyp gewählt wurde, weil sich Aufgaben mit mehreren Lösungen sehr schnell formulieren lassen.
- Aufgaben mit mehreren Lösungen können in Mehrfachwahlaufgaben mit einer Richtig-oder Best-Antwort umgewandelt werden. Die einzelnen Antwortmöglichkeiten enthalten Kombinationen von Teilantworten, die richtige Antwortmöglichkeit enthält nur die zutreffenden Teilantworten, die anderen Antwortmöglichkeiten sind Zusammenstellungen von richtigen und falschen Teilantworten.
- Wenn diese Umwandlung zu sehr umständlichen Aufgaben führt (etwa wenn die Teillösungen zu umfangreich sind) ist eine Serie von Mehrfachantwortaufgaben mit gleichem Modus (zum Beispiel von fünf Antwortmöglichkeiten sind immer zwei auszuwählen) anzustreben, wobei eine Bewertung nur erfolgt, wenn alle richtigen Lösungen (und nur diese!) angekreuzt wurden. Schließlich kann

<sup>1)</sup> Eine weitere Veröffentlichung mit ergänzenden Beispielen ist in Vorbereitung.

man – soweit genügend Aufgaben dieser Art vorliegen – eine Serie von Richtig-Falsch-Aufgaben zusammenstellen, wobei jede richtig eingeordnete Antwortmöglichkeit bewertet wird. Wegen der sehr hohen Ratawahrscheinlichkeit müssen Serien mit Richtig-Falsch-Aufgaben recht umfangreich sein.

### 1.3 Die Aufgaben müssen so einfach und übersichtlich gestaltet werden, daß sich bei der Kennzeichnung und Eintragung der Lösungen keine Schwierigkeiten ergeben.

Fehler entstehen etwa dann, wenn die Anordnung der Aufgaben und der Antwortmöglichkeiten auf dem Aufgabenblatt anders ist als auf dem (gesonderten) Antwortbogen. Besonders bei Zuordnungs- und noch mehr bei Umordnungsaufgaben (als einem weiteren Aufgabentyp) muß deutlich sein, welche Buchstaben oder Ziffern oder deren Kombination das Ergebnis darstellen sollen und wie diese Lösung einzutragen ist.

Solche mehr formalen Konstruktionsprinzipien können nur dann vernachlässigt werden, wenn mehr die Aspekte der lernzielbezogenen Übung, der Aktivierung durch interessante Aufgaben und der Selbstkontrolle und nicht der Prüfungs- und Entscheidungsaspekt im Vordergrund steht. Daher sind beispielsweise Aufgaben, bei denen die Formulieringsregeln nicht genügend beachtet wurden, nach Ansicht des Verf. noch tolerierbar bei Lernkontrollen als Teil von Lehrmaterialien, nicht aber in Prüfungssätzen für Abschlußprüfungen.

Zu den einzelnen Aufgabentypen, ihren Eigenschaften und Verwendungen und zur Formulierung und äußeren Gestaltung von „Aufgaben in gebundener Form“ (programmierte Aufgaben) gibt es – meist auf der Grundlage des klassischen Aufsatzes von EBEL 1951 – auch in deutscher Sprache mehrere brauchbare Anleitungen. Wer mit der Konstruktion von programmierten Aufgaben zu tun hat, wird nicht umhin kommen, die eine oder andere der folgenden Veröffentlichungen durchzuarbeiten: Hinweise finden sich in den schon genannten Arbeiten von WENDELER 1970, BIGL-MAIER 1969, LIENERT 1969 (bes. S. 25-41 und 63 f) und GRONLUND 1974. Von den Tonbildreihen des FWU zu objektivierten Leistungsprüfungen soll nur auf die Tonbildreihe „Aufgabenkonstruktion“ (Reihe B) hingewiesen werden, die über die Konstruktion von Mehrfachwahlaufgaben informiert. Es ist daran zu erinnern, daß die in diesen Unterlagen enthaltenen teststatistischen Empfehlungen sich eher auf normorientierte Tests beziehen. Eine umfassende Systematik und Darstellung der unterschiedlichsten Formen von Testaufgaben enthält das Buch von RÜTTER 1973. Da aber nach Meinung des Verfassers bei programmierten Prüfungen eher zu viel als zu wenig unterschiedliche Aufgabentypen verwendet werden, sei es hier nur mit dieser Einschränkung genannt. Eine gute Übersicht gibt der Aufsatz von HERBIG 1973.

### 1.4 Die Aufgaben sind in der Regel vor der Prüfung zu erproben.

Praktische Erfahrungen bei der Testkonstruktion und wissenschaftliche Untersuchungen haben gezeigt, daß sich bei programmierten Aufgaben (Testaufgaben) auch bei sehr sorgfältiger Formulierungsarbeit Fehlerquellen (beispielsweise durch mißverständliche Aussagen) nicht vermeiden lassen und daß man manche Mängel solcher Aufgaben nur durch eine Erprobung herausfinden kann. LIENERT betont daher zu Recht, daß die Verwendung nicht-analyzierter Mehrfachwahlaufgaben das Testverfahren disqualifiziert (1969, S. 34). Außerdem müssen auf diese Weise statistische Kennwerte für die einzelnen Aufgaben berechnet werden.

**Für jede Aufgabe müssen folgende statistische Kennwerte vorliegen:**

- Umfang und Zusammensetzung der Erprobungsgruppe
- Schwierigkeitsgrad (Prozentanteil der richtigen Lösungen)
- Häufigkeitsverteilung der Antwortmöglichkeiten
- Trennschärfe (die Aufgabe muß geeignet sein, Prüfungsteilnehmer mit guter Leistung von Prüfungsteilnehmern mit schlechter Leistung zu unterscheiden).

