

MASCHINELL AUSWERTBARE AUFGABEN IM HOCHSCHULBEREICH

Ein Überblick über Forschung und Anwendung

Autor: Dr. Christian **Ucke** (Dipl.-Phys.),
Technische Universität München

Redaktion: Helmuth Küffner (Dipl.-Psych.), Fernuniversität Hagen

Die vorliegende Broschüre wurde vom Zentrum für Fernstudienentwicklung herausgegeben.

Alle Rechte vorbehalten

© Fernuniversität - Gesamthochschule - Hagen 1979

Inhaltsverzeichnis

	Seite
Vorwort	3
1 Einleitung und Begriffsbestimmung	4
2 Prüfungsformen und maschinell auswertbare Aufgaben	6
2.1 Prüfungsarten	6
2.2 Prüfungsfunktionen	9
3 Formen von maschinell auswertbaren Aufgaben	11
3.1 Begriffserklärung	11
3.2 Antwort-Auswahl-Aufgaben	13
3.3 Sonstige maschinell auswertbare Aufgaben	16
4 Entwicklung und Speicherung von maschinell auswertbaren Aufgaben	19
4.1 Konstruktion	19
4.2 Verbreitung	21
4.3 Itembanken	22
5 Bewertung und Auswertung von maschinell auswertbaren Aufgaben	23
5.1 Berücksichtigung der Ratewahrscheinlichkeit	23
5.2 EDV-Auswertungssysteme	24
6 Möglichkeiten der Anwendung von maschinell auswertbaren Aufgaben im Hochschulbereich	26
6.1 Komplexität von Aufgaben in der Hochschule	26
6.2 Anwendungsbereich nach Stoffklassen und Lehrzielkategorien	27
7 Empirische Untersuchungen zu maschinell auswertbaren Aufgaben	31
7.1 Vergleich von maschinell und durch Korrektoren auswertbaren Aufgaben	31
7.2 Einstellungen	35
7.3 Auswirkungen auf Lehren und Lernen	36
7.4 Lösungsstrategien und Testwiseness	38
7.5 Anzahl von Aufgaben in Tests	39
8 Fazit: Maschinell auswertbare Aufgaben im Hochschulbereich	41
9 Literatur	46
10 Anhang 1: Liste mit Beispielen für Aufgabensammlungen	50
Anhang 2: Übersicht über Testauswertungssysteme im Hochschul- und Fernstudienbereich	52
Anhang 3: Glossar Test- und Aufgabengütekriterien	53

Vorwort

Der Umfang des Einsatzes von Aufgabenheften im Rahmen des maschinellen Korrektursystems für Einsendeaufgaben der Fernuniversität (LOTSE) ist seit den Anfängen im Herbst 1976 auf inzwischen über 100 Aufgabenhefte pro Jahr gestiegen und in der deutschen Hochschullandschaft ohne Beispiel. Die Erfahrungen bei der Entwicklung von maschinenauswertbaren Aufgaben zeigen jedoch häufig, daß die Kenntnisse über die Möglichkeiten dieser Aufgabenformen nicht dem Stand der Forschung entsprechen. Vorurteile - vielleicht erzeugt durch die eigene Führerscheinprüfung oder durch flache, quizähnliche Fragen in Illustrierten oder im Fernsehen - oder auch einfach liebgewordene Gewohnheiten tun ein übriges zu einer Bevorzugung herkömmlicher Prüfungsformen.

Für den weiteren Auf- und Ausbau des maschinellen Korrektursystems an der Fernuniversität erscheint es auch im Hinblick auf die Einführung weiterer Studiengänge in den Natur- und Geisteswissenschaften erforderlich, einmal einen Überblick über die Möglichkeiten und Grenzen maschinenauswertbarer Aufgaben im Hochschulbereich zu geben.

Da eine Magisterarbeit von Herrn Dr. Ucke aus dem Jahre 1976 über dieses Thema vorliegt und das Zentrum für Fernstudienentwicklung den Autor für eine Überarbeitung und Aktualisierung gewinnen konnte, wurden eigene Vorarbeiten eingestellt bzw. sind in die vorliegende Arbeit eingeflossen, so daß der geforderte Überblick bereits Ende 1978 inhaltlich abgeschlossen werden konnte.

Die Arbeit von Herrn Ucke stellt die Vor- und Nachteile von maschinell auswertbaren Aufgaben in verschiedenen Disziplinen unter Aspekten der Didaktik, der Testtheorie und der Ökonomie zusammen, wobei die Darstellung der Aspekte maßvoll und ausgewogen ist. Dabei wird auch die im angelsächsischen Bereich zu dieser Thematik besonders stark vertretene Literatur berücksichtigt, auch wenn eine Übertragung von Ergebnissen aus Gründen unterschiedlicher Hochschulsysteme wie auch unterschiedlicher Einstellungen zu objektiven Prüfungen mit Antwort- Auswahl-Aufgaben usw. nur äußerst behutsam geschehen darf.

Die vorliegende Schrift dürfte für alle Hochschullehrer und Prüfer von Interesse sein, die sich mit der Erstellung und Auswertung von Hausarbeiten, Übungsarbeiten und Klausuren in großem Umfang beschäftigen, sicher auch über den Bereich der Fernuniversität Hagen hinaus.

Es ist der Verdienst des Autors, daß diese Arbeit ohne Vorkenntnisse lesbar ist und alle Fachbegriffe erläutert werden, so daß auch der in Pädagogischer Psychologie nicht so heimische Leser auf ein Fachwörterbuch verzichten kann.

H. Küffner

1 Einleitung und Begriffsbestimmung

Tests mit sogenannten Multiple-Choice-Aufgaben¹, die man im Deutschen auch Antwort-Auswahl-Aufgaben (AAA) nennt, werden schon seit langem durch Rechner ausgewertet. Sie eignen sich dafür besonders günstig aufgrund ihrer eindeutig vorgegebenen Lösung, die von dem Prüfling selbst in eine von einer Maschine direkt lesbare Form (Markierungsbogen, Lochkarte) kodiert werden kann.

Vom Rechner her gesehen sind letztlich alle Aufgaben auswertbar, deren Lösungsmenge eindeutig verschlüsselt werden kann, mit welchem Aufwand auch immer. Es gibt schon Ansätze, Aufsätze mit Rechnern auszuwerten. Es ist allerdings sehr aufwendig, Aufsätze oder gar technische Zeichnungen direkt von einer Maschine lesen und vom Rechner auswerten zu lassen.

Interessant sind jene Aufgaben, deren Lösung von einem opt. Leser automatisch und mit vertretbarem ökonomischen Aufwand in eine computergerechte Form aufgenommen und der Auswertung zugeführt werden können. Derartige Aufgaben werden im folgenden als Maschinell-Auswertbare-Aufgaben (MAA) bezeichnet.

Bei Einbeziehung der zur Zeit vorhandenen Möglichkeiten für eine ökonomische Auswertung umfassen die MAA nicht viel mehr als die AAA, nämlich praktisch nur noch die numerisch codierbaren Aufgaben, d.h. Aufgaben, deren Lösung direkt als Zahl angebar ist. Die Klasse von Aufgaben, die ökonomisch maschinell auswertbar ist, erweitert sich mit der technischen Entwicklung; bislang sind allerdings die Handschriften- und Klarschriftleser zum Einsatz in der Testauswertung kaum geeignet, da hier eine spezielle Schulung der Testteilnehmer bzw. die, Ausstattung mit Spezialschreibmaschinen erforderlich ist.

Wegen des großen Anteils der AAA bei den MAA wird sich dieser Bericht hauptsächlich auf die Verwendung der AAA im Bildungswesen konzentrieren und dabei schwerpunktmäßig auf den Hochschulbereich.

Die Verwendung von MAA nimmt in Deutschland in den letzten Jahren zu. Erwähnt seien hier die Staatsprüfungen in Medizin und Pharmazie, Prüfungen bzw. Klausuren bei Funkkollegs, Telekollegs, bei der Fernuniversität Hagen sowie dem Forschungs- und Entwicklungsinstitut für objektivierte Lehr- und Lernverfahren, Zertifikate bei Prüfungen der Volkshochschule usw.

Erst durch die Verwendung von MAA und deren Auswertung durch Rechner ist es überhaupt möglich, tausend oder zehntausend Prüflinge mit vertretbarem Aufwand zu beurteilen. Die damit verbundene und zum

¹ Anmerkung: Folgende nicht besonders schöne Abkürzungen wurden aus ökonomischen Erwägungen heraus verwendet:

AAA = Antwort-Auswahl-Aufgabe

MAA = Maschinell-Auswertbare Aufgabe

KAA = Korrektor-Auswertbare-Aufgaben

Teil weitgehende Minderung individueller Prüfungs- und Entfaltungsmöglichkeiten ist ein Problem heutiger Massenveranstaltungen im Hochschulbereich, das hier nicht vertieft werden kann, aber auch nicht aus den Augen verloren werden soll.

Eine Zusammenstellung mehrerer Aufgaben kann einen Test bilden, wobei unter Test mit INGENKAMP et al. (1973) folgendes verstanden wird:

"Ein Test ist ein Untersuchungsverfahren, das unter möglichst konstanten Bedingungen interindividuell unterschiedliche Reaktionen möglichst treffsicher (gültig) und zuverlässig erfassen und einer möglichst objektiven und genormten Auswertung zugänglich machen will."

Keineswegs besteht jeder Test nur aus AAA oder MAA, sondern je nach Art des Lehrziels bzw. Prüfungsgegenstands sind auch Aufgaben mit freiformulierten Antworten oder Werkleistungen gefragt.

Aufgaben, die im Rahmen eines Tests gestellt werden, bezeichnet man als Testitem, Aufgabenitem oder einfach Item.

Prüfung bzw. Examen sind übergeordnete Begriffe; eine Prüfung kann einen oder mehrere Tests enthalten, aber auch Teile, die nicht unter die strenge Interpretation eines Tests fallen. Jeder Test ist umgekehrt irgendeine Art von Prüfung, mit welcher Funktion auch immer.

2 Prüfungsformen und maschinenauswertbare Aufgaben

Aufgaben dienen zur Prüfung, sei es eine Prüfung durch jemand anderen (Fremdkontrolle) oder eine Selbstüberprüfung (Selbstkontrolle).

Im folgenden wird kurz auf eine mögliche Klassifikation und einige Funktionen von Prüfungen eingegangen, da sich schon daraus Abgrenzungen für den Einsatz von MAA bzw. AAA ergeben.

2.1 Prüfungsarten

Es ist sicher nicht möglich, Prüfungen in einem eindimensionalen Schema zu klassifizieren. Im weiteren wird der Versuch unternommen, an hand einiger gliedernder Leitideen eine Klassifizierung von Prüfungen zu erleichtern. Die Gliederung lehnt sich zum Teil an BORNE-MANN (in SCHÜTZ et al 1969) an.

Man kann folgende Grundtypen von Prüfungsformen unterscheiden:

Praktische Prüfungen (Herstellung eines Entwurfes oder Konstruktion bei Ingenieuren)
Mündliche Prüfungen (z.B. als Prüfungsgespräch)
Schriftliche Prüfungen (z.B. Klausuren, Hausarbeiten)

Diese Grundtypen von Prüfungsformen lassen sich anwenden bei

Eingangsprüfungen
studienbegleitenden Prüfungen
Zwischen- oder Vorprüfungen
Studienabschluß-/Berufseingangsprüfungen

Sie sind als

Einzelprüfung oder
Gruppenprüfung

durchzuführen.

Es kann sich um eine

Hochschulprüfung	(z. B. Diplom)
Staatsprüfung	(z.B. Staatsexamen)
Handwerkskammerprüfung	(z.B. Meisterprüfung)

handeln.

Man kann unterscheiden zwischen

- Schnelligkeitsprüfungen (mit Zeitbeschränkung) (Schnelligkeitstests)
- Niveauprüfungen, Kapazitätsprüfungen (ohne Zeitbeschränkung) (Niveautests)
- Produktprüfungen (es kommt auf das Ergebnis an)
- Prozeßprüfungen (es kommt auf den Lösungsweg an)
- normorientierten Prüfungen (Orientierung der Prüfung am Durchschnitt)
- lehrzielorientierten Prüfungen (Feststellung, ob ein vorgegebenes Lehrziel erreicht ist)

Mit den zuletzt angeführten Punkten sind auch schon teilweise Funktionen von Prüfungen angesprochen.

Aus den angegebenen Prüfungsformen und -arten lassen sich eine Unzahl von Kombinationen bilden, die hier nicht alle untersucht werden können oder brauchen: auch sind sicher Überschneidungen zwischen einzelnen Arten vorhanden.

MAA dominieren bei den schriftlichen Prüfungen, sind aber auch bei praktischen Prüfungen (z.B. Psychomotorik mit Gerätetests) möglich. Mündliche Prüfungen lassen sich kaum mit MAA durchführen. Die Ergebnisse solcher Prüfungen können evtl. einer Auswertung durch einen Rechner zugeführt werden. Die für praktische oder mündliche Prüfungen wichtigen Kenntnisse oder Fähigkeiten lassen sich teilweise mit MAA erfassen, womit aber wieder die schriftliche Prüfung eingeführt ist. Möglichkeiten der maschinellen Auswertung technischer Zeichnungen (als Ergebnis) oder z.B. irgendwelcher Meßergebnisse einer Apparatur zum Zwecke der Prüfung sind noch recht wenig entwickelt, jedoch nicht ausgeschlossen.

MAA werden bei allen genannten Prüfungsarten eingesetzt, wenn auch in unterschiedlicher Intensität. Sie sind weiter bei Eingangsstudienbegleitenden Prüfungen verbreitet, weniger bei Studienabschlußprüfungen. Sie sind primär bei den Produktprüfungen eingesetzt: für Prozeßprüfungen eignen sie sich nur sehr begrenzt.

Diese wenigen Andeutungen an dieser Stelle mögen genügen, um die Vielfalt möglicher Anwendungen von MAA aufzuzeigen. Die Vor- und Nachteile im einzelnen werden in weiteren Abschnitten angeführt.

Im Zusammenhang mit der Leistungsmessung im Studium hat sich die Unterscheidung von kriteriums- oder lehrzielorientierten Tests zu normorientierten Tests als bedeutsam erwiesen, weshalb im folgenden dazu noch einige Erläuterungen sinnvoll erscheinen.

Man unterscheidet zwischen lehrziel- bzw. kriteriumsorientierten und normorientierten Tests bzw. Prüfungen, da die ersteren ein Ziel oder ein Kriterium festsetzen, das erreicht bzw. getestet werden soll, die zweiten messen die Leistungen der Prüflinge an einer Bezugsgruppe (KLAUER 1974). Eng im Zusammenhang damit steht das Konzept des mastery-learning, d.h. jeder kann im Prinzip ein gesetztes Lehrziel erreichen, nur der Zeitbedarf je Individuum ist verschieden.

Konsequenterweise entfällt auch eine Notengebung herkömmlicher Art bei lehrzielorientierten Tests: das Ziel ist eben erreicht oder nicht.

Aus der häufig vorgenommenen Gegenüberstellung lehrziel- versus normorientierter Tests darf man aber nicht folgern, es handele sich in jeder Hinsicht um einen Gegensatz. Die Setzung eines Lehrziels unterliegt einerseits sich mit der Zeit ändernden Normen bzw. wissenschaftlichen Erkenntnissen. Andererseits richten sich auch normorientierte Tests letztlich nach irgendeinem explizit oder implizit formulierten Lehrziel.

PRICKE (1974) formuliert in Anlehnung an LIENERT folgende Definition:

"Ein lehrzielorientierter Test ist ein wissenschaftliches Routineverfahren zur Untersuchung der Frage, ob und eventuell wie gut ein bestimmtes Lehrziel erreicht ist. Die hierbei verwendeten Testaufgaben sind nicht identisch mit dem Lehrziel, sondern repräsentieren es nur und dienen dazu, den individuellen Fähigkeitsgrad eines Schülers mit einem gewünschten Fähigkeitsgrad zu vergleichen. Für diesen Vergleich sind erforderlich:

- 1. eine Quantifizierung des Lehrziels.*
- 2. eine quantitative Erfassung der Schülerleistung*
- 3. ein Meßmodell für die zufallskritische Entscheidung darüber, ob das Lehrziel erreicht ist.*

Zur Berechnung der Testgütekriterien für diese Tests ist eine spezielle lehrzielorientierte Testanalyse notwendig".

Besonders die Meßmodelle für die Entscheidung, ob ein Lehrziel erreicht ist, sowie die Definition und Berechnung der Testgütekriterien sind noch Gegenstand der Forschung (HERBIG in KLAUER, 1978). Die Quantifizierung von Lehrzielen wird darüber hinaus von einer Reihe von Hochschullehrern als problematisch angesehen. Eine präzise Beschreibung von Lehrzielen z.B. in operationalisierter Form läßt sich sicher nicht in allen Fällen für jedes Niveau gleich gut durchführen. Durch eine zu weitgehende Spezifizierung können übergreifende Lehrziele leicht aus dem Auge verloren werden. Auch kann es personell und ökonomisch fragwürdig werden, die sich im Hochschulbereich in fast jedem Fach immer wieder verändernden Ziele ständig zu überarbeiten.

Diese Feststellungen lassen aber auf keinen Fall den Schluß zu, vage Zielformulierungen würden genügen, den Rest könne und müsse der Student selbst leisten.

Klare Zielbeschreibungen werden nicht nur von Studenten gewünscht. Sie ermöglichen eine bessere Verständigung aller an der Lehre beteiligten Personen. Eine Festlegung der Unterrichts- und Prüfungsmethoden ist damit ja noch nicht vorweggenommen.

Hinweise für die Quantifizierung von Lehrzielen finden sich z.B. in FRICKE (1974). Der mühsame Prozeß der Lehrzielformulierung läßt sich

als Hochschullehrer kaum alleine bewerkstelligen. Lehrzielorientierte Tests bzw. die Ansätze dazu lassen sich nur in Zusammenarbeit mit einem leistungsfähigen Zentrum für die Entwicklung und Auswertung der Tests durchführen.

Aus der Unterscheidung zwischen lehrziel- und normorientierten Tests ergeben sich auch Folgerungen für die Schwierigkeiten von Aufgaben. Während bei lehrzielorientierten Tests die Aufgaben so beschaffen sein sollen, daß sie von denjenigen gelöst werden können, die das Lehrziel erreicht haben, kann es bei normorientierten Tests durchaus sinnvoll sein, Aufgaben zu verwenden, die aufgrund ihrer Schwierigkeit nur von sehr wenigen gelöst werden können, um damit auch noch unter den sehr guten Prüflingen unterscheiden zu können.

