

**Tabelle 1: Vermeintliche Herzinfarktursachen (100 Männer)**

Ursache	Anzahl der Nennungen
1. Arbeitsüberlastung, Hetze	33
2. Streß	30
3. Ärger	14
4. Rauchen	13
5. keine Erklärung	12
6. falsche Ernährung	10
7. Aufregung	9
8. körperliche Überlastung (akute)	6
9. Angst, Kummer, Sorgen	6
10. viele körperliche Anstrengungen i. d. Vergangenheit	5
11. Veranlagung	3
12. Bewegungsmangel	3
13. Infekte	3
14. Operation, Unfall	3

100 Männer gaben für 115 durchgemachte Herzinfarkte 157 Gründe an. Je einmal wurden u. a. angegeben Hochdruck, Fett im Blut (aus Tobiasch, V. „Die Vermeidung d. Risikofaktoren“ Bindernagel Verlag, Friedberg/Hessen 1977)

**Tabelle 2: Kenntnisse von „Rheuma“-Patienten zu den Arteriosklerose-Risikofaktoren**

Risikofaktor	Anzahl der Nennungen
1. Rauchen	71
2. Übergewicht	63
3. Streß	61
4. Alkohol	51
5. Bewegungsmangel	39
6. Essen (falsches, zu fett, zu gut, unregelmäßig)	32
7. Blutdruck (zu hoch)	14
8. Blutfettwerte (zu hohe)	10
9. Blutzucker (zu hoher)	4
10. Schlaf (zu wenig)	2

100 Männer gaben 354 Gründe an

**Tabelle 3: Kenntnisse von „Rheuma“-Patienten zu den Arteriosklerose-Risikofaktoren**

Risikofaktor	Anzahl der Nennungen
1. Streß	81
2. Rauchen	74
3. Übergewicht	67
4. Alkohol	48
5. Bewegungsmangel	40
6. falsches Essen	40
7. Hochdruck	25
8. Blutfette (zu hohe)	22
9. Sauerstoffmangel (schlechte Luft, smog)	5
10. Vererbung	4
11. Blutzucker (zu hoher)	2
12. Freizeitgestaltung (falsche)	2

100 Frauen gaben 420 Gründe an

seinen Herzinfarkt darauf zurück, daß er drei Monate vor dem Ereignis das Rauchen aufgegeben hatte (Rauchen zur Streßbekämpfung). Aus dieser Einstellung unserer Patienten zu den Risikofaktoren (zum

Beispiel Übergewicht durch schwere Knochen, Rauchen zur Beruhigung) erklärt sich zwanglos unsere schlechte Statistik. Die Streßthese führt zu weiterem Fehlverhalten. Die Streßanhänger sollten u. a. beden-

**Risikofaktoren**

ken, daß der Herzinfarkt im und nach dem Kriege in Deutschland viel seltener war als heute, obgleich das deutsche Volk damals unvergleichlich härteren Streß-Bedingungen ausgesetzt war.

Erneute Befragungen, die wir in den ersten Monaten 1979 durchgeführt haben, weisen auf eine im Gang befindliche (?) Korrektur der Fehlmeinungen der Betroffenen hin (Tabellen 2 und 3). Der Streß steht – wenn man die Ansichten der Männer und Frauen zusammennimmt – nicht mehr an erster Stelle. Zwar handelt es sich bei den Befragten der Tabelle 1 bzw. 2 und 3 um unterschiedliche Patienten-Gruppen (Herzinfarkt – Rheuma), doch glauben wir, daß die aus den neueren Tabellen hervorgehende Sachkenntnis gerade der nicht unmittelbar vom Herzinfarkt bedrohten Bürger bemerkenswert ist. Die auf einem Fehlverhalten beruhenden Risikofaktoren rücken nach oben. Diese Tabellen zeigen, daß die Bemühungen der Präventiv-Mediziner (etwa M. Kunze/Wien und F. Schmidt/Mannheim) und der Medien in den letzten Jahren offenbar nicht erfolglos waren.

Da wir der Ansicht sind (2), daß eine Verbesserung unserer Morbiditäts- und Mortalitäts-Statistik nur über eine Korrektur der Meinung unserer Patienten und vor allem auch der Noch-nicht-Patienten zu erreichen sein wird, sehen wir nach diesem Befragungsergebnis, verglichen mit dem in der Tabelle 1 niedergelegten, etwas optimistischer in die Zukunft.