Die Qualität einer Aufgabe (als Voraussetzung für ihre Aufnahme in einen Prüfungssatz) ergibt sich auch aus ihren

statistischen Kennwerten. Zunächst sollte (unter Umständen generell für einen Aufgabenkatalog) gesagt werden, wie groß die Erprobungsgruppe war und inwieweit sie in ihrer Zusammensetzung der Gruppe entsprach, für die der Test (der Aufgabensatz) gedacht ist. Eine Erprobung mit ein oder zwei Berufsschulklassen wird zwar dazu beitragen, mißverständliche Formulierungen aufzudecken, für die Ermittlung von Kennwerten sollte man aber möglichst größere Gruppen (mindestens 200 Auszubildende) heranziehen. Bei den einzelnen Antwortmöglichkeiten einer Aufgabe ist dann prozentual anzugeben, wieviel Antworten auf sie entfallen (die „falschen“ Antwortmöglichkeiten sollten in etwa mit gleicher Häufigkeit gewählt worden sein). Damit liegt auch der „Schwierigkeitsgrad“ (Schwierigkeitsindex) fest (als Prozentanteil der richtigen Lösungen) wobei noch zusätzlich anzugeben ist, wieviel Auszubildende die Aufgabe nicht bearbeitet haben. Hier muß man beachten, daß nach dieser Festlegung – im Gegensatz zu dem üblichen Wortgebrauch – ein hoher Schwierigkeitsgrad (-index) (z. B. 85 %) eine Aufgabe als leicht kennzeichnet (85 % der Prüfungsteilnehmer haben die Aufgabe richtig gelöst). Eine wichtige Information ist auch mit der „Trennschärfe“ der Aufgaben gegeben. Eine trennscharfe Aufgabe wird von „guten“ Prüfungsteilnehmern (Teilnehmer mit insgesamt mehr richtigen Lösungen) sehr viel häufiger richtig gelöst als von weniger qualifizierten. Der umgekehrte Fall deutet an, daß die Aufgabe für die „guten“ Prüflinge unbekannte Schwierigkeiten enthält. Es läßt sich leicht sehen, daß ein Aufgabensatz mit zahlreichen nicht-trennscharfen Aufgaben das Leistungsbild einer Prüfungsgruppe verschwimmen läßt.

Es gibt verschiedene Verfahren, um Kennwerte für die Trennschärfe zu bestimmen. Da aber zur Zeit diskutiert wird, ob bei lernzielorientierten Tests spezielle Verfahren angebracht sind, erscheint es riskant, hier das eine oder andere Verfahren besonders zu empfehlen. Soweit programmierte Aufgaben für Prüfungen mit relativ großen Teilnehmerzahlen entwickelt und angeboten werden, sollte man die Berechnung dieser und anderer statistischer Kennwerte durch Teststatistiker vornehmen lassen und das jeweils gewählte Verfahren im Aufgabenkatalog erläutern und begründen.

Vereinfachte Verfahren für Trennschärferechnungen und Aufgabenanalysen allgemein finden sich u. a. bei WENDELER 1970 und GRONLUND 1974, die sich in dieser Hinsicht an dem normbezogenen Testmodell orientieren. Dagegen versucht FRICKE 1972 in einer teststatistischen Veröffentlichung ein spezielles Maß für die Trennschärfe von Aufgaben in lernzielorientierten Tests abzuleiten. BÜSCHER 1974 gibt eine Übersicht über Verfahren für Aufgabenanalysen bei lernorientierten Tests.

Über diese Angaben hinausgehend wäre es für die Zusammenstellung von Aufgabensätzen (vgl. 2.3) sehr nützlich, wenn bei jeder Aufgabe zu ersehen wäre, für welchen Lernzielbereich (z. B. Wissen, Verständnis, Anwendung) sie vermutlich geeignet ist. Schließlich könnte man – als Fernziel – an Lernziel-Testaufgaben-Sammlungen denken, die zu jedem Lernziel eine Anzahl sorgfältig formulierter und erprobter Testaufgaben enthalten.

## 2. Zusammenstellung von Prüfungs- und Aufgabensätzen

### 2.1 Ein Prüfungssatz besteht in der Regel aus mehreren Aufgabensätzen.

**Innerhalb eines Aufgabensatzes sollten die Aufgaben hinsichtlich der Lösungstechnik (Aufgabentypen) einheitlich sein.**

**Innerhalb eines Prüfungssatzes können verschiedene Aufgabentypen verwendet werden, wenn dies notwendig ist.**

**Der Wechsel der Aufgabentypen darf nicht zu Verständnisschwierigkeiten führen.**

Zunächst wird hier unterschieden zwischen dem Aufgabensatz (zum Beispiel mit den Aufgaben für ein Prüfungsfach) und dem (gesamten) Prüfungssatz, der alle Aufgaben enthält,

die bei einer Prüfung verwendet werden. In ähnlicher Weise unterscheidet man zwischen einem Einzeltest und einer Testbatterie, bei der mehrere komplette Tests zusammengefaßt werden. Entsprechend der möglichen Aufgliederung eines selbständigen Tests (etwa eines Rechentests) in einzelne Untertests (für unterschiedliche Rechenleistungen) wäre auch noch eine weitere Unterteilung innerhalb eines Aufgabensatzes denkbar. Die Forderung nach einem einheitlichen Aufgabentyp bezieht sich hier auf die Untereinheit „Aufgabensatz“, es spricht aber sicher nichts dagegen, sie auch im Bezug auf eine kleinere Untereinheit anzuwenden. Entscheidend ist jedoch, daß bei einem Wechsel des Aufgabentyps am Anfang einer Untereinheit eine Erläuterung erfolgt, evtl. auch mit einem Aufgaben- und Lösungsbeispiel. Damit wendet sich die Empfehlung gegen eine Vermischung formal unterschiedlicher Aufgaben, die zu Verständnisschwierigkeiten führt, den Aufwand an notwendigen Erläuterungen vergrößert und die Auswertung und einheitliche Bewertung der Lösungen erschwert.