Mit diesen Bemerkungen ist schon eine Funktion von Prüfungen angesprochen, nämlich Qualifikation versus Diskrimination.

2.2 Prüfungsfunktionen

Die Funktionen von Prüfungen sind in den letzten Jahren in allen Bereichen des Bildungswesens unter den verschiedensten Blickwinkeln ausgiebig diskutiert worden (z.B. PRAHL, 1976).

Das folgende von FLECHSIG (1974) stammende Schema (leicht erweitert) bringt eine gewisse Gliederung. Es ist zwar primär für den Hochschulbereich konzipiert, läßt sich aber in andere Bereiche entsprechend übertragen.

A. Rekrutierungsfunktionen

- a. Qualifikationsnachweis
- b. Platzierung (Rangreihe), Diskrimination
- c. Selektion

B. Didaktische Funktionen

- a. zeitliche und inhaltliche Gliederung des Studienganges
- b. Orientierung der Lehrenden und Lernenden über Studienziele
- c. Rückmeldung des Lehr-Lernerfolges an Lehrende und Lernende
- d. Motivation
- e. Diagnose von Lernvoraussetzungen

C. Herrschafts- und Sozialisationsfunktionen

- a. Initiation ("Mutprobe"), Disziplinierung
- b. Statusverleihung
- c. Legitimation (Recht zur Ausübung einer Tätigkeit)

Das Gliederungsschema soll nicht den Eindruck hervorrufen, als wenn die genannten Einzelfunktionen unabhängig voneinander und immer deutlich zu trennen wären. Die einzelnen Punkte sind auch keines-

falls gleichgewichtig und es mag sogar noch weitere Funktionen geben.

Was den Gebrauch der Wörter "Rekrutierungs- bzw. Herrschaftsfunktion" anbetrifft, so lehnt sich FLECHSIG damit an Vorgänger an. Er nimmt sogar bewußt den Terminus Sozialisationsfunktion hinzu, um damit den "emotional-negativen Akzent, den das Wort Herrschaft haben könnte, abzubauen". Auch ist mit der Verwendung dieser Worte keineswegs gesagt, daß Prüfungen solche Funktionen unbedingt haben sollen.

Der Zusammenhang zwischen den Funktionen von Prüfungen und den Aufgabenformen ist sehr locker. Die bisherige Praxis verwendet MAA hauptsächlich für didaktische Funktionen. In der Tat haben ja MAA aufgrund ihrer Aufgabenstruktur den Vorteil der schnellen Auswertbarkeit. Dies kann besonders bei der Rückmeldung des Lehr-Lernerfolges einer Veranstaltung zum Tragen kommen. Auch die Diagnose von Lernvoraussetzungen wird mittels Eingangstests schon häufig mit MAA durchgeführt.

Ihren Vorteil der objektiven Auswertung können MAA dann einbringen, wenn es um die Feststellung einer Platzierung geht. Allerdings wird beim Aufstellen solcher Rangreihen vielfach die Objektivität der Auswertbarkeit mit der Validität und der Reliabilität von Tests durcheinandergebracht¹.

Als Nachweis zur Qualifikation werden MAA bei Abschlußprüfungen fast gar nicht eingesetzt, bei Zwischenprüfungen eher. Ausnahmen bilden hier die Medizin und Pharmazie, deren Zwischen- und Abschlußprüfungen zu einem großen Teil mit AAA durchgeführt werden. Gründe für die Einführung lagen dabei wohl weniger in didaktischen Überlegungen, als vielmehr wegen stark gestiegener Studentenzahlen die Prüfung ökonomisch durchführen zu können sowie die unter dem Punkt A. aufgeführten Rekrutierungsfunktionen eindeutiger und vergleichbarer als bisher erfüllen zu können.

In England hat die Universität von London nach langen Diskussionen über den Anteil von AAA an qualifizierenden Prüfungen festgelegt, daß nicht mehr als ein Drittel der Prüfung mit derartigen Fragen bestritten werden soll (OSBORN, 1976). Wie sinnvoll eine solche Begrenzung ohne ein Eingehen auf den Inhalt und das Niveau der Prüfung ist, mag dahingestellt bleiben. Neue Aspekte sind wohl auch durch die sich erweiternden Möglichkeiten von MAA zu erwarten.

¹ vgl. Glossar im Anhang 3

3 Formen von maschinell auswertbaren Aufgaben

3.1 Begriffserklärung

Man ordnet Aufgaben häufig nach der Form der Beantwortung (RÜTTER 1973, HERBIG in KLAUER 1974) und unterscheidet dabei geschlossene, halboffene und offene Aufgaben. Terminologisch gleichbedeutend sind auch die Bezeichnungen Aufgaben mit gebundener, halbgebundener und freier Beantwortung. Schematisch lassen sich diese Formen folgendermaßen beschreiben (PHILIPPS 1978):

Typ	Beschreibung	Beispiel
geschlossene Aufgabe	die Antwort ist sowohl dem Prüfling wie dem Auswerter vorgegeben	Antwortauswahlaufgaben
halboffene Aufgabe	die Antwort ist nicht dem Prüfling, wohl aber dem Auswerter vorgegeben	Aufgaben mit Korrekturaufforderung, Verbesserungsform
offene Aufgabe	die Antwort ist weder dem Prüfling noch dem Auswerter vorgegeben	Aufsatz, essay-test, Beweisführung

Von der Auswertung durch Mensch oder Maschine ist dabei nicht die Rede. Um das Zusammenspiel von Frage und Antwort im Auswertungsprozeß zu kennzeichnen, schlägt PHILIPPS (1978) folgende Beschreibung vor:

geschlossene Aufgabe	die richtige Antwort ist von vornherein festgelegt
halboffene Aufgabe	was als richtige Antwort angesehen wird, ist inhaltlich nicht festgelegt, wohl aber die Form, in der die Antwort zu erfolgen hat
offene Aufgabe	die richtige Antwort ist nicht festgelegt - weder inhaltlich noch formal

Die Auswertung der Aufgaben ist damit in gewisser Weise in die Beschreibung hineingenommen: Aufgaben, die gewissen Formen genügen, sind durch Maschinen auswertbar. Die Antwort auf eine Aufgabe genügt aber fast immer irgendeiner Form, sei sie explizit festgelegt oder implizit durch äußere Begrenzungen erzwungen. Ob eine Aufgabe als offen oder halboffen gilt, hängt dann von der Interpretation ab, was noch als formale Festlegung bezeichnet wird.

Die technische Entwicklung geht ja durchaus dahin, daß die formale Festlegung immer weniger streng sein muß (Markierungs-, Klarschriftleser bis hin zur Mustererkennung). Es gibt inzwischen schon Ansätze,

Aufsätze maschinell auszuwerten (SLOTNICK 1972), jedenfalls Aufgaben, die im herkömmlichen Sinn offene Aufgaben par excellence sind.

Trivialerweise läßt sich der Begriff MAA von seinem logischen Gegenteil der Nicht-Maschinen-Auswertbaren-Aufgaben abheben. Darunter seien hier - nicht mehr streng logisch, dafür aber verständlicher - Aufgaben verstanden, die nur durch menschliche Korrektoren auswertbar sind (sie seien als KAA - Korrektor-Auswertbare-Aufgaben bezeichnet). Nun sind im Prinzip alle MAA auch durch Korrektoren auswertbar. Zum Zwecke der Gegenüberstellung beider Aufgabenarten seien sie hier einmal, wie beschrieben, getrennt.

Die Menge der MAA - und damit natürlich auch der KAA - liegt aber nicht eindeutig fest. Je mehr Aufwand getrieben wird, um so mehr Aufgabenformen werden maschinell auswertbar. Eigentlich interessant ist die automatische Verarbeitung der unmittelbar vom Prüfling gegebenen Antworten bei vertretbarem ökonomischen Aufwand. Unter Berücksichtigung dieser Randbedingung umfassen die MAA zur Zeit kaum mehr als die AAA, nämlich praktisch nur noch die sogenannten numerisch codierbaren Aufgaben, d.h. Aufgaben, bei denen sich die Lösung eindeutig in Form einer Zahl angeben läßt.

Ist die Lösung zwar als Zahl, aber nicht mehr eindeutig angebbar, wie z.B. bei Aufgaben aus der Rechtswissenschaft, bei denen die für die Lösung eines Falles am ehesten zutreffenden Paragraphennummern angegeben werden sollen, reichen häufig die vorhandenen Auswertungsprogramme noch nicht aus.

Aus solchen Gründen werden sich die folgenden Kapitel primär mit AAA befassen.

Zunächst sei noch eine Zusammenstellung einiger Eigenschaften von MAA und KAA gegeben, aus der sich schon erste Hinweise zur Verwendbarkeit entnehmen lassen:

MAA	KAA (Nicht-MAA)
Aufgabenform und -antwort unterliegen starken Einschränkungen	kaum Einschränkungen für Aufgabenform und -antwort
Aufgabe muß klar formuliert sein und an die maschinelle Auswertung angepaßt sein	unzureichend ausformulierte Aufgabenstellungen können bei der Auswertung noch berücksichtigt werden
Aufgabenerstellung beansprucht viel Zeit	Aufgabenerstellung beansprucht weniger Zeit
Wissens- und Verständnisfragen sind leicht zu stellen. Lösungswege verfolgen ist sehr schwierig	Es ist möglich, die Betonung auf den Lösungsweg zu legen, weniger auf das Resultat
Komplexe Aufgabenstellungen schwieriger (aber in Grenzen auch möglich)	Komplexe Aufgabenstellungen sind leichter möglich
Lösungen müssen eindeutig formulierbar sein	Lösungen brauchen nicht eindeutig festzustehen

Prüfling nimmt hauptsächlich Auswahl aus vorgegebenen Antworten vor	Aktive Formulierung der Lösung notwendig, offen für kreative, originelle Lösungen
Möglichkeit des Einprägens von bzw. Verwirrung durch falsche vorgegebene Antworten vorhanden	Möglichkeit des Einprägens von falschen selbstformulierten Antworten bzw. Verwirrung durch unklare Aufgabenstellung vorhanden
ausführliche Aufgaben- und Testanalyse mittels Testgütekriterien möglich	individuelle Kommentare des Korrektors möglich
Prüflinge haben häufig noch wenig Übung mit MAA	Prüflinge sind an diese Aufgabenform gewöhnt
Objektive Auswertung	objektive Auswertung in der Regel nicht möglich
Gesamtzeit von Aufgabenerstellung und Auswertung bei vielen Prüflingen (> 100) und vielen Aufgaben (> 10) geringer (Zahlenangaben dienen nur als grober Anhaltspunkt!)	Gesamtzeit von Aufgabenerstellung und -auswertung bei wenigen Prüflingen (< 100) und wenigen Aufgaben (< 10) geringer

3.2 Antwort-Auswahl-Aufgaben

Eine einigermaßen vollständige Typologie einzelner Formen von AAA mit Kommentaren und Beispielen findet sich in RÜTTER (1973). Es ist nicht möglich und nicht sinnvoll, an dieser Stelle alle Varianten aufzuführen, da es einige Grundformen gibt, auf die sich die meisten anderen Formen zurückführen lassen bzw. die von so ähnlicher Problemstruktur sind, daß sich eine Aufzählung nicht lohnt.

Die folgende Aufzählung unterscheidet nicht nach logischen Gesichtspunkten, sondern nach grundsätzlicher Bedeutung bzw. einfach nach Häufigkeit der Nennung in der Literatur.

Alternativaufgabe	Die Antwort ist als ja/nein richtig/ falsch o.ä. vorgegeben
Einfachauswahlaufgabe auch (1-a-n)-Aufgabe	von n vorgegebenen Alternativen ist eine richtige auszuwählen. Die Alternativaufgabe ist im Grunde eine (1-a-2)-Aufgabe
Mehrfachauswahlaufgabe auch (m-a-n) bzw. (x-a-n)-Aufgabe	m Antworten sind aus n vorgegebenen Alternativen auszuwählen; die Anzahl m ist dem Prüfling bekannt. von (x-a-n)-Aufgaben spricht man, wenn die auszuwählende Anzahl dem Prüfling nicht bekannt ist. Einfachauswahlaufgaben sind im Prinzip (m-a-n)-Aufgaben mit $m=1$.

Bei Antwortauswahlaufgaben dieser Art unterscheidet man den Aufgabenstamm, das ist die Frage, die Feststellung, der Text, zu dem Stellung genommen werden soll, und die Antwortalternativen, deren Zahl zwischen zwei und zehn schwanken kann. Die falschen Alternativen heißen Ablenker oder auch Distraktoren.

Reihenfolgeaufgabe	Es wird nach der besten oder richtigen Reihenfolge gefragt (z.B. eine Anzahl historischer Ereignisse in eine zeitliche Reihenfolge bringen)						
Zuordnungsaufgabe	Aus vorgegebenen Mengen von Elementen werden Elemente einander zugeordnet (z.B. eine Menge mit Hauptstädten einer zweiten Menge mit Ländernamen zuordnen) im Prinzip eine Zusammenstellung zweier oder mehrerer Mehrfachauswahlaufgaben in eine Aufgabe						
Kausale Verknüpfungsaufgabe	<p>Von zwei Aussagen soll festgestellt werden, ob sie selbst falsch oder richtig sind und ob ihre Verknüpfung richtig begründet ist.</p> <p>(Beispiel:</p> <table border="1" data-bbox="711 1099 1321 1373"> <tr> <td data-bbox="711 1099 1015 1200">Berlin ist die Hauptstadt Deutschlands</td> <td data-bbox="1018 1099 1321 1200">Aussage 1</td> </tr> <tr> <td data-bbox="711 1205 1015 1238">weil</td> <td data-bbox="1018 1205 1321 1238">Verknüpfung</td> </tr> <tr> <td data-bbox="711 1243 1015 1373">Berlin die größte Stadt Deutschlands ist</td> <td data-bbox="1018 1243 1321 1373">Aussage 2</td> </tr> </table> <p>Aussage 1 richtig/falsch Aussage 2 richtig/falsch Verknüpfung richtig/falsch</p>	Berlin ist die Hauptstadt Deutschlands	Aussage 1	weil	Verknüpfung	Berlin die größte Stadt Deutschlands ist	Aussage 2
Berlin ist die Hauptstadt Deutschlands	Aussage 1						
weil	Verknüpfung						
Berlin die größte Stadt Deutschlands ist	Aussage 2						

Abwandlungen bzw. Erweiterungen obiger Formen sind z.B. die sogenannte Fallbeschreibung bzw. Interpretationsaufgaben. Bei ihr wird ein Lesetext, eine Tabelle, eine Zeichnung o.ä. an den Beginn gestellt auf den folgend sich dann eine Reihe von Einfach- oder Mehrfachauswahlaufgaben beziehen.

Insbesondere zählen dazu auch jene Aufgaben, bei denen stellvertretend für eine der oben zitierten Aufgabenformen eine Einfach- oder Mehrfachauswahlaufgabe konstruiert wird.

Dies sei an einem Beispiel erläutert:
(1-a-n)-Aufgabe stellvertretend für eine Reihenfolgeaufgabe

Geben Sie die zeitliche Reihenfolge der folgenden Ereignisse an:

- I. Eroberung Galliens durch Caesar**
- 11. Verschwörung des Catilina**
- 111. Bürgerkrieg zwischen Caesar und Pompejus**
- IV. G.S. Gracchus wird Volkstribun**

- Antwortaufgabe:**
- a) I, 11, V, 111
 - b) 11, I, 111, IV
 - c) IV, I, 111, 11
 - d) IV, 11, I, 111

Es ist zu erkennen, daß sich durchaus noch mehr Alternativen zusammenstellen ließen. Durch die Vorgabe der Alternativen in dieser Form werden die prinzipiellen Möglichkeiten der Reihenfolgeaufgaben eingeschränkt.

Meist werden solche Konstruktionen nur deswegen benutzt, weil ein vorhandenes Auswertungsprogramm nur Einfachauswahlaufgaben auszuwerten gestattet oder weil vorhandene Markierungsbogen keine anderen Aufgabenformen zulassen. Im Einzelfall kann dies bedeuten, auf sinnvolle Aufgaben verzichten zu müssen, nur weil sich nicht genügend attraktive Alternativen finden lassen oder zu Hilfskonstruktionen greifen zu müssen, die häufig genug die Aufgabe leichter oder schwieriger machen, ohne am Inhalt etwas zu ändern.

"Nicht die Struktur, sondern alleine der Inhalt einer Aufgabe soll Probleme für den Prüfling aufwerfen" formulieren ALLINGER et al. (1970). Auch wenn man sich nicht voll dieser Meinung anschließt, weil die Struktur einer Aufgabe zur Aufgabe selbst gehören kann, wird man im allgemeinen den Vorrang des Inhalts kaum bezweifeln.

Eine Abhängigkeit der Schwierigkeit einer Aufgabe von seiner Struktur leuchtet nicht nur ein, sondern ist auch experimentell nachzuweisen (MUELLER 1975, WRIGHT et al. 1977). Sie führt auch zur Empfehlung, schwierigere Aufgabenstrukturen (z.B. die kausale Verknüpfung) sparsam zu verwenden (Hinweise für Sachverständige des IMPP, 1974). Eine eindeutige Reihenfolge für die Schwierigkeit der Aufgaben in Abhängigkeit von der Form läßt sich aber nicht angeben, da die Schwierigkeit einer Aufgabe von sehr viel mehr Faktoren als nur der Struktur abhängt (Inhalt, Vorkenntnisse und Übung der Prüflinge, Anordnung der Aufgabe in einem Test usw.).

Zu den Verfechtern von Alternativaufgaben gehört EBEL. In seinen Arbeiten (EBEL 1971, 1972) wägt er deren Vor- und Nachteile ab. Er konstatiert, daß sich mit Alternativaufgaben das Gleiche messen läßt, wie mit Mehrfachauswahlaufgaben, man muß allerdings mehr Aufgaben konstruieren. Interessant für den Einsatz an Hochschulen z.B. für lehr-lernsteuernde Funktionen ist der geringe Konstruktionsaufwand. Beim Einsatz für diagnostische Möglichkeiten spielt das Raten ebenfalls eine geringere Rolle, da Studenten wenig daran interessiert sind, sich selbst zu betrügen. Bei einem Einsatz in qualifizierenden Prüfungen ist eher Vorsicht geboten.

Die Frage der optimalen Anzahl von Alternativen in einer Mehrfachauswahlaufgabe muß unter den Aspekten des Lehrziels bzw. der Funkti-

on der Prüfung betrachtet werden. Mit steigender Zahl von Alternativen ist die Möglichkeit zum Raten verringert, andererseits ist es aber häufig schwierig, mehr als drei attraktive Distraktoren zu finden. Bei feststehender Anzahl der Aufgaben in einem Test sinkt die Reliabilität, je weniger Alternativen vorhanden sind. Hält man die Summe aller Antworten aller Aufgaben konstant, kann man entsprechend mehr Aufgaben geben bis hin zu Alternativaufgaben. Ist man nicht an eine feststehende Anzahl von Antworten gebunden, läßt sich die Trennschärfe von AAA häufig steigern, indem man schlechte Alternativen wegstreicht. Günstig sind Tests bzw. Auswertungsprogramme, die in einem Test Aufgaben mit verschiedener Anzahl von Antworten zulassen (z.B. das Auswertungssystem LOTSE der Fernuniversität Hagen).

Ist die Zahl der Alternativen festgelegt (wie z.B. bei den Prüfungen des Instituts für medizinische und pharmazeutische Prüfungsfragen (IMPP) in Mainz mit (1-a-5)-Aufgaben), so verzichtet man z.T. auf sinnvolle Fragen bzw. muß zu gekünstelten Aufgabenkonstruktionen greifen.

In einem Test sollen die Aufgaben normalerweise unabhängig voneinander sein, d.h. die Lösung einer Aufgabe soll nicht die Lösung einer anderen Aufgabe beeinflussen. Um komplexere Zusammenhänge bzw. sogar Lösungswege zu prüfen, kann man davon bewußt abgehen und ganze Serien aufeinander aufbauender und sich verzweigender AAA konstruieren. Diese Aufgabenfolgen bzw. Testform haben starke Ähnlichkeit mit einer verzweigten Programmierten Unterweisung bzw. mit dem Computerunterstützten Unterricht. Der Unterschied besteht schwerpunktmäßig nur noch darin, daß bei der Programmierten Unterweisung etwas gelernt werden soll und geprüft werden kann und bei dieser Testform

etwas geprüft werden soll und gelernt werden kann. Der Konstruktionsaufwand für derartige Aufgabenfolgen ist allerdings sehr hoch.

Schon dadurch wird der Einsatz beschränkt werden. Auch ist eine sehr große Zahl von Prüflingen notwendig, um die Lösungswege validieren zu können. Immerhin findet diese Testform im Bereich der Medizin in den Vereinigten Staaten Anwendung (HUBBARD 1974, darin auch ein Beispiel).

3.3 Sonstige maschinell auswertbare Aufgaben

In diesem Abschnitt ist ausschließlich die Rede von MAA, die nicht AAA sind.

Dazu gehören zunächst Aufgaben mit numerischen Antworten, wie sie beispielsweise die Fernuniversität Hagen im System LOTSE verwendet. Bei diesem Aufgabentyp ist das Antwortspektrum nicht auf 5 Antwortalternativen beschränkt, sondern es kann ein maximal 9-stelliges Zahlenergebnis (natürliche Zahlen oder Dezimalzahlen) direkt auf dem Markierungsbogen vermerkt werden.

Zur maschinellen Bewertung dieses Aufgabentyps kann als Lösung eine einzelne Zahl oder ein Intervall angegeben werden, der dem gültigen Lösungsbereich entspricht. Numerische Aufgaben werden z.B. gern im Kurs Statistik eingesetzt, in dem in der Aufgabenstellung häufig das Berechnen von statistischen Kennwerten verlangt wird. Beim Ergebnis, das je nach Rechengenauigkeit des verwendeten Taschenrechners unterschiedlich ausfallen kann, ist eine gewisse Toleranz berücksichtigt.

So könnten bei Berücksichtigung eines Minuszeichens auch negative Zahlenwerte lesbar sein oder durch Erfassung einer endlichen Zahl von Lösungswerten neue Anwendungsgebiete der numerischen Aufgabenformen ermöglicht werden. So wäre denkbar, daß in der Rechtswissenschaft als Lösung von Aufgaben die Nummern der Paragraphen erwartet werden, die in einem vorgegebenen Fall herangezogen werden müssen. Die Zahlen könnten zusätzlich gemäß der vorher eingeschätzten Bedeutung für die Lösung des Falles gewichtet werden.

Im Rahmen der maschinellen Korrektur von Hausarbeiten der Kurse Buchhaltung und Rechnungswesen hat MÖLLERS (1977) für die Fernuniversität Hagen ein Programmsystem entwickelt, das es erlaubt, die Buchungsvorgänge auf einem speziell dafür entwickelten Markierungsbogen (vgl. die Abb. auf S. 41) zu erfassen und dem Studenten eine Bilanz in üblicher Form zu erstellen. Buchungsfehler werden erkannt und können ggf. vorn Programm mit voreingestellten Werten überschrieben werden. Da in der Praxis des Rechnungswesens heute ohnehin meist EDV eingesetzt wird, gewöhnt sich der Student auf diese Weise schon an praxisgerechte Verfahren.

Erweiterungen dieser Aufgabenform bzw. Auswertungsform zu Buchstaben oder Wörtern als Antwort bieten sich an und eröffnen Möglichkeiten auch für jene Fächer, in denen Antworten in nichtnumerische Form auftreten.

Das wesentliche ist, daß vor der Auswertung festgelegt werden muß, was und gegebenenfalls mit welchem Gewicht als Lösung gilt.

Je nach Stand der Technik können auch handschriftliche Normschriften oder standardisierte Ziffern direkt vorn Lesegerät erkannt und interpretiert werden (z.B. im KÖLN-Programm, vgl. SIMONS, 1969).

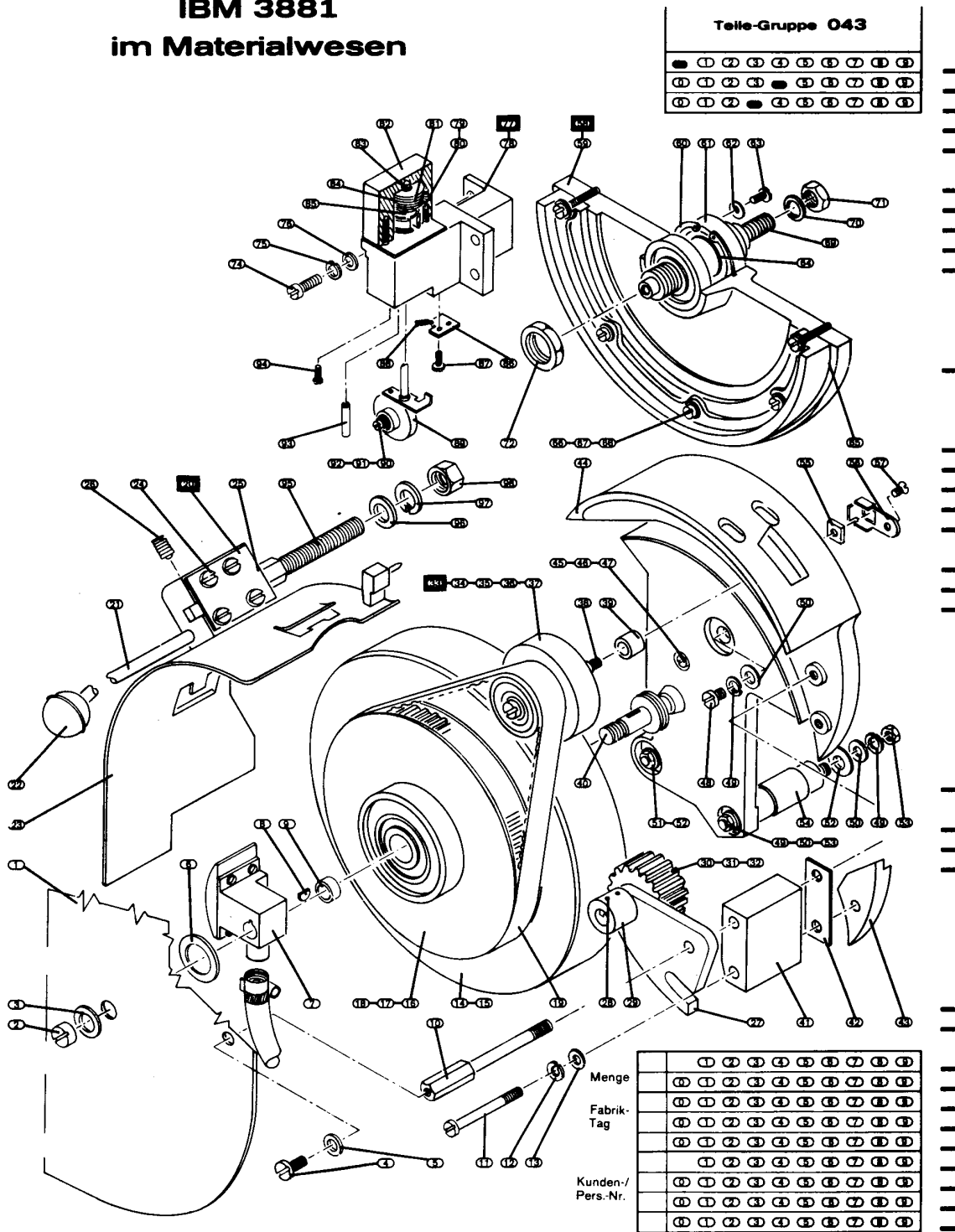
Im Bereich der Ingenieurwissenschaften sind auch Aufgabenstellungen möglich, bei denen auf einem entsprechend gestalteten Antwortbogen mit vorgedruckten Zeichnungen Antworten markiert werden können, z.B. bei der Angabe von Verschleißteilen einer Maschine. Ein auf der nächsten Seite abgebildetes Beispiel aus der Materialbestellung soll hier andeuten, was gemeint ist.

Konsequent weitergedacht könnte auch das Anfertigen einfacher Strichzeichnungen der Auswertung durch optische Lesegeräte zugänglich sein, allerdings sind uns dazu noch keine Anwendungen im Bereich der Testauswertung bekannt.

Die in diesem Abschnitt beschriebenen MAA, die nach herkömmlicher Terminologie als offene Aufgaben gelten, verändern die bisherige

Diskussion über Vor- und Nachteile von AAA versus offene Aufgaben erheblich.

Der optische Markierungsleser IBM 3881 im Materialwesen



IBM 102 628 102 629

4 Entwicklung und Speicherung von maschinell auswertbaren Aufgaben

4.1 Konstruktion

Eine entscheidende Rolle für die Qualität von MAA spielt die Erfahrung des Testkonstruktors. Das gilt insbesondere für die AAA, da durch deren festgelegtere Form einige Probleme auftreten, die bei offenen Aufgaben nicht in der Weise vorhanden sind.

Damit auch weniger Erfahrene nicht alle Anfängerfehler wiederholen, gibt es eine Reihe von kürzeren oder längeren Anleitungen bzw. Regeln zum Aufgabenverfassen. Zur Information sei hier eine aus HUBBARD (1974) entnommene, leicht veränderte Liste wiedergegeben:

Zusammenfassung wichtiger Regeln für die Fragenherstellung:

Wir bieten die folgende Liste als Hilfe an, um einen annehmbaren Standard für Prüfungsfragen aufrechtzuerhalten. Genügt ein Item nicht jeder dieser Standardanforderungen, so sollte man es verwerfen oder revidieren, bzw. überprüfen, ob man zugunsten eines für wichtig gehaltenen Lehrziels bewußt von diesen Ratschlägen abweicht.

1. *Behandelt die Frage einen oder mehrere wichtige Aspekte des Faches?*
Unerhebliche Kleinigkeiten fachlichen Wissens sind zu vermeiden.
2. *Verlangt das Item Informationen, die der Kandidat kennen sollte oder die er ableiten können sollte, ohne ein Nachschlagewerk zu konsultieren?*
Numerische Angaben, terminologische Definitionen u.ä. sollten nur geprüft werden, wenn sie wichtige Informationen darstellen, die zum täglichen Gebrauchswissen des Kandidaten gehören.
3. *Ist das Item dem Wissensniveau angepaßt, das man vom Kandidaten erwarten kann?*
Items, die zu schwierig oder zu leicht sind, taugen nicht zur Diskrimination zwischen guten und schlechten Kandidaten.
4. *Ist das zentrale Problem klar und exakt formuliert?*
Zweideutige oder verschwommene Formulierungen können den Prüfling in die Irre führen und die Gültigkeit der Frage zerstören. Alle erforderlichen Informationen für eine vernünftig begründete Antwort sollten vorhanden sein.
5. *Sind versteckte Hinweise für die richtige Antwort vermieden worden?*
Die häufigsten Hinweise dieser Art sind: die richtige Antwort ist länger und präziser als die Distraktoren, der Gebrauch gleicher Elemente im Stamm und in der richtigen Antwort, unbeabsichtigte Hinweise in der grammatischen Konstruktion, der Gebrauch spezifischer "Determinanten" wie "alle", "keine", "immer" und "nie".
6. *Enthält das Item nicht mehr Wörter als nötig, um es klar und vollständig zu machen?*
Unnötige Wörter erhöhen die Lesezeit; die Prüfung soll fachliches Wissen erfassen, nicht die Lesegeschwindigkeit.
7. *Ist der gewählte Fragentyp der beste für das betreffende Problem?*
Ein Thema, das mit einem Fragentyp schwierig zu prüfen ist, kann vielleicht viel leichter mit einem anderen Fragentyp bewältigt werden.
8. *Ist das Item formal richtig geschrieben?*
Die Auswahlantworten müssen grammatisch mit dem Hauptsatz (Stamm) und untereinander übereinstimmen, usw.

Ähnliche, teils viel ausführlichere Hinweise sind enthalten in:

HERBIG (in KLAUER 1974): Aufgabentypen zur Leistungsüberprüfung

HENRYSSON (in INGENKAMP 1968): Methoden der Konstruktion und Analyse
von Testaufgaben

IMPP (Institut für med. und pharmazeut. Prüfungsfragen): Konstruktion und Auswertung von Multiple-Choice-Prüfungsaufgaben

So einfach und einleuchtend die wenigen, oben angeführten Hinweise erscheinen mögen, um so häufiger werden sie nicht beachtet. Dies wohl weniger aus Unkenntnis als einfach aus Zeitnot, mangelnder Übung oder Möglichkeiten.

Gilt es beispielsweise eine studienbegleitende Zwischenklausur mit AAA bei einer Vorlesung oder Seminar zu erstellen, so tauchen oft die ersten Schwierigkeiten bei der Feststellung auf, was der Student eigentlich konkret können bzw. die Lehrveranstaltung leisten soll (Lehrzieloperationalisierung). Häufig ist es nicht möglich, daß dann konstruierte AAA mit anderen Sachverständigen vor dem Einsatz im Team besprochen werden, abgesehen davon, daß jene wohl etwas vom Inhalt verstehen können aber nicht von Testfragen. In einer mehr forschungsorientierten Hochschule zahlt sich außerdem ein Einsatz zur Verbesserung der Lehre für die eigene Karriere kaum aus.

Der von HYLLEA (in INGENKAMP 1968) ausgesprochene Rat, sich an test-erfahrene Kollegen in anderen Fachbereichen (Pädagogik, Psychologie, Statistik o.ä.) zu wenden, ist zwar gut gemeint aber kaum praktikabel. Zum einen wegen Personalmangel und Arbeitsbelastung gerade in diesen Fächern, zum anderen aber wegen des engen Zusammenhangs zwischen Form und Inhalt einer Aufgabe.

Interessant ist die von MARSTON (1965) untersuchte Möglichkeit, die zu prüfenden Studenten selbst an der Konstruktion der Aufgaben zu beteiligen. Er fand, daß nicht nur die Zufriedenheit am Unterrichtsprozeß bei den so beteiligten Studenten gegenüber den Nichtbeteiligten größer war, sondern daß auch die Leistungen objektiv meßbar stiegen. Immerhin doch wesentliche Ziele eines Unterrichts.

Erwähnt sei auch die Möglichkeit, aus veröffentlichten Aufgabensammlungen (vgl. nächsten Abschnitt) insbesondere auch aus fremdsprachigen Büchern durch Umstellung der vorhandenen Antwortalternativen oder durch Übersetzen der Aufgaben mit verhältnismäßig wenig Aufwand valide Aufgaben zu erhalten.

Aufgabensammlungen mit AAA erscheinen in wachsender Zahl auf dem deutschen Büchermarkt. Beispielsweise hat der Springer-Verlag für die in den medizinischen Examina geprüften Fächer eine ganze Serie herausgegeben; sie enthält Fragen zu Physik, Chemie, Psychologie, Physiologie, Biomathematik, usw. (Reihe Examensfragen in der Medizin). Anlaß der Herausgabe sind die mit AAA abgehaltenen Examen in Medizin. Die Studenten haben die Möglichkeit, sich mit Hilfe derartiger Sammlungen auf die Prüfungen vorzubereiten. Es gilt allerdings dabei Vorsicht walten zu lassen, denn die Prüfungsinhalte werden immer wieder verändert, so daß die Sammlungen nicht mehr angepaßt sein können. Auch ist bei den Büchern häufig nicht auszumachen, ob die

darin enthaltenen Fragen beim gleichen Adressatenkreis getestet sind, Angaben über Schwierigkeit, Trennschärfe usw. fehlen. Derartige Bücher erweisen nach FLÖRKEMEIER u. KERSCHBAUM (1974) den Studenten einen schlechten Dienst. HOFFMANN (1962) spricht von den Gefahren einer Testindustrie, hat allerdings die viel umfangreicheren Aktivitäten in den USA vor Augen.

Im Anhang 1 sind exemplarisch einige Aufgabensammlungen aus verschiedenen Verlagen zusammengestellt, an denen sich der interessierte Leser zunächst orientieren kann. Die Liste erhebt keinen Anspruch auf Vollständigkeit, da es im Rahmen dieser Schrift unmöglich ist, alle derartigen Aufgabensammlungen aufzuführen.

Da besonders im angelsächsischen Bereich zahlreiche Veröffentlichungen zu diesem Thema vorhanden sind, kann hier nur allgemein der Rat gegeben werden, selbst bzw. über die Universitätsbibliothek recherchieren zu lassen.

Schließlich sei noch auf die in den hochschuldidaktischen Zentren oder Arbeitsgruppen mancher Hochschulen vorhandenen qualifizierten Mitarbeiter hingewiesen, die für eine Beratung bei Konstruktion der Aufgaben und damit zusammenhängender Probleme zur Verfügung stehen.