## **2.2 Die Aufgaben müssen in ihrer Gesamtheit hinsichtlich der Prüfungsanforderungen (z. B. nach den Ausbildungsordnungen und Rahmenlehrplänen) repräsentativ und in ihrem Schwierigkeitsgrad ausgewogen sein.**

Je nachdem, mit welchem Gewicht einzelne Lerninhalte und Lernziele in den jeweiligen Curricula (zum Beispiel Ausbildungsordnungen und Rahmenlehrplänen) vertreten sind, sollten auch entsprechende Aufgabenzahlen für die einzelnen Prüfungsgebiete festgelegt werden.

Für normorientierte Testverfahren – und zum Teil auch schon für programmierte Prüfungen – wird empfohlen, die Aufgaben so zusammenzustellen, daß die mittlere Aufgabenschwierigkeit bei 50 % liegt. Wenn die wenigen Schwierigkeitsanalysen von programmierten Aufgabensätzen verallgemeinert werden können (mittlere Aufgabenschwierigkeit bei 70 %, vgl. IFOL-Institut 1972) und gleichzeitig das konstante Umrechnungsschema „Erreichte Punktzahl-Note-Bestehen der Prüfung“ beibehalten wird, würde dadurch der Anteil der „Prüfungsversager“ beträchtlich ansteigen. Daher wird in der Empfehlung lediglich ein „ausgewogener Schwierigkeitsgrad“ gefordert. Es sind also Aufgaben mit unterschiedlichem Schwierigkeitsgrad(-index) zu verwenden und eine massive Häufung von Aufgaben eines Schwierigkeitsbereiches – besonders der Extrembereiche – zu vermeiden. Im übrigen spielt die Aufgabenschwierigkeit bei der Konstruktion lernzielorientierter Verfahren auch eine geringere Rolle. Wenn man hierbei die Aufgaben mehr nach den Lernzielen auswählt und abschließend den Lernerfolg prüfen will, muß sich bei einem halbwegs erfolgreichen Ausbildungsgang eine mittlere Aufgabenschwierigkeit ergeben, die über 50 % liegt.

## **2.3 Zwischen Wissens- und Verständnisfragen soll ein angemessenes Verhältnis bestehen.**

**Ein Aufgabensatz sollte höchstens zwei Drittel Aufgaben enthalten, die sich ausschließlich auf Wissensfragen beziehen und**

**mindestens ein Drittel Aufgaben, die Verständnis, Interpretation, Problemlösen, Anwendung und sonstige intellektuelle Leistungen betreffen.**

Untersuchungen haben immer wieder gezeigt, daß Prüfungen aller Art, auch nicht sorgfältig konstruierte testartige Prüfungen, überwiegend oder sogar fast ausschließlich Wissensprüfungen darstellen. Zwischen der Fähigkeit, Wissen zu reproduzieren und es anzuwenden, besteht jedoch ein erheblicher Unterschied. Zwar wird bei Prüfungsanforderungen häufig von „Kenntnissen“ gesprochen, aber der Kenntnis-Begriff wird hier in einem sehr viel weiteren Sinn verwendet als der Ausdruck „Wissen“ in der Empfehlung. Da es kaum Ziel der Berufsausbildung sein kann, „totes“ Wissen bei den Auszubildenden anzuhäufen, ist dieser Absatz der Empfehlung von zentraler Bedeutung.

Daß der geforderte Prozentsatz von 30% Nicht-Wissens-Fragen nicht unreal ist, konnte bei gut konstruierten Verfahren demonstriert werden. So steigerten KRUMM und SEIDEL 1970 in einem standardisierten Wirtschaftslehretest den Anteil der Verständnis-, Anwendungs- und Beurteilungsaufgaben gegenüber 3,4% bei Klassenaufgaben und 7% bei Prüfungsaufgaben auf 28% bei den Testaufgaben, die außerdem noch nach ihren Kennwerten den Anforderungen normorientierter standardisierter Verfahren gerecht werden (vgl. S. 24 f dieses Beitrags).

Entsprechend dieser Forderung der Empfehlung sind generell Aufgabenformen vorzuziehen, bei denen möglichst praxisnahe Leistungen verlangt werden (zum Beispiel Aufgaben, die Situationsbeschreibungen, Abbildungen, Zeichnungen, Belege und sonstige Materialien verwenden). Bei allen Aufgaben (auch Wissensfragen) sollte die Formulierung neu sein (d. h. etwa von Lehrbuchformulierungen abweichen) oder das Problem in einem anderen Zusammenhang auftauchen.

Zahlreiche Beispiele für Testaufgaben für ein höheres intellektuelles Niveau sind – nach Lernzielkategorien geordnet – bei BLOOM 1972 dargestellt. Instrukтив sind auch die durch Analyse- und Interpretationen ergänzten Aufgabenbeispiele bei CHAUNCEY u. DOBBIN 1968. Obgleich diese Beispiele dem allgemeinen Bildungsbereich entnommen sind, lassen sie sich un schwer auf die berufliche Bildung übertragen, besonders soweit sie sich auf naturwissenschaftliche Fächer beziehen. Beispiele für gute Textaufgaben zur Leistungsmessung in der medizinischen Ausbildung finden sich bei HUBBARD 1974.

Es hat sich als zweckmäßig erwiesen, bei einem Prüfungsinstrument mit Hilfe einer „Spezifikationstabelle“ gleichzeitig den Inhalts- und den Lernzielaspekt darzustellen oder zu untersuchen. In einer solchen Tabelle wird in Form einer Inhalts-Lernzielmatrix angegeben, wieviel Aufgaben des Aufgabensatzes sich auf die einzelnen Inhalts- und Lernzielbereiche beziehen, wobei für die Lernziele eine grobe Einteilung etwa in die Kategorien Wissen, Verständnis, Anwendung ausreichend ist, (als Beispiel vgl. RÜTTER 1973, S. 244).

Besonders in den Abschnitten 2.2 und 2.3 werden Gesichtspunkte für die Auswahl und Zusammenstellung von Aufgaben dargestellt. Diese Kriterien sind sicher nicht gleichgewichtig: Zunächst wären die Gesichtspunkte „Repräsentativität der Inhalte“ und „Verringerung der Wissensfragen“ zu beachten, dann eine adäquate Schwierigkeitsverteilung und schließlich wäre noch eine hohe Trennschärfe anzustreben, soweit das bei Bevorzugung der anderen Kriterien möglich ist (vorausgesetzt wird, daß nicht-trennscharfe Aufgaben oder solche mit negativer Trennschärfe vorher aus den Aufgabenkatalogen entfernt wurden). Oder mit anderen Worten: Eine Aufgabe mit ungünstigem Schwierigkeitsgrad und geringer Trennschärfe ist trotzdem in einen Prüfungssatz aufzunehmen, wenn sie sich auf ein wichtiges Lernziel bezieht und keine bessere Aufgabe zur Verfügung steht. Zu ergänzen ist noch, daß in einem Aufgabensatz die Aufgaben nach abnehmendem Schwierigkeitsgrad (index) angeordnet werden (man beginnt also mit den leichteren Aufgaben).

## **2.4 Den Aufgaben ist ein Bewertungsvorschlag beizufügen. Aufgaben innerhalb eines Aufgabensatzes sind gleich zu gewichten. Bewertungen von Teillösungen sind gemäß Ziff. 3.6 möglich.**

In diesem Absatz wird nahegelegt, eher ein möglichst einfaches und einheitliches Bewertungsschema zu verwenden (für jede richtige Antwort wird ein Punkt gegeben). Es hat sich gezeigt, daß der Aufwand für komplexe Bewertungsfragen meistens nicht zu einer Erhöhung der Aussagekraft führt.

## **2.5 Wegen der bei programmierten Prüfungen erhöhten Abschreibefahr wird empfohlen, Parallelformen von Aufgabensätzen zusammenzustellen; beispielsweise**

○ die gleichen Aufgaben in verschiedener Reihenfolge,

- die gleichen Aufgaben mit Antwortmöglichkeiten in unterschiedlicher Anordnung,
- verschiedene Aufgabensätze mit gleichschweren Aufgaben.

Hier wäre noch zu ergänzen, daß Parallelformen gesonderter Auswertungen bedürfen, wobei Verwechslungen ausgeschlossen werden müssen. „Echte“ Parallelformen (d. h. verschiedene Aufgabensätze mit gleichschweren Aufgaben) müssen übrigens empirisch überprüft werden, ob sie nach ihrer durchschnittlichen Schwierigkeit und nach anderen Merkmalen, auch wirklich gleich sind. Es ist nach allen bisherigen Erfahrungen nicht möglich, aufgrund der geschätzten Schwierigkeiten unterschiedliche, aber gleichschwere Aufgabensätze zusammenzustellen.

### 3. Bedingungen für die Durchführung und Auswertung programmierter Prüfungen

Häufig wird angenommen, daß eine Prüfung dadurch „objektiv“ wird, daß man programmierte Aufgaben verwendet und die Ergebnisse mit Hilfe von EDV-Anlagen auswertet. Sofern man den im Zusammenhang mit Leistungsmessungen üblichen Begriff der „Objektivität“ verwendet, reichen diese Maßnahmen nicht aus. Unter „Objektivität“ wird hier die Unabhängigkeit der Prüfungsergebnisse von den Prüfern und Auswertern, der Prüfungssituation und den jeweiligen Auswertungsverfahren verstanden. Da aber auch bei Tests und testähnlichen Verfahren das Ergebnis von der Art der Durchführung abhängig ist, (es ist hierbei sehr leicht, bei mangelnder Aufsicht Ergebnisse „weiterzusagen“) liegt es nahe, diesem Punkt besondere Aufmerksamkeit zuzuwenden, mit dem Ziel, diese Einflußgrößen zu vereinheitlichen und damit zu kontrollieren.

#### 3.1 Die Prüfungsteilnehmer sollen vor der Prüfung über das Prüfungsverfahren und die Art der Aufgabenbeantwortung (Lösungstechnik) informiert sein.

Anleitungen zur Bearbeitung von Tests sollen grundsätzlich so abgefaßt sein, als ob der Prüfungsteilnehmer noch nie mit solchen Verfahren in Berührung gekommen wäre. Trotzdem ist es sicher günstig, wenn nach dieser Empfehlung durch entsprechende Vorbereitungen die Ausgangssituation in dieser Hinsicht für alle Teilnehmer möglichst gleich ist. Für eine Vorbereitung dürfte es ausreichen, wenn ein- oder zweimal mit solchen Verfahren (etwa als Übungsarbeit im Berufsschulunterricht oder unter Verwendung von speziellen „Vorbereitungsaufgaben“) gearbeitet wird.