4.2 Verbreitung

Vielfach werden Tests im Hochschulbereich zur Lehr-Lern-Steuerung bzw. zur Qualifikationsermittlung nach Vorlesungen eingesetzt. Diese Tests werden oft von den Anwendern selbst konstruiert. Da das ein mühsames Geschäft ist, kann man es verstehen, wenn die benutzten Aufgaben "geheimgehalten" werden, d.h. konkret, daß die Aufgaben den Studenten nur während der Prüfung selbst zur Verfügung stehen, sei es schriftlich oder projiziert, um danach wieder eingesammelt und nicht weiter zugänglich gemacht zu werden.

Wenn dieses Vorgehen aus vordergründig ökonomischen Gründen vielleicht noch zu vertreten ist, so ist es aus didaktischen Gründen abzulehnen. Es werden dadurch nicht nur unerfreuliche Nebenwirkungen stimuliert (Abschreiben der Items während der Prüfung und das möglicherweise in der Eile auch noch falsch, Beschaffen der Aufgaben auf anderen Kanälen), sondern es muß in weitem Ausmaß auf die pädagogisch wertvolle und bei MAA leicht mögliche schnelle Rückmeldung über richtige oder falsche Lösungen verzichtet werden, da man bei einer Nachbesprechung die gleiche Geheimhaltung praktizieren muß.

Werden andererseits einmal benutzte Items öffentlich zugänglich gemacht, so muß der Testanwender viel Zeit in die Konstruktion neuer Aufgaben investieren. Das kann oft die für die Studenten unerfreuliche Konsequenz haben, daß die Aufgaben schwieriger werden. Meist gibt es nur einen endlichen Vorrat sinnvoller Aufgaben in einem kleinen Teilgebiet, das getestet werden soll. Man kann sich dann helfen, indem man zu komplizierteren Aufgabenstrukturen übergeht; eben damit steigt aber die Schwierigkeit.

4.3 Itembanken

Als Item- oder Fragenbanken bezeichnet man Testfragensammlungen, die einen, nach gewünschten Kriterien festlegbaren Zugriff zu den Fragen innerhalb der Bank gestatten. Dabei braucht es sich nicht nur um AAA oder MAA zu handeln. üblicherweise werden in eine Itembank nur schon erprobte Fragen inklusive der zugehörigen Aufgaben- und Testgütedizes und weiterer Klassifikationsmerkmale (z.B. Lehrziel, Schlüsselworte, Taxonomiestufe ...) aufgenommen.

Itembanken sollen die schnelle Konstruktion von validen und reliablen Tests für ein vorgegebenes Ziel ermöglichen. In einfachen Fällen wird eine "handbetriebene" Sammlung von Fragen auf Karteikarten ausreichen. Ein Rechner zur Speicherung der Fragen und ihrer Kennwerte bietet sich an, wenn die Zahl der Fragen groß ist und die Zahl der Auswahlgesichtspunkte für die Zusammenstellung der Fragen verfeinert wird. Jedoch wird auch im letzteren Fall aus Kostengründen bei Fragen mit mehrfarbigen Zeichnungen, Fotos o.ä. eine Dokumentation mit üblichen Karteikarten neben dem Rechner betrieben.

Die Kosten für eine Itembank mit extensivem Rechnereinsatz sind ziemlich hoch und lohnen sich nur bei großen Anzahlen von Fragen und Studenten und dabei sinnvollerweise im Zusammenhang mit der Auswertung der Fragen durch Rechner (BYRNE 1976). In der Bundesrepublik Deutschland gibt es nach INGENKAMP (1975) im Hochschulbereich keine entsprechende Itembank.

Inzwischen hat sich die Situation jedoch verändert. Die Fernuniversität Hagen hat 1977 mit dem Aufbau eines derartigen Systems begonnen. Das Institut für medizinische und pharmazeutische Prüfungsfragen in Mainz betreibt für die Staatsprüfungen in Medizin und Pharmazie bereits Itembanken. Außerhalb des Hochschulbereichs gibt es in der Bundesrepublik vor allem in der Pädagogischen Arbeitsstelle der VHS in Frankfurt Itembanken für die Zusammenstellung von Zertifikatsprüfungen.

Im Hochschulbereich werden sich im allgemeinen nur "handbetriebene" Fragensammlungen lohnen, z.B. wegen der mit der Lehrfreiheit verbundenen Prüfungsfreiheit, wegen der für die gleiche Veranstaltung häufig von Semester zu Semester wechselnden Dozenten und schließlich den sich gerade im Hochschulbereich immer wieder verändernden, weil an den Stand der Forschung angepaßten Lehrinhalten. Besondere Verhältnisse können sich allerdings im Fernstudium oder Volkshochschulen bei einführenden Veranstaltungen ergeben, die gesichertes Wissen vermitteln.

Bestimmendes Kriterium ist das langfristig gesehene Verhältnis Aufwand zu Nutzen einer Itembank. Hochschullehrer können sich trivialerweise nicht nur mit der Konstruktion und Analyse von Testfragen befassen.

5 Bewertung und Auswertung von maschinell auswertbaren Aufgaben

5.1 Berücksichtigung der Ratewahrscheinlichkeit

Sogenannte Rateformeln werden immer wieder zitiert, um bei der Auswertung von Tests mit AAA für Examenszwecke zu berücksichtigen, daß durch reines Raten auch richtige Lösungen erbracht werden können. Die Anwendung derartiger Formeln benachteiligt jedoch "nicht-ratefreudige" Kandidaten, wie umgekehrt die Nichtanwendung die "ratefreudigen" bevorzugt.

Die Rateformeln gehen meist von der falschen Voraussetzung aus, es gäbe nur zwei Wege zu einer richtigen Lösung: Wissen oder Raten (INGENKAMP 1971). Tatsächlich aber gibt es vom klaren Wissen bis zum blinden Raten fließende Übergänge. Häufig sind Teilinformationen vorhanden, die ein "gezielteres" Raten ermöglichen, insbesondere wenn dazu weitere Überlegungen über die logische Struktur der Aufgabe bzw. der Lösungen treten oder einfach Übung mit diesen Aufgabenformen gegeben ist.

Raten kann schon wesentlich durch die Aufgabenform beeinflusst werden. (1-a-n)-Aufgaben begünstigen Raten eher als (x-a-n)-Aufgaben. Die Art der Auswertung kann weiter dazu beitragen, den Einfluß des Ratens zu vermindern. Wird z.B. bei einer (x-a-n)-Aufgabe jede richtige Entscheidung (d.h. das Markieren jeder richtigen Lösung und das Nicht-Markieren jeder falschen Lösung) mit einem Punkt bewertet und jede entsprechend falsche Entscheidung mit einem Minuspunkt (die Minuspunkte werden nur innerhalb einer Aufgabe verrechnet), so dürfte das Raten keine nennenswerte Rolle spielen (z.B. System LOTSE, Fernuniversität Hagen).

Mittels des sogenannten "confidence weighting" hat man versucht, das Raten noch weiter in den Griff zu bekommen. Der Prüfling kreuzt nicht nur die Antwortalternativen selbst an, sondern soll beispielsweise in Prozent angeben, wie sicher er sich seiner Entscheidung ist. Dieses Verfahren erbringt jedoch praktisch kaum zusätzliche Informationen. Die Korrelation zwischen den Ergebnissen von Zwischen-Tests mit und ohne Angabe des Sicherheitsniveaus ist nahe bei eins; der Aufwand für die Prüflinge und Auswerter aber sehr viel größer (FREIBICHLER 1976).

Ist das Raten schon durch die Verwendung von AAA in (x-a-n)-Form bei entsprechender Bewertung verringert, so spielt es schließlich keine Rolle bei den MAA, die offene Aufgaben sind (z.B. Aufgaben mit numerischer Beantwortung).

5.2 EDV-Auswertungssysteme

Für die Auswertung von MAA gibt es in vielen Programmbibliotheken an Hochschulen schon fertige Programme. Sie sind z.T. auch käuflich bzw. man kann die Auswertung als Auftrag durchführen lassen. Dabei sollte man aber deutlich auf die Auswertungsmöglichkeiten der jeweiligen Programme und auch der verwendeten Markierungsbogen achten (vgl. ausführlich: KUFFNER 1976, STALLER et al. 1976). Einige Testauswertungssysteme aus dem Hochschul- und Fernstudienbereich sind im Anhang 2 aufgeführt. Konkret erläutert sei dies an dem System LOTSE (Lernzielorientierte Testaufgaben zur standardisierten Erfolgsbeurteilung; vgl. auch: LOTSE, Informationsheft für Fernstudenten, Hagen 1978/79).

Erlaubt sind bei diesem System (1-a-n)- und (x-a-n)-Aufgaben sowie Aufgaben mit numerischer Beantwortung. Andere Systeme erlauben häufig nur (1-a-n)-Aufgaben!

Beim System LOTSE ist auch eine Gewichtung der Aufgaben z.B. entsprechend ihrer Schwierigkeit möglich. Die Gewichtung von Aufgaben stellt ein gewisses Problem dar. Erfolgt sie entsprechend der Schwierigkeit, so muß diese entweder schon aus früheren Tests bekannt sein oder geschätzt werden. Die Schätzung von Aufgabenschwierigkeiten mißlingt aber selbst erfahrenen Aufgabenkonstrukteuren häufig. Bei schon getesteten Aufgaben muß man sich im klaren sein, daß ihre Schwierigkeit von den anderen Aufgaben im Test (Zahl und Art) abhängt, vom Ort der Aufgaben im Test, von den Testbedingungen usw., d.h. Parametern, die sich immer wieder verändern. Soll schließlich eine doppelt so schwierige Aufgabe zweifache Gewichtung erhalten? Das bedeutet nämlich implizit, daß man einen Bezugspunkt dieser Verhältnisskala braucht. Für eine Gewichtung könnte man ja auch noch die inhaltliche Bedeutsamkeit einer Aufgabe heranziehen, auch wenn die Aufgabe leicht ist. Viele andere Programme verzichten aus solchen Gründen auf eine Gewichtung.

Das System LOTSE sieht ferner eine differenzierte Bewertung auch von Teilleistungen bei Mehrfachauswahlaufgaben vor. Auch wenn sich bei dem hier verwendeten Verfahren der Verrechnung von falschen und richtigen Antworten noch mancherlei Fragen ergeben (z.B. Verhältnis Aufwand/Nutzen), so ermöglicht es doch eine genauere Analyse der individuellen Leistungen bei einzelnen Aufgaben als eine pauschale Richtig-Falsch-Bewertung.

Für die Studenten ist es sehr wichtig, früh genug und möglichst verständlich zu erfahren, wie die Bewertung und Auswertung der Aufgaben erfolgt. Ausführliche Informationshefte leisten beim System LOTSE diese Aufklärungsarbeit.

Bei der Auswertung sogenannter Einsendaufgaben der Fernuniversität Hagen, d.h. Aufgaben zur Selbstkontrolle, werden neben den tatsächlich markierten und den richtigen Lösungen auch noch Kommentare zum Erfolg ausgedruckt, aufgegliedert nach Lehrzielgruppen und bezogen auf ein festgesetztes Kriterium sowie den Durchschnitt der Leistungen der Vergleichsgruppe.

Damit wird dem Prüfling eine gezielte Hilfe für ein weiteres Lernen bzw. Wiederholen an die Hand gegeben. Intern lassen sich weitere Aufgaben- bzw. Testgüteindizes berechnen und den Aufgabenkonstruktoren bzw. Testanwendern zur Verfügung stellen.

Bezieht man sich auf das Schema der Funktionen von Prüfungen (im Abschnitt 2.2), so erfüllt das System LOTSE primär didaktische Funktionen.

Auswertungssysteme, wie z.B. das des Instituts für medizinische und pharmazeutische Prüfungen in Mainz für Staatsprüfungen in Medizin lassen sich eher den Rekrutierungsfunktionen von Prüfungen zuordnen. Entsprechend geben sie für den Studenten keine oder nur verhältnismäßig spärliche Ausdrücke der Ergebnisse ohne Kommentar wieder.

Ein besonderes Problem bilden bei der automatischen Auswertung von Tests mittels Markierungsbogen fehlerhafte Markierungen sowie Übertragungsfehler der Lösung vom Aufgabenheft auf den Markierungsbogen. Fehlerhafte Markierungen führen zur Zurückweisung des Markierungsbogens durch den Markierungsleser bzw. zu einer verfälschten Auswertung. Übertragungsfehler sind besonders ärgerlich für die Prüflinge, wenn eine richtig gelöste Aufgabe als falsch gewertet wird. Beide Fehler können durch Übung der Prüflinge unter Testbedingungen bzw. auch durch adäquate Gestaltung der Markierungsbogen verringert werden (KÜFFNER 1977). Naturgemäß wirken sie sich besonders gravierend bei Prüfungen mit Rekrutierungsfunktionen aus.

Zwischen der Auswertung von Tests per Hand oder per Großrechner, die meist beide zur Folge haben, daß einige Tage vergehen, bis der Prüfling sein Ergebnis weiß, gibt es noch weitere Möglichkeiten, die sich durch geringeren (finanziellen) Aufwand und sofortige Rückmeldung der Ergebnisse an Lernende und Lehrende auszeichnen.

Im einfachsten Fall werden zu normalen Markierungsbogen Auswerteschablonen vorbereitet, so daß durch Übereinanderlegen beider Bogen eine schnelle Übersicht über die Lösungen gewonnen werden kann. Bei diagnostischen Tests kann die Auswertung sogar von den Prüflingen selbst vorgenommen werden. Dies ist bei weitem die billigste Lösung, die auch noch den Vorteil hat, daß die Markierungsbogen später einer genaueren statistischen Analyse zugeführt werden können.

Aufwendiger ist es schon, wenn in einem Hörsaal an jedem Sitzplatz eine Tastatur mit fünf Druckknöpfen angebracht ist, die den fünf Antwortalternativen der vom Dozenten beispielsweise mit Overheadprojektor gestellten AAA entsprechen. Jeder Student drückt den Knopf der ihm richtig erscheinenden Alternative; der Dozent kann sofort quantitativ das Ausmaß der richtigen Lösungen übersehen, den Studenten bekannt geben und seinen Lehrprozeß darauf abstellen (GRÄNICHER 1975).

Wieder andere Systeme arbeiten mit perforierten Lochkarten, die dann ebenfalls für eine weitere maschinelle Auswertung verfügbar sind (PHILIPPS 1978).

6 Möglichkeiten der Anwendung von maschinell auswertbaren Aufgaben im Hochschulbereich

6.1 Komplexität von Aufgaben in der Hochschule

Ein ganz äußerliches Merkmal bei Prüfungsarbeiten bzw. Testaufgaben bildet deren Länge und der Zeitbedarf zur Fertigstellung. In erster Näherung kann man davon ausgehen, dass längere Arbeiten einen höheren Grad von Komplexität besitzen, d.h. je länger eine Arbeit ist, um so mehr lassen sich mit ihr "höhere" Fähigkeiten wie Organisation, Integration, Synthese, Analyse, Originalität und Kreativität aufzeigen und testen.

Folgendes, leicht veränderte Schema von FLECHSIG u. RITTER (1970) möge dies verdeutlichen:

Kontinuität der Komplexität von Testaufgaben

Aufgabentypus	Auswertung/Bewertung
Forschungsprojekt, Examensarbeit (Dissertation, Diplom usw)	durch Schätzskalen (rating scales)
Wochenarbeit, Referat	
Aufsatz, essay-test	
statement (drei bis fünf Sätze) Kurzaufgabe	
Ein-Satz-Antworten	mit Hilfe eines Auswerteschlüssels/automatisch
Ein-Wort-Antworten, MAA	
Auswahl vorgegebener Lösungen	automatisch

Das Schema kann dazu verleiten anzunehmen, für jeden Grad von Komplexität gäbe es nur einen geeigneten Aufgabentypus, der zudem deutlich von anderen zu trennen wäre. Dem ist aber sicher nicht so. Bei der Skala Komplexität ist ein kontinuierlicher Übergang vorhanden; auch ist eine Vertauschung einzelner Aufgabentypen möglich, denn es ist ohne weiteres denkbar, dass mit einer AAA eine komplexere Leistung verlangt wird, als mit einer Ein-Wort-Aufgabe.

Wenden wir uns zunächst den Forschungs- und Examensarbeiten, den Wochenarbeiten und Referaten zu. Man wird darin übereinstimmen, daß sie sich mit MAA nicht mehr durchführen lassen. Diese Arbeiten sollen u.a. die Fähigkeit erweisen, daß man wissenschaftliches Arbeiten gelernt hat und anwenden kann. In ihnen stecken Fähigkeiten wie selbständige Organisation, Integration, Synthese und Analyse vorhandenen oder selbsterstellten umfangreicheren Materials, eventuell Originalität und Kreativität, Einfallsreichtum, ein Problem bzw. Fragen eigenständig finden, präzisieren und lösen können, d.h. auch einen längeren Lösungsweg als sich entwickelnden Prozeß bearbeiten können, schließlich Fertigkeiten wie Bibliographieren und zitieren

können, sprachlich einwandfreies Formulieren, stilistische Fähigkeiten usw.

Es ist hiermit klar genug ausgedrückt, daß es Prüfungen mit kognitiven Anforderungen im Hochschulbereich gibt, die mit MAA nicht durchzuführen sind. Solche Arbeiten bei einer Abschluß- oder Zwischenprüfung o.ä. bilden aber normalerweise nur einen Teil der Anforderungen; andere Teile werden mit Klausuren, mündlichen Prüfungen usw. absolviert. Es sei kurz angemerkt, daß das Staatsexamen in Medizin keine längere Abschlußarbeit erfordert, wie es bei wohl allen anderen Fächern der Fall ist; die Dissertation ist eine freiwillige Angelegenheit, es promovieren ja auch nicht alle Mediziner.

Die Anwendung von MAA bei Abschlußprüfungen ist aber mindestens zum Teil bei jenen Prüfungsabschnitten möglich, in den Klausuren bzw. Aufsätze geschrieben werden müssen. Und viele Prüfungsordnungen enthalten solche Teile. Von weiteren Prüfungen (studienbegleitende Prüfungen, Diagnostikprüfungen) wird später noch die Rede sein.

6.2 Anwendungsbereich nach Stoffklassen und Lehrzielkategorien

Im folgenden ist primär die Rede von AAA, da über die AAA hinausgehenden MAA bisher kaum Forschungsgegenstand sind und uns daher noch keine verallgemeinernden Erfahrungen bekannt geworden sind.