Nach Meinung des Verfassers ist es aber abzulehnen, als Prüfungsvorbereitung extensiv mit Aufgabensammlungen zu „pauken“, da

- auf diese Weise im Unterricht pädagogisch fragwürdige Lernprozesse (Gedächtnistraining, Erlernen von Tricks zur Lösung von Testaufgaben) in Gang gesetzt werden,
- in der Prüfung Aufgaben, die höhere intellektuelle Leistungen voraussetzen, auf diese Weise zu „Wissensaufgaben“ werden können,
- das Prüfungsergebnis durch unkontrollierbare Einflüsse verfälscht werden kann.

Aus gutem Grund sollte man es beibehalten, Prüfungs- und besonders Testaufgaben nicht vorher zugängliche zu machen (vgl. zu dieser Frage KRUMM 1971, S. 104).

**3.2 Es ist sicherzustellen, daß die Prüfungsteilnehmer aufgrund schriftlicher Vorlagen zu Beginn der Prüfung einheitliche und ausreichende Instruktionen zum Prüfungsablauf und zur Prüfungstechnik erhalten. Zusätzliche mündliche Erläuterungen zu den einzelnen Aufgaben sind nicht zulässig.**

Wie bereits erwähnt, ist die „Instruktion“ ein wesentlicher Bestandteil des gesamten Verfahrens. Sie enthält alle notwendigen Informationen einschließlich Übungsbeispiele und soll den Prüfungsleitern, den Aufsichtspersonen und den Prüfungsteilnehmern schriftlich vorliegen. Zu Beginn wird sie üblicherweise laut vorgelesen, wobei die Prüfungsteilnehmer still mitlesen. Dabei werden Übungsbeispiele gelöst und restliche Fragen geklärt, so daß es gerechtfertigt ist, während der Durchführung keine Erläuterungen mehr zu geben.

**3.3 Gerade bei programmierten Prüfungen kommt es darauf an, daß auch die äußeren Bedingungen für die Abwicklung der Prüfung geregelt werden und bestimmten Mindestanforderungen genügen wie zum Beispiel**

- einheitliche Prüfungstermine bei überregionalen Prüfungen,
- Platzbedarf der Prüfungsteilnehmer,
- erlaubte Arbeits- und Hilfsmittel,
- störungsfreier Ablauf,
- Einhaltung der vorgegebenen Bearbeitungszeiten,
- vollständige und rechtzeitige Unterrichtung der Aufsichtspersonen über die Abwicklung der Prüfung.

Eine gut durchgeführte Prüfung bedarf – außer der Zusammenstellung der Prüfungssätze – erheblicher organisatorischer Vorbereitungen, wie beispielsweise

- Zeitplanung,
  - Zusammenstellung und Ordnung aller notwendigen Materialien,
  - Raumbeschaffung,
  - Vorbereitung der Prüfungsleiter und Aufsichtspersonen.
- Jede Unterlassung kann die Durchführung stören und damit das Ergebnis verfälschen. Beispielsweise muß man bei der Auswahl eines Raumes u. a. prüfen, ob
- die Größe ausreicht,
  - genügend Mobiliar zur Verfügung steht,
  - die Sitzordnung geändert werden muß,
  - die Instruktion auch ganz hinten im Raum zu verstehen ist,
  - Lautsprecher für Durchsagen ausgeschaltet werden können,
  - der Pausenraum nicht neben dem Prüfungsraum ist,
  - nicht gerade Bauarbeiten unter dem Fenster des Prüfungsraumes durchgeführt werden.

Zwingend notwendig ist es auch, jede Störung zu vermeiden. Es geht nicht an, daß die Lösung der Aufgaben durch Begrüßungs- oder Ermunterungsansprachen unterbrochen wird. Noch störender für die Bearbeitung dürften sich Erläuterungen auswirken, die für andere Prüfungsteilnehmer gedacht sind, die im gleichen Raum geprüft werden.

Für die Unterrichtung der einzelnen Prüfungsleiter und Aufsichtspersonen und für einen einheitlichen und kontrollierten Ablauf sind „Durchführungsanleitungen“ notwendig, in denen alle Einzelheiten der Abwicklung der Prüfung beschrieben sind.

**3.4 Auch bei der Durchführung programmierter Prüfungen ist die in Ziff. 2.5 zum Ausdruck gebrachte Forderung zu beachten. Im übrigen ist durch die Sitzordnung ein einwandfreier Ablauf der Prüfung zu gewährleisten.**

Parallelformen können in mehrfacher Weise ausgenutzt werden: Meistens wird man die Formulare so verteilen, daß Sitznachbarn unterschiedliche Formen bearbeiten.

Wenn aber ein einheitlicher Prüfungstermin ausnahmsweise nicht möglich war, kann man auch bei der ersten Prüfungsgruppe Form A verwenden, bei der zeitlich späteren Gruppe Form B.

**3.5 Bei der Auswertung durch EDV-Anlagen sind Maßnahmen zu treffen, die Auswertungsfehler ausschließen, zum Beispiel sind:**

- die Belege (Datenträger) vor der Auswertung darauf zu überprüfen, ob die Eintragungen den Erfordernissen der Datenverarbeitung entsprechen,
- bei Beleglesern geeignete Kontrollverfahren zu verwenden, durch die die objektive und zuverlässige Durchführung der Auswertung nachgewiesen wird.

Erfahrungen in den USA bei der jährlichen Auswertung von Millionen Testergebnissen über EDV-Anlagen (vgl. dazu LINDQUIST 1968) zeigen, daß es riskant ist, diesen Abschnitt des Verfahrens zu vernachlässigen. So ist vergleichsweise zu fragen,

- ob durch das verwendete Kennungs- bzw. Identifikationsverfahren Vertauschungen und falsche Zuordnungen (z. B. bei mehreren Aufgabensätzen und Parallelformen) völlig auszuschließen sind;
- wie Belege auf ihre Auswertbarkeit geprüft wurden, wie hoch der Anteil nicht maschinell auswertbarer Belege war und wie sie schließlich verarbeitet wurden;
- in welchem Umfang sich Korrekturen und unterschiedliche Strichstärken auswirken;
- ob Differenzen zwischen den Ergebnissen einer EDV-Auswertung und einer Auswertung mit Schablonen lediglich durch Fehler des konventionellen Verfahrens zu erklären sind;
- ob bei einer wiederholten Auswertung einer Stichprobe von Belegen durch einen anderen Belegleser und/oder ein anderes EDV-System die Ergebnisse identisch sind;
- ob bei kritischen Fällen (den „Prüfungsversagern“) noch zusätzliche Kontrollen erfolgen.

**3.6 Unvollständige bzw. nur teilweise richtige Lösungen sind zu berücksichtigen, wenn es sich um unabhängige und sinnvolle Teillösungen im Rahmen der gestellten Aufgabe handelt.**

Es wurde bereits erwähnt, daß die Bewertung schon bei der Konstruktion einer Aufgabe berücksichtigt werden muß und daß komplizierte Bewertungsvorschriften häufig nicht sehr ergiebig sind. Komplexe Aufgaben mit Teillösungen sind nur dann zu vertreten, wenn das jeweilige Lernziel vereinfachte Verfahren ausschließt. In solchen Fällen sollte man prüfen, ob die Teillösungen nicht als eine Serie selbständiger Aufgaben aufgefaßt werden können, die sich auf das gleiche Problem beziehen.

**3.7 Die Ergebnisse der Prüfung sollen so aufbereitet werden, daß sie sowohl individuell als auch in ihrer Gesamtheit ausgewertet werden können; sie sollen auch in Bezug auf statistische Gütekriterien (z. B. Objektivität und Zuverlässigkeit) Aussagen erlauben.**

Maschinenlesbare Unterlagen von programmierten Prüfungen enthalten eine Vielfalt von Informationen, die mit Hilfe von Standard-EDV-Programmen relativ leicht aufbereitet werden könnten. So ist es möglich, über eine globale Auswertung hinausgehend jeden Prüfungsteilnehmer darüber zu unterrichten, welche — durch die einzelnen Aufgaben repräsentierten — Lernziele er erreicht hat und wo noch besonders schwerwiegende Lücken vorhanden sind. Solche lernzielbezogenen Ergebnismitteilungen sind in gleicher Weise für Betrieb und Berufsschule notwendig,

wobei noch eine zusätzliche Zusammenfassung („Leistungsprofil“ einer Berufsschulklasse oder einer Ausbildungsgruppe in einem Betrieb) sinnvoll ist.

Informationen über den Leistungsstand allgemein lassen sich aus umfassenderen Auswertungen gewinnen. An dieser Stelle erscheint jedoch eine Warnung angebracht: Ergebnisse dieser Art werden nicht selten überinterpretiert in dem Sinne, daß man die Schuld an vermeintlichen oder tatsächlichen Mängeln dem jeweiligen Interessengegner (wahlweise Hauptschule, Berufsschule, Betrieb) zuweist, obwohl die Untersuchungen solche Aussagen nicht erlauben.

Die Qualität eines Instruments für Leistungsmessungen ist besonders aus einer Reihe von Kennziffern (den „statistischen Gütekriterien“) zu ersehen, die bei einer statistischen Verarbeitung der Daten von Stichproben gewonnen werden. Auch hierfür dürfte der Aufwand relativ gering sein, wobei allerdings die Frage zu klären ist, welche statistischen Verfahren, beispielsweise für die Ermittlung der Zuverlässigkeit bei Prüfungsinstrumenten mit Lernzielorientierung angemessen sind. Unter Zuverlässigkeit wird dabei die Genauigkeit der Messung verstanden, unabhängig von dem, was gemessen wird.

**Die Gesamtauswertung ist auf Anforderung den an der Ausbildung und Prüfung beteiligten Institutionen zugänglich zu machen.**

Die Empfehlung spricht sich an dieser Stelle für eine größere Publizität aus. Auch hier ist auf das Beispiel der USA hinzuweisen, wo bei einem sehr förderalistischen bzw. differenzierten Schulsystem zur Zufriedenheit der Beteiligten seit Jahrzehnten zentral entwickelte und ausgewertete Prüfungen in vergleichsweise extrem großen Umfang durchgeführt werden und eine Vielzahl entsprechender Veröffentlichungen vorliegt.