Folgende durch AAA erfaßbaren Lern- und Denkbereiche gibt NUNNALLY (1964) (in GAUDE u. TESCHNER, 1970) an:

- Definieren, z.B.: Elemente, die sich mit Metallen ohne Sauerstoff verbinden heißen ... (Halogene)
- Fakten reproduzieren, z.B.: In Meeresspiegelhöhe beträgt die Siedetemperatur des Wassers ... (100° C)
- begründendes Denken, z.B.: Ein wichtiger Grund für Schwachsinn ist ... (Vererbung)
- Bewerten und Vergleichen, z.B.: Am wichtigsten ist folgendes Gesetz ... (Grundgesetz)
- finales Denken, z.B.: Eine wichtige Aufgabe des Parlaments ist ... (Gesetzgebung)

INGENKAMP (1971) zeigt anhand von Beispielen, daß AAA sich gemäß der Taxonomie von BLOOM (1972) einsetzen lassen zur "Prüfung von Verständnis, Anwendung des Gelernten in neuen Situationen, zur Analyse von Beziehungen, zur Überprüfung von Leistungen im Bereich der Synthese und der Beurteilung. Die Grenzen ihrer Anwendbarkeit hängen

vorn Einfallsreichtum und der Erfahrung der Testkonstrukteure ab. Grenzen der AAA sind im Augenblick aber ohne Zweifel vorhanden, wenn wir Originalität, Kreativität und sprachliche Spontanleistungen erfassen wollen."

GAUDE u. TESCHNER (1970) weisen darauf hin, daß mit AAA folgende Denkleistungen stimuliert werden können:

- Assoziationen zu reproduzieren oder verschiedene Alternativen nach Kriterien klassifizieren
- Problemlösungen selbsttätig zu erarbeiten und mit vorgegebenen Lösungen zu vergleichen (vor allem mathematische und naturwissenschaftliche Themen)
- ein Problem kritisch zu bearbeiten oder Fehler zu finden und Widerspruch zu Text oder Aufgabe anzeigen.

AAA können demnach viele einfache aber auch komplexere Bereiche auf der Verständnis- und Analyseebene messen. Da ein Sachverhalt, der in einer Alternativaufgabe, Zuordnungsaufgabe oder Kurzwortform formuliert ist, häufig in eine oder mehrere AAA umgewandelt werden kann, ist diese Aufgabenform außerordentlich vielseitig und flexibel. AAA in Verbindung mit der Interpretation eines gegebenen Textes, einer Graphik, Tabelle oder ähnlichem erweitern das übliche Spektrum von Aufgaben.

Trotz der Vielseitigkeit der AAA sind die Grenzen nicht zu übersehen (FREIBICHLER 1976):

- AAA beschränken sich auf verbalisierbare Lernergebnisse. Gerade in den höheren Lernzielebenen würde sich oft eine realistischere Aufgabenstellung anbieten (Simulationen, Spiele, diagnostische Problemlösetests).
- AAA sind nicht geeignet für Problemlösungsaufgaben auf mathematischem und naturwissenschaftlichem Gebiet, wenn ein Vergleich der Lösung nicht möglich sein soll (jetzt allerdings geringfügig anders durch MAA).
- AAA sind nicht geeignet für die Prüfung der Fähigkeit, Ideen zu organisieren und zu präsentieren.
- AAA sind nicht geeignet für die Prüfung von Endzielen der Erziehung.
- Mit AAA können nicht längere Prozesse überprüft werden, z.B. Sequenzen des Problemlösens.

FREIBICHLER (1976) führt folgende Übersicht über die Anwendungsschwerpunkte einzelner Aufgabenarten aufgegliedert nach Lehrzielklassen an:

KENNTNISSE	VERSTÄNDNIS	ANWENDUNG	ANALYSE	SYNTHESE
Kurzantwortform Alternativaufgabe Zuordnungsaufgabe Auswahlaufgabe	Auswahlaufgabe z.T. Kurzantwort-Form	Auswahlaufgabe	Auswahlaufgabe Essay-Test (wenn Freiheitsgrad bedeutsam ist)	Kurzantwortform Essay-Test
Bei mathematischen und naturwissenschaftlichen Themen ist gegebenenfalls nur die Kurzantwort-Form angemessen				

Auch aus dieser Tabelle sind die Grenzen der AAA zu ersehen. Erwähnt sei hier auch, daß die Zuordnung von Problemen zu verschiedenen Komplexitätsstufen auch von den subjektiven Erfahrungen der Prüflinge abhängt. Eine Aufgabe in der Klasse Anwendung kann bei entsprechenden Vorerfahrungen eines Prüflings zu einer reinen Kenntnisaufgabe werden.

Je höher die Lehrzielklasse, die man testen möchte, um so mehr Aufgaben sollten auch vorhanden sein, um mit einiger Wahrscheinlichkeit entscheiden zu können, ob das Ziel erreicht ist oder nicht.

Bevor wir uns im nächsten Abschnitt der genaueren Analyse zuwenden, was MAA wirklich erfassen können, soll noch die Diskrepanz zwischen Anspruch und Wirklichkeit von Prüfungen beleuchtet werden.

BLACK (1968) analysierte in England die Inhalte der Fragen bei Physikexamen und stellte dabei fest, daß 40% der Fragen gänzlich aus dem Gedächtnis beantwortet werden konnten und weitere 40% Erinnerungsvermögen und die Fähigkeit, gut bekannte technische Verfahren durchzuführen, ansprachen. Er gelangt zum Schluß, daß die Studenten allein schon durch straff organisiertes Auswendiglernen gute Noten erzielen konnten.

BEARD (1972) berichtet über eine Reihe weiterer Untersuchungen an Universitäten, in denen die gleiche Tendenz gefunden wurde. INGENKAMP (1971) bestätigt das auch für Schulen.

Im Lichte solcher Untersuchungen nimmt sich die Meinung vieler Studenten und Hochschullehrer, MAA könnten "nur" isoliertes Faktenwissen und reproduktives Denken erfassen, merkwürdig aus. Offensichtlich nutzt die Praxis der üblichen Prüfungen mit offenen Fragen die ihnen innewohnenden Möglichkeiten nur sehr begrenzt aus. Das ist natürlich auch kein Freibrief für die Einführung von MAA.

FERNUNIVERSITÄT

Postfach 9 40
5800 Hagen

Kurs Buchhaltung

Identifikationsfeld
Ziffern (auch führende
Nullen) erst schreiben,
dann markieren!

Name _____

Vorname _____

Wohnort _____

Absenddatum _____

Unterschrift _____

Matrikelnummer						
0	0	0	0	0	0	0
1	1	1	1	1	1	1
2	2	2	2	2	2	2
3	3	3	3	3	3	3
4	4	4	4	4	4	4
5	5	5	5	5	5	5
6	6	6	6	6	6	6
7	7	7	7	7	7	7
8	8	8	8	8	8	8
9	9	9	9	9	9	9

Kursnummer						
0	0	4	6	0	4	
1	1	1	1	1	1	
2	2	2	2	2	2	
3	3	3	3	3	3	
4	4	4	4	4	4	
5	5	5	5	5	5	
6	6	6	6	6	6	
7	7	7	7	7	7	
8	8	8	8	8	8	
9	9	9	9	9	9	

Kurseinheit		
0	1	
1	1	
2	2	
3	3	
4	4	
5	5	
6	6	
7	7	
8	8	
9	9	

interne Vermerke		
0	0	
1	1	
2	2	
3	3	
4	4	
5	5	
6	6	
7	7	
8	8	
9	9	

1. Vorgang

Bitte beachten:
Bleistift Härtegrad 2
oder HB verwenden
und kräftig markieren!



Buch-Nr				Kto. Soll				Kto. Haben			
1	1	1	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9	9	9	9

Betrag in DM							Menge			
0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9	9	9

2. Vorgang

Weitere Hinweise:
Diesen Markierungs-
bogen sorgfältig
behandeln und
nicht knittern
oder heften.
Zum Radieren
einen Plastik-
radiergummi
verwenden!

Buch-Nr				Kto. Soll				Kto. Haben			
1	2	0		0	0	0	0	0	0	0	0
1	1	1		1	1	1	1	1	1	1	1
2	2	2		2	2	2	2	2	2	2	2
3	3	3		3	3	3	3	3	3	3	3
4	4	4		4	4	4	4	4	4	4	4
5	5	5		5	5	5	5	5	5	5	5
6	6	6		6	6	6	6	6	6	6	6
7	7	7		7	7	7	7	7	7	7	7
8	8	8		8	8	8	8	8	8	8	8
9	9	9		9	9	9	9	9	9	9	9

Betrag in DM							Menge			
0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9	9	9

3. Vorgang

Engpassfeld

Buch-Nr				Kto. Soll				Kto. Haben			
1	3	0		0	0	0	0	0	0	0	0
1	1	1		1	1	1	1	1	1	1	1
2	2	2		2	2	2	2	2	2	2	2
3	3	3		3	3	3	3	3	3	3	3
4	4	4		4	4	4	4	4	4	4	4
5	5	5		5	5	5	5	5	5	5	5
6	6	6		6	6	6	6	6	6	6	6
7	7	7		7	7	7	7	7	7	7	7
8	8	8		8	8	8	8	8	8	8	8
9	9	9		9	9	9	9	9	9	9	9

Betrag in DM							Menge			
0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9	9	9

7 Empirische Untersuchungen zu maschinell auswertbaren Aufgaben

7.1 Vergleich von maschinell und durch Korrekturen auswertbaren Aufgaben

Zu dieser Frage gibt es eine außerordentlich umfangreiche Literatur im angelsächsischen Bereich. Zusammenfassende Berichte mit zahlreichen Literaturhinweisen findet man z.B. in INGENKAMP (1971), EBEL (1972), FREIBICHLER (1976).

Der Vergleich wird auch häufig zwischen objektiven Prüfungen und Essay-Tests vorgenommen. Prinzipiell die gleiche Fragestellung ist auch bei der Gegenüberstellung MAA - KAA vorhanden.

Betrachtet man Faktoren wie Objektivität und Reliabilität, so ist die AAA bei entsprechend sorgfältiger Konstruktion der offenen Frage überlegen. Die Validität ist bei beiden Aufgabenformen eine Funktion der Lehrzielbestimmung und -analyse. Die Auswertung von AAA mittels Computer oder Schablone und die damit mögliche schnelle Rückmeldung lassen diese Aufgabenform für viele Zwecke überlegen erscheinen. Dafür ist der Konstruktionsaufwand bei AAA größer als bei offenen Aufgaben.

Bei den letzteren ist wiederum die Auswertung schwieriger und zeitaufwendiger.

Vergleicht man einen Test mit einem Aufsatz (mit einer oder mehreren offenen Aufgaben), so kommt es nach EBEL (1972) beim Aufsatz darauf an, daß man zu schreiben versteht, beim Test darauf an, daß man zu lesen versteht. Da man schneller lesen als schreiben kann, lassen sich bei gleicher Testzeit mit AAA größere Bereiche des allgemein zu prüfenden Wissens überdecken.

Ein Aufsatz bringt insgesamt gesehen Zeitersparnis, wenn es um die Prüfung weniger Probanden geht und kein Itempool erprobter AAA vorliegt; bei großen Probandenzahlen bringen Tests ihre Zeitvorteile zur Geltung, z.T. sind bei großen Studentenzahlen Prüfungen kaum noch anders durchzuführen (Funkkollegs mit mehreren zehntausend Teilnehmern), ein Problem, welches erst jetzt mit der Zunahme von Massenuniversitäten akut geworden ist.

Aufgrund der rechnerisch beliebig genau zu ermittelnden Rangfolge von Prüfungskandidaten gestatten Tests eine äußerlich exakte Unterscheidung zwischen den Probanden. Sie eignen sich dadurch vorzüglich zur Selektion, was wahrscheinlich Tests mehr Vorurteile als Vorteile einbringen wird. Äußerlich ist die Unterscheidung insofern, als wegen niemals erreichbarer vollkommener Reliabilität von Tests Änderungen in der ermittelten Rangfolge mit einer gewissen Wahrscheinlichkeit durchaus möglich sind.

Ein Aufsatz ist nicht zu ersetzen, wo es um Wortgewandtheit, Einfallsreichtum, Kreativität, Originalität, stilistische Fähigkeiten,

selbständiges Formulieren, Darstellung eines kompletten Lösungsweges mit Ansatz und Ergebnis usw. geht.

Die AAA scheint häufig gut geeignet zu sein, Differenzierungsvermögen, subtile sprachliche oder logische Unterscheidungen und auch Präzision des Denkens zu erfassen. Die Vorgehensweise, die im Abwägen und Beurteilen vorhandener Alternativen besteht, kommt durchaus in der Praxis vor und ist deshalb keineswegs lebensfern (Diagnose in der Elektronik, Medizin usw., Wirtschaftsentscheidungen o.ä.).

Neben den angeführten teils formalen Vor- und Nachteilen zwischen beiden Formen spitzt sich der Vergleich auf die Frage zu, ob nicht aufgrund der Aufgabenform selbst etwas prinzipiell Verschiedenes gemessen wird. LIENERT (1967) spricht davon, daß AAA das latente Lerngut erfassen, Ergänzungsaufgaben das manifeste Lerngut. Man könnte hinzufügen, offene Aufgaben das frei verfügbare Lerngut.

Wenn solche Unterscheidungen sinnvoll sind, müßten sich in Unterscheidungen, in denen versucht wird, sie zu operationalisieren, auch Unterschiede aufzeigen lassen.

COOK (1955; zitiert in INGENKAMP 1971) hat bei Collegeanfängern im Fach Geschichte je einen Test mit 60 inhaltlich gleichen Fragen in AAA - bzw. offene Fragenform durchgeführt und eine korrigierte Korrelation von 0,99 zwischen den Leistungen in beiden Aufgabenformen erhalten. Dies läßt nur den Schluß zu, daß beide Aufgabentypen dasselbe gemessen haben.

Eine gleichartige Untersuchung hat AULT (1972) an Schülern in amerikanischer Geschichte, allgemeinen Naturwissenschaften und Arithmetik durchgeführt. Es ergab sich nicht nur der gleiche Korrelationskoeffizient von $r = 0,99$. Eine Faktorenanalyse erbrachte keine spezifischen Faktoren, die der AAA oder der offenen Aufgabenform zuzuschreiben wären.

CROMBAG (1970) führte eine Vergleichsuntersuchung bei Jurastudenten im Fach Strafrecht durch. Er erhielt eine korrigierte Korrelation von $r = 0,94$. Eine ebenfalls vorgenommene Faktorenanalyse erbrachte auch in diesem Fall keine aufgabenform-spezifischen Faktoren.

Zitate dieser Art ließen sich vermehren. Es soll aber nur noch eine Untersuchung hier angeführt werden. INGENKAMP (1971) berichtet über eine Arbeit von BRACHT u. HOPKINS (1968), in der die Vorliebe von Studenten zu Tests bzw. Prüfungen in Aufsatzform gemessen wurde und zwar in einem Kurs zu Pädagogischer Psychologie. Danach bevorzugten 61 % der Studenten eine Kombination beider Examenstypen, 26 % bevorzugten Tests, 13 % Aufsatzprüfungen. Interessanterweise ließ sich aber kein signifikanter Zusammenhang zwischen bevorzugtem Examenstyp und Examensergebnis feststellen.

INGENKAMP (1971) resümiert: "Überblickt man die Forschungsergebnisse, so sprechen sie überwiegend dafür, daß objektive Tests die gleichen Fähigkeiten und Leistungen erfassen können, wie sie in den meisten üblichen Prüfungen einschließlich der Aufsätze erfaßt werden".

In einer Fußnote gibt er allerdings zu, daß ein Aufsatz zusätzlich Aufschlüsse über Einstellungen, Eigenschaften, Werthaltungen u.ä. des Verfassers geben könne.

Mir scheint, daß man vorsichtiger formulieren muß, daß Aufsatzprüfungen (bzw. offene Fragen) inhaltlich all das erfassen können, was objektive Tests bzw. AAA messen können, aber nicht umgekehrt; die Menge der von AAA inhaltlich erfaßbaren Kenntnisse und Fähigkeiten sind eine echte Teilmenge der mit offenen Aufgaben erfaßbaren Bereichen.

Die früher angeführten Fähigkeiten der selbständigen Organisation umfangreicheren Materials, einen Lösungsansatz z.B. in der Mathematik aufstellen, den Lösungsweg als Prozeß verfolgen und das Ergebnis diskutieren und werten können, die sich mit Aufsatzprüfungen bzw. offenen Fragen messen lassen, sind ein Indiz für obige Formulierung. Umgekehrt ist mir keine Fähigkeit bekannt, die sich inhaltlich mit AAA aber nicht mit offenen Fragen messen ließe. Wenn manchmal gewisse Aufgabenformen für einige Bereiche sich inhaltlich besser zu eignen scheinen, so betrifft das nur den Intensitätsgrad der Messung, nicht aber die prinzipiellen Möglichkeiten. Der gemeinsam von AAA und offenen Fragen überdeckte Bereich scheint allerdings groß zu sein. Dafür könnten die mitgeteilten, fast unglaublich hohen Korrelationen zwischen beiden Aufgabenformen sprechen. Diese Korrelationen sind aber unter dem Aspekt zu betrachten, daß sich der Vergleich jeweils auf Dinge erstreckte, die eben mit beiden Aufgabentypen zu formulieren und zu messen waren. Stilistische Ausdrucksfähigkeit oder Einfallsreichtum läßt sich praktisch nicht mit AAA messen; ein Vergleich zwischen beiden Aufgabenformen im Hinblick auf diese Fähigkeiten würde wohl sehr viel niedrigere Korrelationen ergeben.

Die obige Formulierung darf allerdings auf keinen Fall so interpretiert werden, daß Tests bzw. AAA überflüssig wären. Nicht nur spricht alles dafür, vorhandene Prüfungsformen wirklich einzusetzen um den gleichen Inhalt zu prüfen: sozusagen verschiedene Wege zum gleichen Ziel. Was eine Reihe formaler Faktoren betrifft, die für die pädagogische Praxis von großer Wichtigkeit sind, besitzt die AAA wesentliche Vorzüge.

Auf obige Relation zwischen AAA und offener Frage kann auch derjenige nicht stolz sein, der Prüfungen in Aufsatzform unreflektiert einsetzt und dabei allzu häufig die gängigen Fehler macht; er verspielt die prinzipiellen Möglichkeiten von Aufsatzprüfungen. Ein durchdachter Einsatz von AAA in Tests kann sicherlich mehr erfassen, als eine unreflektierte Handhabung von Aufsätzen.