## 4. Anwendung und Erfüllung der Empfehlungen

**4.1 Vorbehaltlich der gesetzlich festgelegten Aufgaben und Zuständigkeiten kommen bei der Anwendung und Erfüllung der obigen Empfehlungen insbesondere folgende Stellen in Betracht:**

- Anbieter des Prüfungssystems bzw. Aufgabenersteller bei: 1.1, 1.2, 1.3, 1.4, 2.1, 2.2, 2.3, 2.4, 2.5, 3.2, 3.5, 3.6, 3.7
- Aufgabenauswahlkommission bzw. Prüfungsausschuß bei: 1.2, 1.3, 2.1, 2.2, 2.3, 2.4, 2.5, 3.6
- Prüfungsausschuß bzw. beauftragte Aufsicht bei: 3.2, 3.3, 3.4, 3.7
- Ausbildungsstätte bzw. Schule bei: 3.1

In diesem Abschnitt wird eine Tendenz zum Ausdruck gebracht, die mit der Entwicklung übereinstimmt, die bei der Leistungsmessung mit Tests im allgemeinen Bildungsbereich vor einigen Jahren zu beobachten war: Die Anwender behalten ihre Entscheidungsbefugnisse, müssen aber von den notwendigerweise recht aufwendigen Konstruktionsaufgaben entlastet werden, wenn die Verfahren hinreichend qualifiziert sein sollen.

## 5. Hinweise für die weitere Entwicklung programmierter Prüfungen

**5.1 Die bei Prüfungen verwendeten Verfahren sollten wegen der Bedeutung der damit getroffenen Entscheidungen hinsichtlich ihrer Güteermkmale standardisierten Berufsleistungstests mit Lernzielorientierung möglichst nahekommen.**

Informelle, nichtstandardisierte Tests, wie sie die programmierten Prüfungen zur Zeit noch darstellen, sind sicher für manche Formen von Lernkontrollen in der beruflichen Bil-

dung geeignet. Langfristig ist es aber nicht zu vertreten, informelle Verfahren zu verwenden für umfassende Prüfungen mit entsprechenden Konsequenzen für die Beteiligten. In Analogie zu dem Terminus „Schulleistungstest“ wird hier der Ausdruck „standardisierter Berufsleistungstest“ vorgeschlagen für qualifizierte Verfahren zur Leistungsmessung mit Tests in der Berufsbildung. Zur Zeit sind dem Verfasser zwei Verfahren in der Bundesrepublik bekannt, die als „standardisierte Berufsleistungstests“ bezeichnet werden könnten, je eins für Metallberufe (BERG und SCHAD 1971) und den kaufmännischen Bereich (KRUMM und SEIDEL 1970). Wie eingangs erwähnt wurde, ist bei standardisierten (im Gegensatz zu informellen) Tests empirisch nachgewiesen, daß die Gütemerkmale erfüllt sind, beispielsweise Kennwerte für die Zuverlässigkeit bestimmter Grenzen nicht unterschreiten. Zu erwähnen wäre noch, daß dabei der Aspekt der Lernzielorientierung mehr in den Vordergrund rücken muß. Ein Unterschied ist noch zu erwähnen: Bei standardisierten Tests wird im allgemeinen immer wieder das gleiche einmal konstruierte und nur in großen Abständen revidierte Instrument eingesetzt. Bei Prüfungen in der beruflichen Bildung verwendet man jedoch Materialien, die speziell für eine Prüfung zusammengestellt werden. Naheliegende Gründe sprechen gegen die wiederholte Verwendung vollkommen identischer Prüfungssätze. Auch hier zeigt die Prüfungspraxis mit Tests in den USA, daß es durchaus möglich ist, auch standardisierte Tests immer wieder modifiziert zusammenzustellen, sofern ein genügend umfassender Katalog empirisch überprüfter Aufgaben zur Verfügung steht und alle Konstruktionsprinzipien beachtet werden (vgl. dazu SEELIG 1967).

**5.2 Es ist zweckmäßig, bei der Entwicklung und Verwendung von Prüfungsverfahren Experten für die Testkonstruktion zu beteiligen und mit öffentlichen und privaten Institutionen zusammenzuarbeiten, die Erfahrungen bei der wissenschaftlichen Entwicklung von Testprogrammen haben.**

Der vorliegende Beitrag macht deutlich, daß die Konstruktion von Tests als qualifizierten Instrumenten zur Leistungsmessung sich im Lauf der Zeit als selbständiges und recht umfassendes Spezialgebiet entwickelt hat. Wenn man das „know-how“ aus diesem Bereich zur Weiterentwicklung der programmierten Prüfungen heranziehen will, sollte man sich der entsprechenden Experten bedienen, in der Regel Diplompsychologen mit dem Arbeitsschwerpunkt Testkonstruktion, für manche Aufgaben auch Pädagogen mit entsprechenden Zusatzqualifikationen. Testentwicklung ist eine typische Teamarbeit. Vorschläge, wie Psychologen bei der Entwicklung von Verfahren zur Lern- und Leistungskontrolle im Bildungsbereich stärker beteiligt werden können, liegen schon lange vor (vgl. SÜLLWOLD 1960), auch zeigen standardisierte Schulleistungstests für allgemeine Schulen (vgl. die Reihe DEUTSCHE SCHULTESTS) wie vorteilhaft sich eine Kooperation von Pädagogen, Fachspezialisten, Testpsychologen und Statistikern auswirkt.

Als Abschluß unserer Überlegungen sind noch einige Bemerkungen über die Möglichkeiten und Grenzen der programmierten Prüfungen und des standardisierten Berufsleistungstests mit Lernzielorientierung als deren Weiterentwicklung angebracht. Die Probleme sind dadurch erheblich vereinfacht worden, daß eingangs lernzielorientierte Messungen als Hauptanwendungsbereich in den Mittelpunkt gestellt wurden. Prüfungen und Kontrollen in der beruflichen Bildung können jedoch höchst unterschiedliche Funktionen erfüllen, die jeweils anders geartete Verfahren erfordern. Dazu kommt noch, daß sich Leistungsmessungen auch auf Verhaltensaspekte beziehen müssen (beispielsweise manuelle Fertigkeiten, allgemeines Arbeitsverhalten, sprachliche Ausdrucksfähigkeit, Sozialverhalten) die

nicht oder kaum durch schriftliche Verfahren testähnlichen Charakters erfaßt werden können. Schließlich entsprechen technisch perfekte, aber punktuelle Prüfungen nicht mehr dem jetzigen Diskussionsstand in der Erziehungswissenschaft. Trotz dieser Einschränkungen könnten die programmierten Prüfungen bzw. die standardisierten Berufsleistungstests mit Lernzielorientierung einmal zentrale Teile eines auch zeitlich differenzierten Systems von Lernkontrollen und Bewertungen mit unterschiedlichen Zielsetzungen darstellen — wenn es gelingt, sie in der Richtung weiterzuentwickeln, die in der vorliegenden Empfehlung verdeutlicht wird.

**Literatur**

- Berg, M. und Schad, E., 1971: Fachkundetest Metall für die Unterstufe. FMB (U). Beih. mit Anleitung, Normtab. Weinheim: Beltz.
- Biglmair, F., 1969: Leistungsmessung durch informelle Lehrertests. *erziehung*, Jg. 2, Heft 3, S. 22–26, Heft 4, S. 26–28.
- Bloom, S. et al., 1972: *Taxonomie von Lernzielen im kognitiven Bereich* (Übers.) Weinheim: Beltz.
- Büscher, P., 1974: Einige testtheoretische Aspekte kriterienbezogener Leistungsmessung. In: *Leistungsbeurteilung in der Schule*, Hrsg. von K. Heller. Heidelberg: Quelle & Meyer.
- Chauncey, H. u. Dobbin, J. E., 1968: *Der Test im modernen Bildungswesen* (Übers.). Stuttgart: Klett.
- Deutsche Schultests. Weinheim: Beltz.
- Ebel, R. L., 1951: *Writing the Test Item*. In: Lindquist, E. F. (Edit.): *Educational Measurement*, Washington: American Council on Education.
- Fricke, R., 1972: Testgütekriterien bei lernzielorientierten Tests. *Zeitschrift für erziehungswissenschaftliche Forschung*, Heft 3, Jg. 6, S. 150–175.
- Gronlund, N. E., 1974: *Die Anlage von Leistungstests* (Übers.). Frankfurt: Diesterweg.
- Herbig, M., 1972: Aufgabentypen zur Leistungsüberprüfung. In: *Lernzielorientierte Tests*, Hrsg. von K. J. Klauer et al. Düsseldorf: Schwann, S. 74–100.
- Hubbard, J. P., 1974: *Erfolgsmessung der medizinischen Ausbildung* (Übers.). Bern: Hans Huber.
- IFOL, 1972: *Programmierte Prüfungen. Dokumentation im Rahmen eines Forschungsauftrags des BBF* (Unveröffentlichtes Manuskript). Gießen: Institut für objektivierbare Leistungskontrolle.
- Ingenkamp, K. (Hrsg.), 1971: *Tests in der Schulpraxis*. Weinheim: Beltz.
- Ingenkamp, K., 1973: Brauchen wir noch normorientierte Tests? *Lernzielorientierter Unterricht*, Heft 1, Jg. 1, S. 2–6.
- FWU, 1970: *Aufgabenkonstruktion. Objektivierbare Leistungsprüfungen Reihe B (TBR 10)*. München: Institut für Film und Bild in Wissenschaft und Unterricht.
- Klauer, K.-J., 1972: Einführung in die Theorie lernzielorientierter Tests. In: *Lernzielorientierte Tests*, Hrsg. von K. J. Klauer et al. Düsseldorf: Schwann, S. 13–43.
- Krumm, V., 1971: *Testaufgaben an die Schüler verkauft! Die berufsbildende Schule*, Jg. 23, S. 101–105.
- Krumm, V. und Seidel, G., 1970: *Betriebswirtschaftslehretest BWL*. Beih. mit Anleitung u. Normtab. Weinheim: Beltz.
- Lienert, G. A., 1969: *Testaufbau und Testanalyse* (3. Aufl.) Weinheim, Berlin, Basel: Beltz.
- Lindquist, E. F., 1968: *Automatisierte Testauswertung und elektronische Testdatenverarbeitung*. In: Ingenkamp, K. und Marsolek, Th. (Hrsg.). *Möglichkeiten und Grenzen der Testanwendung in der Schule*. S. 477–493, Weinheim: Beltz.
- Rutter, Th., 1973: *Formen der Testaufgabe*. München: Beck.
- Seelig, G. F., 1967: *Schullaufbahnlenkung durch Leistungsmessung „Educational Testing in America“*. *Der Schulpsychologe*, Jg. 14, Heft 1, S. 21–24.
- Seelig, J. F., 1970: *Arbeitsanweisungen für objektivierbare Leistungsprüfungen*. *Die Deutsche Schule*, Jg. 62, S. 51–60, 118–127.
- Süllwold, F., 1960: *Die Mitarbeit des Psychologen bei der Konstruktion von Schulleistungstests*. *Psychologie und Praxis*, Jg. 4, S. 173–175.
- Süllwold, F., 1964: *Schultests*. In: Heiß, R. (Hrsg.): *Psychologische Diagnostik Hdb. d. Psychol.*, Bd. 6, Göttingen: Hogrefe, S. 352–384.
- Wendeler, J., 1969: *Standardarbeiten — Verfahren zur Objektivierung der Notengebung*. Weinheim: Beltz.