In diesem Sinne sind AAA schon deshalb von Nutzen, weil sie zu einer Diskussion über Faktoren, die Prüfungen beeinflussen, mit Anlaß gegeben haben - ähnlich wie die Programmierte Unterweisung über Unterrichtsverfahren. Daß diese Diskussion in der Hochschule zu vielen Fächern noch kaum Zugang gefunden hat, geschweige denn zu einer Änderung der Prüfungspraxis geführt hat, deutet auf einen Nachholbedarf der Hochschuldidaktik in Deutschland.

Manchmal hört man das Argument, daß gute Studenten unabhängig von der Examensform immer gut abschneiden würden. ILLINGWORTH (1963), der im Fach Chirurgie einen objektiven Test mit traditionellen Prüfungsmethoden verglich, zog genau den Schluß, "daß die Studenten, die in einem Test über faktische Kenntnisse sehr gute Noten erhielten, auch die Fähigkeit besitzen, jeden anderen Examenstyp erfolgreich zu bestehen".

BEARD (1972), die diese Arbeit zitiert, berichtet aber auch über andere Experimente, in denen sich obige Vermutung nicht bestätigen ließ: "Offenbar müssen kreative Menschen nicht notwendigerweise auch gute Examenskandidaten sein". Wie man erkennt, hängt die Beurteilung, ob jemand "gut" ist, offenbar von vielen Faktoren ab.

Immerhin kann man doch den Schluß ziehen, daß es nur der Erhöhung der Prüfungsgerechtigkeit dienen kann, die zur Verfügung stehenden Prüfungsformen wirklich auszunutzen.

Daß das Examensergebnis von der Bevorzugung eines Examenstyps unabhängig ist, wurde schon berichtet (BRACHT u. HOPKINS 1968). Dennoch scheint es so, daß gute Studenten eher Aufsatzprüfungen bevorzugen, während es bei schlechteren Studenten so ist, daß sie objektive Tests vor Aufsatzprüfungen wählen (SHARMA 1970).

HOFFMANN (1962) widmet sich besonders der Frage, ob gute Studenten durch AAA bzw. Tests nicht verwirrt und schlechte, oberflächlich denkende Studenten belohnt würden. Er diskutiert eine Reihe von Aufgaben, die aus renommierten amerikanischen Tests stammen, in denen in der Tat die Alternativen so gewählt sind, daß sich bei oberflächlicher Betrachtung die "richtige" Antwort ergibt. Bei näherem Hinsehen und tieferem Nachdenken aber - und dadurch zeichnen sich eben die guten Studenten aus - kommen Zweifel, so daß schließlich eine andere Alternative attraktiver erscheint. Insbesondere die "schwierigen" Aufgaben würden ihre Schwierigkeit durch eine doppeldeutige Formulierung erhalten und dadurch pädagogisch mehr als zweifelhaft werden.

Die angesprochenen Probleme kulminieren in der pädagogisch bedeutsamen Frage, ob es überhaupt sinnvoll und gerechtfertigt ist, Distraktoren anzubieten. Nicht nur, daß man als Student auf diese Weise gezwungen ist, falsche und irreführende Antworten zu durchdenken, selbst wenn man die richtige Lösung bei einer offenen Frage auf Anhieb gewußt hätte, schlimmer noch, daß man möglicherweise etwas falsches besser behält und sich einprägt, als die später mitgeteilte oder ermittelte richtige Lösung. In der Sprache der Lerntheoretiker wird man falsch konditioniert.

SKINNER (1961) schreibt: "Jede falsche Antwort bei einem Test mit mehreren Wahlmöglichkeiten erhöht die Chance, daß ein Student eines Tages die falsche anstelle der richtigen Antwort aus einem lückenhaften Gedächtnis hervorholt".

INGENKAMP (1971) nimmt am Beispiel eines Diktats mit angebotenen falsch geschriebenen Wörtern recht eindeutig Stellung: "Diese Testform steht nicht im Gegensatz zu pädagogischen Bemühungen... Es gibt

auch keinen Beweis dafür, daß sich ein falsch geschriebenes Wort besser einprägt, wenn es mit der Einstellung zum Korrigieren gelesen wird und häufig übersehene Falschschreibungen hinterher noch einmal besprochen werden. Wir hängen hier noch an einem Aberglauben, nach dem Schüler passiv aufnehmende Automaten seien, in die nur richtige Fakten hineingesteckt werden dürfen... Die Kritik, AAA seien pädagogisch nicht zu verantworten, weil sie auch falsche Antworten vorlegten, stammt von einer didaktisch völlig überholten Position und steht im Gegensatz zu dem Bestreben, Schüler zur kritischen Prüfung und Selbstüberprüfung anzuregen".

Man kann sich auch fragen, wie es mit den falschen Antworten steht, die ein Prüfling auf eine offene Frage selbst formuliert. MARTON (1967) formuliert: "Erhält ein Student eine Korrektur im unmittelbaren Anschluß an die Prüfungssituation, so kann diese Korrektur für Fragen mit Auswahlantworten zumindest ebenso effektiv sein, wie für selbstformulierte Antworten".

Es kann pädagogisch sogar sehr sinnvoll sein, auf falsche Antworten einzugehen, anstatt sie möglichst gar nicht zu erwähnen.

In diesem Zusammenhang kann man auch die Frage diskutieren, ob eine schnelle Rückmeldung über die Richtigkeit einer Lösung in jedem Fall positiv zu bewerten ist. Üblicherweise wird diese bei den AAA vorhandene Möglichkeit als Vorteil eingestuft. Auch hier wird eine Antwort davon abhängen, welches pädagogische Ziel gesetzt ist. Ziele wie z.B. Zähigkeit, Sicherheit, Selbständigkeit rechtfertigen auch eine verzögerte Rückmeldung.

7.2 Einstellungen

FLÖRKEMEIER u. KERSCHBAUM (1974) befragten deutsche Studenten der Medizin zu ihren Meinungen über AAA. Fast alle der insgesamt ca. 200 befragten Studenten kannten schon AAA. Eine leichte Mehrheit von diesen (56 %) hatte positive Erfahrungen mit AAA gesammelt. Dennoch fühlten sich die Studenten bei der Bearbeitung von AAA eher verunsichert, frustriert und unterschätzt. Als Prüfungsform wurde die mündliche Prüfung deutlich bevorzugt; man könne sich besser entfalten und habe den größeren Erfolg.

Diese mehr ablehnende Haltung rührt möglicherweise von enttäuschten Hoffnungen her, daß Prüfungen mit AAA leichter wären.

Vor 1970 forderten die Studenten nämlich objektive Prüfungen und betonten die Nachteile mündlicher Prüfungen. FLÖRKEMEIER u. KERSCHBAUM (1974) interpretieren: "Zwischenzeitlich hat es sich wohl herumgesprochen, daß auch die neuen Prüfungen (mit AAA) die Leistungen kontrollieren - noch umfangreicher, repräsentativer und transparenter als bisher - und somit hart gearbeitet werden muß, um erfolgreich zu bestehen".

Studenten der Fernuniversität Hagen im Fach Mathematik für Wirtschaftswissenschaftler stehen nach einer Untersuchung von SCHUMER (1977) den MAA durchaus positiv gegenüber, u.a. weil die Bewertung objektiver und damit gerechter sei, als bei anderen Prüfungsformen.

Die Einstellung zu MAA hängt demnach von der Vorerfahrung sowohl mit MAA als auch mit anderen Aufgabenformen, von individuellen Persönlichkeitsfaktoren, fachstrukturellen Fragen, von den Informationen über die Möglichkeiten verschiedener Aufgabenformen usw. ab.

Dies gilt nicht nur für die Prüflinge, sondern auch für die Anwender (z.B. Hochschullehrer, wissenschaftliche Mitarbeiter). Von ihnen hängt es größtenteils ab, ob MAA überhaupt zum Einsatz gelangen. Ausführliche Untersuchungen dazu sind nur für die allgemeinbildenden Schulen für den Zusammenhang zwischen Einstellungen über Tests und Zahl der eingesetzten Tests vorhanden, (HAASE 1975) wobei davon ausgegangen werden kann, daß die meisten Schultests aus MAA-Aufgaben bestehen.

7. 3 Auswirkungen auf Lehren und Lernen

"Wenn man nicht messen kann, was man lehrt, lehrt man, was man messen kann". (POPHAM nach ELBERS, 1971)

Bereitet man sich als Prüfling auf ein Examen vor, so wird man dabei wohl berücksichtigen, um welche Art von Prüfung es sich handelt. Im Verbalen geschickte Studenten glauben - nicht ohne Grund - z.B. bei mündlichen Prüfungen Nichtwissen zumindest teilweise überspielen zu können. Das zu Prüfende wird möglicherweise nicht so präzise gelernt.

Ein Lehrer wird - bewußt oder unbewußt - bei seinem Unterricht die späteren Prüfungsformen berücksichtigen.

HOPF (1973) zitiert, daß in England aufgrund weitgehender Verwendung von Tests bei der Übergangsauslese zur Sekundarstufe massive Effekte auf den Lehrplan und die Unterrichtsformen aufgetreten seien. Die Schüler wurden einseitig auf die Prüfung vorbereitet. Es bestand keine Möglichkeit, innovative Unterrichtsformen zu erproben.

In Deutschland beklagen viele Hochschullehrer für das vorklinische und klinische Studium der Medizin, daß die in der Approbationsordnung vorgesehenen Staatsexamen mit AAA den Unterricht stark beeinflussen bzw. einengen. Vom Institut für medizinische und pharmazeutische Prüfungsfragen werden sogenannte Gegenstandskataloge herausgegeben, die praktisch einen inhaltlichen Lernzielkatalog für die Prüfungen mit AAA darstellen. Zwar wird betont, daß die Gegenstandskataloge unverbindlich für die - freie - Lehre seien. Dennoch ist zu beobachten, dass Vorlesungen, Übungen u.ä. darauf ausgerichtet werden. Bücher, die genau auf diese Kataloge abgestimmt sind, erscheinen auf dem Markt. Damit kann sehr leicht die von HOPF (1973) beschriebene Verfestigung eintreten. ZENZ (1975) fragt, ob nicht

letztlich durch die Rückkopplung der Prüfungsanforderungen auf das Studierverhalten eine Erziehung zum konvergenten Denken unterstützt wird.

Wenn beispielsweise vorklinischen Medizinstudenten aufgrund der späteren kognitiven Prüfungsanforderungen primär Kenntnisse in Psychologie eingepaukt werden statt affektive und wohl ursprünglich auch intendierte Verhaltensweisen im Umgang mit Patienten zu üben, so stellt das einen erheblichen Eingriff auf Freiheit und Sinn der Lehre dar.

Das Beispiel der Medizin führt uns augenblicklich besonders drastisch vor Augen, welche Rückwirkungen auf die Lehre und das Lernen die Einführung von Tests hat. Jedoch wird ähnliches aus vielen anderen Fächern auch berichtet.

STUMMER (1974) sieht eine große Gefahr darin, daß Tests mit einem Überwiegen von Sachfragen im literarischen Bereich (Anglistik, fremdsprachliche Lektüre) eine Konzentration des Lese- und Lernverhaltens auf das Quasi-Faktische zur Folge hat; keineswegs ein erstrebenswertes Ziel in diesem Bereich.

MELLENBERGH (1972) kommt allerdings in einer Literaturstudie zu dem Ergebnis, daß das Studierverhalten nicht wesentlich von erwarteten Aufgabentyp in der Prüfung beeinflußt werde.

GARDNER (in INGENKAMP 1968) berichtet, daß "wenn Schüler wissen, daß sie mit objektiven Tests geprüft werden, dazu neigen, sich weniger auf größere Abschnitte des Stoffgebietes zu konzentrieren und darauf vorbereiten, ihre Fähigkeit, Gedanken zu gliedern, zu integrieren und auszudrücken unter Beweis zu stellen.

Komplexe intellektuelle Prozesse stehen weniger im Vordergrund; die Betonung liegt auf einfacheren geistigen Tätigkeiten, z.B. dem Lernen von Fakten und speziellen Kenntnissen".

Die widersprüchlichen Aussagen - es handelt sich teils um geäußerte Meinungen, teils um fundierte Untersuchungen - können bedeuten, daß nicht alle wichtigen Variablen hinreichend genau erfaßt sind oder daß der Schwerpunkt der Betrachtungsweise verschieden ist. Auch ist denkbar, daß sich bei Studenten und Dozenten, die an eine Rückwirkung auf Lernen und Lehren glauben, dieser Glaube im Sinne einer sich selbst erfüllenden Prophezeiung bewahrheitet.

Man sollte auch an dieser Stelle berücksichtigen, welche Rückwirkungen denn wohl eine Prüfung in Aufsatzform oder eine mündliche Prüfung haben. Bei der Vorbereitung auf Aufsatzprüfungen kann es passieren, daß die unzweifelhaft vorhandene Bedeutung von Fakten unterschätzt wird und man in der Prüfung ins Spekulieren kommt; bei mündlichen Prüfungen mit einem bekannt launischen Prüfer lernt man womöglich aus Frustration weniger oder aus Angst mehr als sonst.

Ein problematisches Feld ist auch das Verhältnis von notwendigen Fakten und "höheren" Fähigkeiten im Hochschulbereich. Die einen werden allzu leicht unterbewertet und deswegen vernachlässigt, die an-

deren allzuoft beschworen aber nicht adäquat im Unterrichtsprozeß oder Prüfungen wirklich berücksichtigt.

Der steuernde Effekt, den die Art der Prüfungsaufgaben auf die Art des Lernens der Prüflinge haben können, läßt sich durch bewußte Handhabung auch positiv einsetzen. Die mehr Fakten und Grundlagen enthaltenen Teile einer Vorlesung lassen sich zum Beispiel mit objektiven Prüfungen mit MAA testen. Prüfungen in Aufsatzform können dazu dienen, die Fähigkeiten zur Analyse und Synthese umfangreicheren Materials zu erfassen.

Jede Aufgabenform sollte natürlicherweise für die mit ihr primär erfaßbaren Bereiche eingesetzt werden. Eine Kombination von Aufgabenformen kann bei entsprechendem Einsatz und entsprechender Vorbereitung der Prüflinge Einseitigkeiten des Lernens vermeiden.

7.4 Lösungsstrategien und Testwisseness

Für jede Art von Prüfung ist Übung mit der verwendeten Aufgabenform von Nutzen. Dies gilt ebenso für Aufsätze wie für MAA.

Speziell im Zusammenhang bei Tests mit AAA hat man den Begriff der "testwisseness" geprägt. Darunter ist die Fähigkeit eines Prüflings zu verstehen, unabhängig von dem in den Testaufgaben geforderten Wissen durch Anwendung formaler Entscheidungskriterien auf die Struktur der Aufgaben und/oder durch Überlegungen zu den Umständen der Testdurchführung einen hohen Testwert zu erhalten. Folgende Strategien können dabei mit JONGSMA u. POUND (1977) differenziert werden:

I) Unabhängig vom Testkonstrukteur bzw. Testziel:

- a) Zeit ausnützen
- b) Fehler vermeiden
- c) raten
- d) deduktives Schließen

II) Abhängig vom Testkonstrukteur bzw. Testziel

- a) Überlegungen zum Ziel des Tests
- b) Ausnützen weiterer Hinweise

Das Benutzen derartiger Strategien erscheint durchaus legitim. Das Institut für medizinische und pharmazeutische Prüfungsfragen gibt in seinen "Praktischen Hinweisen zur Durchführung der schriftlichen Prüfungen mit AAA nach der Approbationsordnung" (IMPP, Mainz, fortlaufend seit 1975) die Empfehlungen, zunächst alle scheinbar leicht-

ten bzw. schnell lösbaren Aufgaben anzugehen, sodann bei den schwierigeren Aufgaben alle als sicher falsch erkannten Alternativen zu eliminieren und unter den verbleibenden Alternativen zu raten und schließlich unbedingt jede Aufgabe zu bearbeiten, da andernfalls ratetfreundige Kollegen belohnt würden.

Nur leicht abgewandelte Strategien empfehlen sich ja auch für essaytests (Befolgen eines Gliederungsschemas, Einteilen der Zeit, richtiges Zitieren o.ä.). Zwar ist Raten nicht in gleicher Weise wie bei AAA möglich, dafür ermöglichen Aufsätze aber Bluffen (EBEL 1972).

Jeder Testautor sollte bemüht sein, jene Lösungsstrategien nicht zum Tragen kommen zu lassen, die von ihm selbst abhängen. D.h. vor allem, daß bei AAA nicht Nachlässigkeiten bei der Konstruktion der Aufgaben vorkommen, z.B. absurde Alternativen, grammatische Inkonsistenzen zwischen Stamm und Antwort usw. Bei MAA, die nicht AAA sind, stellt sich dieses Problem in geringerem Ausmaß.

Bei einem einzelnen Testautor, den die Prüflinge persönlich kennen, besteht auch sehr wohl die Möglichkeit, daß die Prüflinge seine Schwächen kennen, z.B. eine Neigung für besonders pfiffige Lösungen, Vorliebe für ganz gewisse Themata o.ä. FREUDENTHAL (1974) spricht auch von der Einsicht in die Psychologie des Testproduzenten, die hilfreich für die Lösung sein könne.

Die bisherigen Untersuchungen zur testwiseness haben eine ganze Reihe von zum Teil widersprüchlichen Zusammenhängen erbracht; so gibt es Korrelationen der testwiseness mit der Intelligenz, dem sozioökonomischen Status, dem Alter der Testteilnehmer usw. (JONGSMA u. POUND 1977). Die Untersuchungen reichen aber noch nicht aus, um hinreichend zuverlässige Empfehlungen über Lösungsstrategien an Testteilnehmer in Abhängigkeit von ihren persönlichen Fähigkeiten und Charakteristika zu geben.

Schließlich hängt die Frage der Lösungsstrategien auch stark von der Funktion des Tests als Prüfung ab. Tests für diagnostische Zwecke zur Selbstkontrolle eines Kandidaten unterliegen da anderen Überlegungen als Tests als Prüfung mit Rekrutierungsfunktion.

7.5 Anzahl von Aufgaben in Tests

Die Anzahl der Aufgaben in einem Test hängt vom Lehrziel, von der Aufgabenform, vom Adressatenkreis und weiteren Parametern ab. Durchaus verbreitet sind Tests zur Qualifikationsermittlung mit 50 bis 90 MAA, die in 1,5 bis 2 Stunden zu bearbeiten sind. Auf eine Aufgabe entfällt also durchschnittlich ca. 1 bis 3 Minuten zur Lösung.

"Können Fragen Tiefe besitzen, von denen eine pro Minute beantwortet werden muß?" fragt HOFFMANN (1962) ironisch. Und weiter: "Wird nicht gerade derjenige Kandidat belohnt, der ein breites oberflächliches Wissen hat und nichts tiefer durchdacht hat?"

Nun sprechen eine Reihe von teststatistischen Gründen für eine hohe Aufgabenzahl, die Reliabilität wird erhöht. Äußere Faktoren begrenzen oft die für die Prüfung zur Verfügung stehende Zeit, so daß man leicht auf die angeführten Zusammenhänge kommt. Im Vergleich mit offenen Aufgaben kann man einfach deswegen mehr MAA in einen Test hineinnehmen, weil die Lesegeschwindigkeit größer als die Schreibgeschwindigkeit ist, ein - wenn auch relativ äußerlicher - Hinweis darauf, daß mit verschiedenen Aufgabenformen auch verschiedenes gemessen wird.

Psychologisch gesehen kann sich die Prüfungsangst verringern, wenn nicht alles von einer oder wenigen Fragen abhängt.

Die schwierige Frage der Validität ist damit natürlich noch nicht erledigt. Während man HOFFMANNs Einwand noch dadurch begegnen könnte, indem man nicht 100 Fragen zu hundert Inhalten stellt, sondern z.B. 10 oder mehr Fragen zur Vertiefung eines Punktes formuliert, muß man klar sagen, daß es Fähigkeiten gibt, die mit MAA, wie sie bis jetzt im Gebrauch sind, nicht zu messen sind. Erst die Kenntnis der Möglichkeiten vorhandener Aufgabenformen gestattet allerdings den problemgemäßen Einsatz.

8. Fazit: Maschinell auswertbare Aufgaben im Hochschulbereich

Um zu beantworten, bei welchen Prüfungen in der Hochschule MAA eingesetzt werden können, orientieren wir uns an den eingangs beschriebenen Schemata von Prüfungsarten und -funktionen. Wir behalten außerdem in Erinnerung, daß die mit MAA-Tests inhaltlich erfaßbaren Prüfstoffe eine Teilmenge der mit offenen Fragen bzw. Aufsätzen prüfbareren Gegenstände sind und daß MAA ganze Bereiche, wie psychomotorische Fertigkeiten und affektive Fähigkeiten gar nicht erfassen können; sie können allerdings - und das sollte man nicht unterschätzen - das Wissen über diese Bereiche messen.

An dieser Stelle soll keine vollständige Analyse des Einsatzes von MAA in jedem spezifischen Studienfach geleistet werden; dies ist auch kaum sinnvoll. Vielmehr werden schwerpunktmäßig Einsatzmöglichkeiten beleuchtet, die sich aber z.T. verallgemeinern lassen.

MAA-Prüfungen sind bei denen von ihnen erfaßbaren Inhalten wohl als Prüfungsform am günstigsten geeignet, Qualifikationsnachweise zu erbringen. Den an sich merkwürdig anmutenden Zusatz - bei den von ihnen erfaßbaren Inhalten - habe ich absichtlich gewählt, weil vielfach von Gegnern dieser Testverfahren deren Nichteignung an Bereichen erwiesen wird, in denen sie in der Tat nicht oder schlecht verwendbar sind und dann eine unzulässige Verallgemeinerung vorgenommen wird.

MAA-Tests ermöglichen eine vergleichende Bewertung von Examina in verschiedenen Hochschulen in viel höherem Ausmaß als andere Prüfungsformen. Ein derartiger Vergleich ist in Deutschland bisher kaum vorgenommen worden und vielleicht auch nicht notwendig gewesen (im Gegensatz zu den USA). Geht man aber von einer Endqualifikation aus, die die Hochschule vermitteln soll (Lehrziel), so erscheint es in vielen Fächern legitim, mit hinreichender Sicherheit zu überprüfen, ob die Qualifikation auch erreicht ist.

Bei Studienabschlüssen ist es wegen der Komplexität der - zumindest geforderten - Fähigkeiten, nicht gut denkbar, eine Qualifikation zur Berufsbefähigung nur aufgrund von MAA-Tests zuzusprechen. Andere Prüfungsformen, die ja auch entsprechend andere Qualifikationen erfassen können und damit der Breite der Hochschulausbildung gerecht werden, müssen einfach hinzutreten. Selbst bei dem deutschen Staatsexamen in Medizin, welches fast ausschließlich mit AAA-Tests vollzogen wird, treten in der Schlußphase mündliche Prüfungen dazu. Allerdings herrscht bei diesem Examen insgesamt noch ein beträchtliches Ungleichgewicht möglicher und sinnvoller Prüfungsformen zugunsten von AAA. Man ist wohl im Überschwang der Euphorie der Möglichkeiten objektiven Testens Ende der sechziger Jahre in Deutschland über das Ziel hinausgeschossen.

Interessant in diesem Zusammenhang ist in der Untersuchung von FLÖRKE u. KERSCHBAUM (1974) eine Umfrage unter Medizinstudenten, ob sie AAA für geeignet für wichtige Prüfungen wie Staatsexamen oder Diplom hielten: ca. 70% der Befragten verneinten dies. Hingegen be-

fürworteten sie den Einsatz für weniger wichtige Prüfungen (z.B. Aufnahmeklausur, Kursabschluß) zu etwa 65 % und für Selbstkontrolle (ohne Ausschlußcharakter) sogar zu fast 90 %.

Bei der Qualifikation von Ärzten muß nach PAULI (1975) vom Sicherheitsanspruch der Gesellschaft ausgegangen werden. In mehr oder weniger stark ausgeprägtem Ausmaß trifft dies auch für Ingenieure, Lehrer, ja wohl fast alle Abschlußprüfungen zu; interessant nur, daß es bei den Medizinern so explizit formuliert wird. Um den Sicherheitsanspruch zu gewährleisten muß eine genügend große Prüfungszuverlässigkeit und damit verbunden Prüfungsobjektivität vorhanden sein, wie dies eben bei Tests in besonderem Ausmaß gegeben ist. PAULI unterlässt nicht darauf hinzuweisen, daß Zuverlässigkeit zwar ein notwendiges aber nicht hinreichendes Kriterium für die Qualität eines Evaluationsverfahrens ist. Es sei hinzugefügt, dass mangelnde Zuverlässigkeit bei Prüfungen in Aufsatzform oder mündlichen Prüfungen ein hinreichendes Kriterium für deren schlechte Qualität darstellt.

Hat die Gesellschaft bei der Qualifikation von Ärzten und neuerdings auch Pharmazeuten stets ein besonderes Sicherheitsbedürfnis gehabt und durch die Einführung von Tests befriedigt, so ist für andere Studienabschlüsse nichts entsprechendes geschehen. Ja selbst, wo bestehende Prüfungsordnungen (Diplom, Staatsexamen usw.) eigentlich die Benutzung von Tests nicht untersagen - wenn auch nicht ausdrücklich erwähnen - werden sie de facto kaum verwendet. Bestehende Ausnahmen (z.B. die Prüfungsordnung für das Höhere Lehramt an Berufsschulen in Bayern erlaubt ausdrücklich allerdings in äußerst begrenzter Weise die Verwendung von AAA) oder sporadische Initiativen (AAA wurden beim Hauptdiplom Psychologie an der Universität München benutzt) bestätigen die Tendenz.

An dieser Stelle sei die Frage aufgeworfen, ob manche Studienrichtungen aufgrund ihrer augenblicklichen Fachstruktur und vielleicht sogar prinzipiell eher geeignet sind, selbst in Abschlußexamina mehr AAA zu verwenden als andere, ohne dass damit eine Diskriminierung gemeint sein soll, weil vielleicht mancher fälschlich denkt, sein Fach würde durch eine derartige Aussage auf bloßes Memorieren o.ä. und damit Minderwertigem beschränkt. Vergleicht man beispielsweise Medizin und Mathematik, so scheint mir, daß in der Mathematik aus wenig Prinzipien und Grundlagen durch lange Schlußketten viel gefolgert werden kann, während in der Medizin viele Details bei einer Diagnose beachtet werden müssen, die mehr Faktenwissen und kleinere logische Folgerungen erfordern. Es könnte sich ergeben, daß die Medizin, die ja im Staatsexamen in verhältnismäßig viele Fächer aufgesplittert ist, für die Anwendung von AAA prädestinierter ist.

Um derartige Fragen genauer behandeln zu können, wäre für jedes Fach ein Lehrziel- und Stoffkatalog notwendig, wie er im Zuge einer Studienreform eventuell einmal entwickelt wird.

Was bis jetzt für Abschlußprüfungen gesagt wurde, gilt in ähnlicher Weise für Zwischenprüfungen u.ä. als Qualifikationsnachweis. Zwar befürwortet HYLLE (in INGENKAMP, 1968) in einem Aufsatz über ökonomische und objektive Zwischenprüfungen an Hochschulen ausdrücklich

die Eignung von Tests für Prüfungen. Die Wirklichkeit an Deutschlands Hochschulen hat aber offenbar bisher wenig Notiz davon genommen.

Eine Reihe hochschulinterner "kleinerer" Qualifikationsnachweise wie erfolgreiche Teilnahme an Vorlesungen, Seminaren, Übungen usw. werden durchaus häufiger mittels objektiver Tests vergeben. Es sei angemerkt, daß für diese kleineren Prüfungen in Hochschulen, die in der Summe natürlich entscheidend einen individuellen Studiengang beeinflussen, zumeist keine offiziellen Prüfungsordnungen existieren, so daß die Anwender unabhängiger und flexibler in der Handhabung von Prüfungsformen sind.

Für Studienzulassungstests, die letztlich die Studienqualifikation nachweisen sollen bzw. für Eingangsprüfungen kommt die Entwicklung in Deutschland erst in Gang. In den USA gibt es eine Vielzahl entsprechender Qualifikationstests (vgl. den Abdruck eines "Scholastic Aptitude Test (SAT) in den Empfehlungen des Wissenschaftsrates 1970, Bd. 11). Hochschuleingangsprüfungen können - geeignet entwickelt und zentral angewendet - bildungspolitisch brisante Wirkungen haben, z.B. im Hinblick auf das Abitur.

Natürlich hängt auch hier vom zu prüfenden Gegenstand, d.h. vom Prüfungsziel ab, ob Tests angemessen sind. Ähnlich wie für Studienabschlußprüfungen gilt aber auch hier eine besondere Verantwortung des Anwenders bei Prüfungen, die so entscheidend einen Lebensweg beeinflussen können. Objektive Tests mit einer hohen Reliabilität bieten sich an.

Geradezu hervorragend aufgrund ihrer rechnerischen Auswertbarkeit und Zuordbarkeit zu Notenstufen oder Punktwerten sind objektive Tests zu Zwecken der Plazierung bzw. Herstellung einer Rangreihe und darauffolgend der Selektion geeignet.

EBEL (1972) beschreibt als Hauptzweck der "Multiple-Choice-Aufgaben" direkt die Möglichkeit, mit ihnen zwischen guten und schlechten Leistungen zu diskriminieren. Nun ist dieser Satz vielleicht eher aus einer amerikanischen Situation heraus zu verstehen, in der Rangreihen unter Studenten nicht nur üblich sondern z.T. sogar gewollt werden. Bei der Selektion von Studenten mittels Tests muß man sich darüber im klaren sein, daß auch hier das Risiko vorhanden ist, geeignete Kandidaten zurückzuweisen und ungeeignete zu akzeptieren. Vergleicht man jedoch eine derartige mögliche Falscheinschätzung mit dem entsprechenden Risiko bei anderen weniger reliablen Prüfungsformen, so haben Tests unzweifelhaft Vorteile.

INGENKAMP (1975) resümiert die europäisch-ausländischen Aktivitäten dahingehend, daß die vorliegenden Tests noch stark für Selektionszwecke eingesetzt werden; die didaktischen Möglichkeiten seien forschungsmäßig und anwendungsmäßig noch keineswegs ausgeschöpft.

Neben den Einsatzmöglichkeiten bei den sogenannten Rekrutierungsfunktionen (FLECHSIG 1974) von Prüfungen, bei denen die formalen Vorzüge von AAA wie Objektivität und hohe Reliabilität stark im Vordergrund stehen, spielen bei den didaktischen Funktionen strukturel-

le Vorzüge wie schnelle und eindeutige Auswertbarkeit und Rückmeldung z.B. für lehr-lernsteuernde Prozesse, evtl. motivationale Gründe eine Rolle. In den MAA innewohnenden didaktischen Möglichkeiten (vgl. INGENKAMP 1975) liegen m.E. die wichtigeren Gründe für deren Anwendung; sie werden aber oft zu wenig betont oder ausgenutzt.

Sei zunächst die im Hochschullehrprozeß so oft vernachlässigte Rückmeldung über den Erfolg einer Veranstaltung genannt, am auffälligsten wohl bei Massenvorlesung, deren darbietend-rezeptive Struktur auch Hochmotivierte frustrieren kann.

Kleine, informelle Tests mit wenigen MAA, eingestreut in die Veranstaltung, bieten die Möglichkeit, schnell und sicher zu erfahren, was Studenten noch mitbekommen haben bzw. wo ihr Verständnis aussetzt und zwar für Dozenten und Studenten. Für derartige Tests ist nicht einmal ein allzu hoher Aufgabenkonstruktionsaufwand notwendig. Eventuelle Aufgabenfehler werden nämlich sofort relevant und können in der Aufgabenbesprechung sogar noch positiv zum Sichern des Lernens beitragen. Die Aufgaben brauchen auch nicht schriftlich verteilt zu werden, man kann sie projizieren.

ECKSTEIN (1972) berichtet über eine den Lehr-Lernprozeß steuernde Anlage in der ETH Zürich. Durch Betätigen einer entsprechenden auf seinem Platz im Hörsaal befindlichen Taste kann jeder Student direkt die Alternativen bei einer AAA wählen - ohne sich vor Dozent oder Kommilitone diskriminiert zu fühlen. Der Dozent kann aus der ihm ersichtlichen Summe aller gedrückten Alternativtasten sofort die Häufigkeiten zu den Alternativen entnehmen und entsprechend seine Vorlesung handhaben. Für die Studenten kann sich das befriedigende Gefühl ergeben, die Lehrveranstaltung mitzubestimmen. Die beschriebene Möglichkeit ist im Hinblick auf die Unmittelbarkeit der Rückmeldung ein spezifischer Vorteil von AAA.

Wie man an diesen Beispielen erkennt, können AAA außerordentlich unmittelbar zur Lehr-Lernsteuerung beitragen. Man kann das in ständigen kleinen Schritten, in studienbegleitenden Prüfungen vornehmen. Es erhebt sich die Frage, wo vielleicht ein Optimum zwischen ständig fordernden Kleinprüfungen und/oder punktuellen größeren Prüfungen liegt. Es scheint so, als ob die Leistungsmotivation durch studienbegleitende Prüfungen mit MAA durchaus angeregt wird; angelsächsische Erfahrungen sprechen dafür. Bei Überforderung kann andererseits auch schnell eine Testmüdigkeit und das Gefühl eines Leistungsdrucks bzw. Konkurrenz- und Wettbewerbsverhalten entstehen.

Die häufig diskutierte Frage, ob Tests anonym sein sollen oder können, hängt damit zusammen, ob man die lernsteuernde didaktische Funktion von der reinen Rekrutierungsfunktion der Qualifikation oder Plazierung trennt.

Eine Schwierigkeit an der Hochschule bildet der große Konstruktionsaufwand guter MAA; für eine Hochschule valide Tests brauchen wegen der akademischen Lehrfreiheit und größtenteils auch Prüfungsfreiheit noch nicht für andere Hochschulen zu gelten. Das Problem stellt sich sogar intrauniversitär von Dozent zu Dozent. Man muß also seine Items wirklich selbst schreiben, wozu aber die Erfahrung häufig fehlt

und aufgrund mangelndem Prestige der Lehre auch wenig erworben wird. Hochschuldidaktische Zentren mit entsprechend qualifizierten Mitarbeitern bilden einen Tropfen auf einem heißen Stein.

Weitgehendst ungenutzt sind die diagnostischen Möglichkeiten erprobter Tests. Obschon gerade beim Übergang von der Schule zur Hochschule, aber auch inter- und intrauniversitär erhebliche Probleme auftreten, sind erst Ansätze zu einer Forschung bzw. Anwendung erkennbar (INGENKAMP, 1975; SCHNEIDER, 1975).

Was schließlich die Herrschafts- und Sozialisationsfunktion von Prüfungen anbetrifft, so können Tests vielleicht dazu beitragen, die bei anderen Prüfungsformen stärker mögliche Willkür und Ungerechtigkeit abzubauen. Ihre objektive Auswertbarkeit kann Studenten weniger prüfungsängstlich machen und damit beitragen, die Prüfungsfunktion eines "Initiationsrituals" bzw. Mutprobe (FLECHSIG, 1974; PRAHL, 1976) in den Hintergrund treten zu lassen.

9 Literatur

Umfassendere Arbeiten mit weiteren Hinweisen sind mit einem Stern * gekennzeichnet

ALLINGER, U.: Computerunterstützte Prüfungen im Quadriga- Funkkolleg. in: DOHMEN u. PETERS (Hrsg.): Hochschulunterricht im Medienverbund 2. Weinheim 1971

AULT, L.H.: A comparison of multiple-choice and created-response test item formats. Diss. Abstr. Int., (1972), 33, (1-B), 416

BÅÅTH, J.: Improving Correspondence Instruction by Means of Electronics. Convergence (1972), 2

*BEARD, R.: Lehren und Lernen an der Hochschule. Düsseldorf 1972

BLACK, P.J.: University Examinations. Physics Education 3, (1968), 2

BLOOM, B.S.: Taxonomie von Lernzielen im kognitiven Bereich. Weinheim 1972

BRACHT, G.H. u. HOPKINS, K.D.: Objective and essay tests: do they measure different abilities. Zit. in INGENKAMP 1971

BYRNE, C.J.: Computerized Question Banking Systems. Brit. J. of Educat. Techn., (1976), 7, 44

CROMBAG, H.F.: Comparison between open question examinations and a choice test. Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden, (1970), 25, 349-359

*EBEL, R.L.: Essentials of educational measurement. New Jersey, Englewood Cliffs 1972

EBEL, R.L.: The Comparative Effectiveness of True-False and Multiple-Choice Achievement Test Items. ERIC ED 050 148 (1971)

ECKSTEIN, B.: Hochschuldidaktik und gesamtgesellschaftliche Konflikte. Frankfurt 1972

FLECHSIG, K.: Prüfungen und Evaluation. Hochschuldidaktische Stichworte. IZDH, Univ. Hamburg 1974

FLECHSIG, K. u. RITTER, U.P.: Konstanzer Werkstattseminar. Blickpunkt Hochschuldidaktik, AHD Harnburg, (1970), 8

FLÖRKEMEIER, V. u. KERSCHBAUM, T.: Meinungen zu Multiple-Choice-Fragen. Ergebnisse einer Umfrage bei Medizin- und Zahnmedizinstudenten. Deutsches Ärzteblatt, (1974), 71, 2581-2586

- FLÖRKEMEIER, V. u. KERSCHBAUM, T.: Objektivierete Leistungskontrollen in der medizinischen und zahnmedizinischen Ausbildung. Deutsches Ärzteblatt, (1974), 71, 2133-2139
- *FREIBICHLER, H.: Aufgabenarten bei objektivierten Lehr- und Testverfahren. Hannover 1976
- *FRICKE, R.: Kriteriumsorientierte Leistungsmessung. Stuttgart 1974
- FREUDENTHAL, H.: Lernzielfindung im Mathematikunterricht. Z. f. Päd., (1974), 20, 719-738
- *GAUDE u. TESCHNER: Objektivierete Leistungsmessung in der Schule. München 1970
- GRÄNICHER, H.: Alma hilft. Uni 75, Univ. Zürich, (1975), 8
- HAASE, H.: Psychologische Tests im Bildungswesen. Göttingen 1975
- HANKE, J. u. VOLLMER, H.W.: TASC-Manual. Freiburg 1975
- HOFFELNER; SEIBT; TAUBER: Demonstrationsmaterial zur Software CAVA. Paderborner Arbeitspapiere. Paderborn 1975
- *HOFFMANN, B.: The tyranny of testing. New York 1962
- HOPF, D.: Möglichkeiten und Grenzen der Anwendung von Tests. Grundlagentexte. Funkkolleg Pädagog. Psychologie Bd.2. Frankfurt 1973
- HUBBARD, J.P.: Erfolgsmessung in der medizin. Ausbildung. Bern 1974
- ILLINGWORTH, H.C.: The multiple choice or objective examination, a controlled trial. Zit. in BEARD 1972
- IMPP: Konstruktion und Auswertung von Multiple-Choice-Prüfungsaufgaben. Hinweise für Sachverständige. Institut für medizinische u. pharmazeutische Prüfungsfragen (IMPP) Mainz 1974
- IMPP: Praktische Hinweise zur Durchführung der schriftlichen Prüfungen nach der Approbationsordnung für Ärzte. IMPP Mainz 1975
- INGENKAMP, K.: Möglichkeiten und Grenzen der Testanwendung in der Schule. Weinheim 1968
- INGENKAMP, K.: Tests in der Schulpraxis. Weinheim 1971
- *INGENKAMP, K.: Pädagogische Diagnostik. Weinheim 1975
- INGENKAMP, K.; PAREY, E. u. TENT, L.: Schätzen und Messen in der Unterrichtsforschung. Weinheim 1973

- JONGSMA, E. u. POUND, R.: Testwiseness: The state of art. Paper presented at the Third International Symposium on Educational Testing. Leyden 1977
- KIRKWOOD, A.: CMA Instructions. Open University, Institute of Educational Technology, Milton Keynes, Großbritannien (unveröffentl.)
- KLAUER, K.J.: Lehrzielorientierte Tests. Düsseldorf 1974
- *KLAUER, K.J.: Handbuch der pädagogischen Diagnostik. Düsseldorf 1978
- KÜFFNER, H.: Einsatz von EDV bei der Datenorganisation in: HELLER, K. (Hrsg.): Handbuch der Bildungsberatung. Stuttgart 1976
- KÜFFNER, H.: LOTSE, Informationsheft für Fernstudenten. Maschinelles Korrekturdienst für Einsendeaufgaben. Fernuniversität Hagen 1977
- KÜFFNER, H.: Optisch-lesbare Markierungsbelege: Erfahrungen mit der ersten Version des LOTSE-Antwortbogens, LOTSE-Erfahrungsbericht Nr. 3. Fernuniversität - Gesamthochschule -, Zentrum für Fernstudienentwicklung. Hagen 1977 (unveröffentl.)
- KÜFFNER, H.: Lehrzielorientierte Tests im Fernstudium: Probleme der Einrichtung von maschinellen Korrekturdiensten in: TAUBER, M.J. (Hrsg.): Der Computer als didaktisches und organisatorisches Hilfsmittel des Fernstudiums. Hannover und Paderborn, 1979
- LIENERT, G.A.: Testaufbau und Testanalyse. Weinheim 1967
- MÅNSSON, N.O.: Datorn i Korrespondensundervisningen. Forsknings-och utvecklings nytt, Hermods Skola, Malmö, (1974), 3
- MASSEY, A.J. u. NEWBOULD, C.A.: Computerised Item Banking System, Handbook. Occasional publication, Cambridge, (1976), 3
- MELLENBERGH, G.J.: A comparison between different kinds of achievement tests. Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden, (1972), 27, 157-158
- *MARTON, F.: Prüfung und Evaluierung in der akademischen Ausbildung. Göteborg 1967
- MARSTON, A. u. MARSTON, M.: The effect of student participation, in the construction of a Multiple-choice achievement examination. The Journ. of Educat. Research, (1965), 59
- MÖLLERS, P.: Programmsystem "Integriertes Rechnungswesen" an der Ruhr-Universität Bochum, Abteilung für Wirtschafts- wissenschaft. Bochum (interne Mitteilung. o.J.)
- MUELLER, D.J.: An Assessment of the Effectiveness of Complex Alternatives in Multiple-Choice Achievement Test Items. Educat. and Psychol. Measurement, (1975), 35, 135-141

- OSBORN, J.W.: The Use of Multiple-choice-questions in University Examinations. University of London, Bulletin, (1976), 37
- PAULI, H.G.: Neue Verfahren in der Medizinerbildung. Schweizerische Ärztezeitung. (1975), 7, 244-247
- PHILIPPS, L.: Testaufgaben in der Rechtswissenschaft. Heidelberg 1978
- POPHAM: Zit. in ELBERS, D.: Curriculumentwicklung in den USA. betrifft erziehung, (1971), 4, 26
- PRAHL, H.W.: Hochschulprüfungen - Sinn oder Unsinn?, München 1976
- *RÜTTER, T.: Formen der Testaufgabe. München 1973
- SCHNEIDER, W.: Diagnostische Prüfsysteme im Hochschulbereich. IBE-Bulletin, (1975), 20, 3-16
- SCHÜMER, R.: Vorläufiger Bericht zu den Begleituntersuchungen des CMA-Projektes. ZIFF, Fernuniversität Hagen 1977
- SCHÜTZ, M. et al.: Prüfungen als hochschuldidaktisches Problem. Blickpunkt Hochschuldidaktik, AHD Hamburg, (1969), 1
- SIMONS, V.: Handschriftliche Korrespondenz mit dem Computer. Versuche mit dem Handschriften-Ziffern-Belegleser. Neue Unterrichtspraxis, (1969), 2, 595-598 u. (1969), 3, 157-159
- SKINNER, B.F.: Teaching Machines. Scientific American (1961), 90-102
- SLOTNICK, H.B.: Toward a theory of computer essay grading. Journ. of Educat. Measurement, (1972), 9, 253-263
- STALLER, R. et al.: Automatische Unterrichtsauswertung. In: LANGENDER, W. (Hrsg.): SIZSOZ Expertisen, Ausgewählte Gebiete sozialwissenschaftlicher DV-Anwendung Bd. 1. Bonn 1976
- STUMMER, P.O.: Untersuchungen zur Hochschuldidaktik: Der literarische Test. Die neueren Sprachen, (1974), 3, 206-216
- VOIGTMANN, K., NEUMANN, D. u. SCHAPPEL, S.: Computerunterstützter Verfahrensablauf der bundeseinheitlichen schriftlichen Prüfungen nach der Approbationsordnung für Ärzte. Medizinstudent, (1976), 11, 160-165
- VORKAUF, H.: Neue Verfahren in der Medizinerbildung: Statistische Analyse von Prüfungsergebnissen. Schweizerische Ärztezeitung, (1975), 9
- WRIGHT, P., ALDRICH, A. u. WILCOX, P.: Same Effects of Coding Answers for Optical Mark Reading on the Accuracy of Answering Multiple-Choice Questions. Human factors, (1977), 19, 83-87
- ZENZ: Zum Punktesammeln ins Bürgerhaus. Analysen, (1975), 5, 28-29

Anhang 1

A Aufgabensammlungen mit Multiple-Choice-Fragen

AVERY, J.H.; INGRAM, A.W.K.: Objective Tests in Advanced Level Physics. London 1977

COOK, J.B.: Multiple Choice Questions in O-Level Physics.
Wheaton 1978

DEMPEWOLF, E.: Multiple-Choice-Fragen Chemie. Herten 1978

ECKERT, K.: 500 Multiple-Choice-Fragen Physik. Herten 1978

HAWKINS, H.J.; POTTER, M.: Multiple Choice Physics. London 1978

HEINICKE, A.: Examens-Fragen Biomathematik. Berlin 1975

HÖHL, M.; NÄGERL, H.: Examens-Fragen Physik für Mediziner. Berlin 1975

HOUSTON, J.G.; WILSON, J.M.: Multiple Choice Questions for Assessment in Physics. London 1978

INDUSTRIE- UND HANDELSKAMMER FÜR DIE PFALZ: Programmierte Prüfungsaufgaben - Betriebswirtschaftslehre, Ludwigshafen 1976

INDUSTRIE- UND HANDELSKAMMER FÜR DIE PFALZ: Prüfungsaufgaben Wirtschafts- und Sozialkunde. Ludwigshafen 1975

JOHNSON, K.: Multiple Choice Physics for C.S.E and O-Level. London 1978

KALISCH, W.E.: Multiple-Choice-Fragen Biologie. Herten 1978

LATSCHA, H.P.; KLEIN, H.A.: Examensfragen Chemie für Mediziner. Berlin 1977

MARSHALL, J. u.a.: Multiple Choice Questions for 'A' Level Physics. London 1978

MELCHINGER, H.: Examens-Fragen Allgemeine Psychologie. Berlin, Heidelberg, New York 1978

NEUBRAND; SCHAAF: Programmierte Prüfungsaufgaben-Einkommensteuer, Lohnsteuer, Körperschaftsteuer. Ludwigshafen 1976

RAHMISCH, S.: Multiple-Choice-Fragen Psychologie und Soziologie. Herten 1978

SCHUSTER, D.: Bankkaufmann, Übungsaufgaben Lösungshilfen Musterlösungen Übungssatz Elektronische Auswertung. Bad Homburg v.d.H. (o.J.)

SCHWERD, W.; WAGNER, H.J.: Examens-Fragen Rechtsmedizin. Berlin 1976

SOUKUP, B.: Speditionskaufmann, Übungsaufgaben Lösungshilfen Musterlösungen Übungssatz Elektronische Auswertung. Bad Homburg v.d.H. (o.J.)

B Beispiele für Multiple-Choice-Fragen finden sich in:

BASTINE, R. u.a. (Hrsg.): Funkkolleg Beratung und Erziehung, Bd. 1 u. 2. Stuttgart 1977

BAUMGÄRTNER, K.; STEGER, H.: Sprache Bd. 1 u. 2, Eine Einführung in die moderne Linguistik. Stuttgart 1976

HANF, T. u.a. (Hrsg.): Funkkolleg Sozialer Wandel, Bd. 1 u. 2. Stuttgart 1975/76

HÄUSER, K.: Funkkolleg Volkswirtschaftslehre. Stuttgart 1975.

IMPP: Aufgaben-Entwicklung-Analysen, Institut für medizinische u. pharmazeutische Prüfungsfragen. Mainz (o.J.)

JURISTISCHE SAMMLUNG (JuS). Zeitschrift. München (seit 1961)

KADELBACH, G. (Hrsg.): Wissenschaft und Gesellschaft, Funkkolleg Bd. 1. Stuttgart 1974

KLAFKI, W. u.a. (Hrsg.): Funkkolleg Erziehungswissenschaft, Bd. 7, 8 u. 9. Stuttgart 1975/76

PHILIPPS, L.: Testaufgaben in der Rechtswissenschaft. Heidelberg u. München 1978

RUEGG, W. (Hrsg.): Funkkolleg Soziologie. Stuttgart 1975

TODT (Hrsg.): Funkkolleg Biologie, Bd. 1 u. 2. Stuttgart 1976

WIETHÖLTER, R. (Hg.): Funkkolleg Rechtswissenschaft. Stuttgart 1976

Anhang 2 Maschinelle Korrekturdienste im Hochschul- oder Fernstudienbereich (entnommen aus Küffner, 1979)

Systembezeichnung		Institution	Bereich	Nähere Angaben
CAVA	Computerunterstützte Analyse und Vergabe von Aufgaben	Forschungs- und Entwicklungszentrum für objektivierte Lehr- und Lernverfahren, Paderborn	Hausaufgaben im Fernstudium Mathematik	TAUBER (1973) HOFFELNER; SEIBT; u. TAUBER (1975)
CIBS	Computerised Item Banking System	Test Development and Research Unit, University of Cambridge (Großbritannien)	Prüfungen zum Hochschuleingang (versch. Fächer)	MASSEY u. NEWBOULD (1976)
CMA	Computer Marked Assignment	Open University, Milton Keynes (Großbritannien)	Hausaufgaben im Fernstudium (versch. Studiengänge)	KIRKWOOD (o.J.)
CADE	Computer Assisted Distance Education	Hermods Skola, Malmö (Schweden)	Hausaufgaben im Fernstudium (versch. Fächer)	MANSSON (1974) BAATH (1972)
IEA	Institut für Ausbildungs- und Examensforschung	Medizin. Fakultät der Universität Bern (Schweiz)	Zwischen- und Abschlußprüfungen im Medizinstudium	VORKAUF (1975)
IGR	Integriertes Rechnungswesen	Ruhr-Universität Bochum, Abteilung für Wirtschaftswissenschaft	Klausuren und Übungsaufgaben zu den Fächern Rechnungswesen, Kostenrechnung u.a.	MÖLLERS (1977)
IMPP		Institut für medizinische und pharmazeutische Prüfungsfragen, Mainz	Vor- und Abschlußexamen der bundeseinheitlichen schriftl. Examen für Mediziner und Pharmazeuten	VOIGTMANN; NEUMANN und SCHAPPEL (1976)
LOTSE	Lehrzielorientierte Testaufgaben zur standardisierten Erfolgsbeurteilung	Fernuniversität (Gesamthochschule), Hagen	Hausaufgaben im Fernstudium Studiengänge: Mathematik, Wirtschaftswissenschaften Erziehungswissenschaften	KÜFFNER (1977)
TASC	Testauswertendes statistikerstellendes Computerprogramm	Modellversuch CUU Mathematik Universität Freiburg	Klausuren an der Universität Freiburg	HANKE u. VOLLMER (1975)
QUADRIGA-FUNKKOLLEG		Deutsches Institut für Fernstudienforschung, Tübingen	Hausaufgaben und Zertifikatsprüfungen im Fernstudium (z.B. Erz.wiss., Mathematik, Biologie, Linguistik, Päd. Psychologie)	ALLINGER (1971)

Anhang 3: Glossar Test- und Aufgabengütekriterien

Zur knappen Beschreibung der Qualität von Tests bzw. Aufgaben sind eine Reihe von Indizes entwickelt worden, von denen die am häufigsten gebrauchten im folgenden angeführt sind. Der Zusammenhang der Indizes untereinander und deren Problematik im einzelnen findet sich ausführlich in LIENERT (1967).

Objektivität

Als Objektivität bezeichnet man den Grad der Unabhängigkeit der Testergebnisse vom Untersucher bzw. Auswerter. Im einzelnen werden Durchführungsobjektivität, Auswertungsobjektivität und Interpretationsobjektivität unterschieden.

Validität

Unter Validität versteht man den Grad der Genauigkeit, mit dem ein Meßverfahren das mißt, was es zu messen vorgibt (LIENERT 1967). Man unterscheidet zwischen Inhaltsvalidität, kriteriumsbezogener Validität und Konstruktvalidität.

Reliabilität

Unter Reliabilität eines Tests versteht man den Grad der Genauigkeit, mit dem er ein bestimmtes Persönlichkeits- oder Verhaltensmerkmal mißt, gleichgültig, ob er dieses Merkmal auch zu messen beansprucht (LIENERT 1967).

Schwierigkeitsindex

Kennzeichnet das Verhältnis der Prüflinge mit richtiger Lösung einer Aufgabe zur Zahl aller Prüflinge eines Tests. Je größer er ist, um so leichter ist die Aufgabe. Wird deswegen auch manchmal anders definiert.

Trennschärfeindex

Der Trennschärfeindex ist ein Maß für den Zusammenhang zwischen der Lösung einer Aufgabe und den Lösungen aller Aufgaben im Test. Die Trennschärfe einer Aufgabe ist hoch, wenn sie häufig von den Prüflingen gelöst wird, die auch viele andere Aufgaben richtig gelöst haben. Der Trennschärfeindex wird üblicherweise als Korrelationsmaß angegeben und kann zwischen -1 und +1 schwanken.

ÜBER DEN AUTOR:

Christian UCKE (Jahrgang 1942)

1961-1968	Studium der Physik in Göttingen und München (Dipl.-Phys.)
1969-1973	Promotion in Physik (Dr.rer.nat.) interdisziplinäre Arbeit im Bereich der medizinischen Optik über "Streulicht im Auge"
1971-1976	Studium der Pädagogik und Psychologie an der Universität München (M.A. in Pädagogik)
seit 1974	Leiter des physikalischen Praktikums für Mediziner an der Technischen Universität München
Arbeitsschwerpunkte	Didaktik naturwissenschaftlichen Unterrichts insbesondere Praktika an der Hochschule Entwicklung von Versuchsanordnungen, Lehrziel-evaluierung, Einsatz von audiovisuellen Medien, spezielle optische Refraktionsverfahren für Augenärzte (Buch "Skiaskopie" Enke-Verlag, Stuttgart 1978) in diesen Gebieten auch Veröffentlichungen
Anschrift	Physik-Department E20 Technische Universität München James-Franck-Straße 8046 Garching bei München