**Anmerkungen**

(1) bis 1977, die Statistik 1978 steht uns noch nicht zur Verfügung – (2) siehe die Publikationen: Tobiasch, V.: „Übergewicht – was tun“, Hippokrates Verlag, Stuttgart, 1974; ders.: „Die Vermeidung der Risikofaktoren“, Verlag Bindernagel Friedberg/Hessen, 1977; ders.: „Die Meinung des Patienten“, Med. Welt 28 (N. F.), 2014–2018, 1977; Tobiasch, V., und H. Prokop: „Streß – biologisches Erbe, Segen oder Fluch?“, Verlag S. Tomek, Monte Carlo, 1979

Anschrift des Verfassers:  
Professor Dr. med. Viktor Tobiasch  
Institut zur Erforschung  
präventiv-medizinischer  
Fragen e. V.  
7972 Neutrauchburg

# Analyse des Multiple-choice-Verfahrens

Prüfung als Denkresultatkontrolle  
oder als denkverlaufsspezifische Rückmeldung

Theo Wehner und Peter Kruse

Wir verzichten auf eine Ausformulierung von Vor- und Nachteilen einer m-c-Prüfung und verweisen auf Flörkemeier und Kerschbaum (1). Sie nannten 12 negative Implikationen, die jedoch allesamt als Vernachlässigung der Subjektseite in der Prüfung angesehen werden müssen, während die positiven Aussagen unter die Kategorien Objektivität und Ökonomie subsumierbar sind und sich vorrangig als Positiva für die Prüfinstitution darstellen. In der Tabelle haben wir eine Auflistung der m-c-Test-Implikationen vorgenommen.

Auf dem Weg, Prüfungen in der Sprache der Prüflinge zur gerechteren Sache, in der Sprache der Wissenschaft zu einem objektiven Meßinstrument zu machen, wurde die mündliche Prüfung aufgrund der sie beeinflussenden Faktoren (Störvariablen) durch ein standardisierteres Prüfverfahren ersetzt. Sowohl Form als auch die Intention der vom Institut für medizinische und pharmazeutische Prüfungsfragen (IMPP) durchgeführten Prüfungen gestatten ihre Klassifikation als Test und damit die Anwendung der drei Hauptgütekriterien, die an einen Test angelegt werden sollten: Objektivität, Reliabilität, Validität.

Anhand dieser Kriterien muß die Frage gestellt werden, ob die Ablösung der mündlichen Prüfung durch den m-c-Test alte Mängel beseitigt sowie neue Qualitäten hinzugewinnt. Wir nehmen an dieser Stelle nur Bezug auf das Kriterium der Objektivität.

Bezüglich der Interpretation, der Durchführung und der Auswertung kamen die Prüfungsergebnisse bei mündlichen Prüfungen nicht unabhängig vom Prüfer und dem Interaktionsprozeß zwischen Prüfer und Prüfling zustande. An Einflußfaktoren wie Geschlechtsspezifik, Erwartungen und Lehrmeinungen des Prüfers entzündete sich vorwiegend die Kritik. Diese wurde von den „Überwindern“ aufgegriffen und beseitigt:

Um die Durchführungsobjektivität zu gewährleisten, werden jedem Prüfling die gleichen m-c-Fragen vorgelegt. Die Bearbeitungszeit wird konstant gehalten, und die zentrale Auswertung in Mainz sichert die Auswertungsobjektivität. Die klassische Testtheorie fordert Objektivität für die Durchführung und die Auswertung. Diese ist unter Einsatz von Technik und der Umsetzung des Postulats der Normierung und Standardisierung als erfüllbar anzusehen.

Wir möchten als Erweiterung des Kriteriums den Begriff der *Konstruktionsobjektivität* einführen, das heißt, daß die Fragen und Antwortvorgaben unabhängig vom Testkonstrukteur zustande kommen sollen, daß sie wissenschaftlich begründbar sein und inhaltlichen Forderungen genügen müssen. Eine gebundene Aufgabe, das m-c-Item zählt zu diesem Typ, besteht aus einem sogenannten Stamm, dem eigentlichen Aufgabentext und z. B. fünf Antwortalternativen. Die falschen Antwortvorgaben, „die von der Lösung ab-

Die Verfasser – beide am Psychologischen Institut der Universität Münster tätig – haben Multiple-choice (im folgenden: m-c)-Prüfverfahren mehrfach analysiert und zum Beispiel die durch den m-c-Test begünstigte Trennung von Studium und Beruf oder die „neue Einheit“ von Prüfungsvorbereitungszeit und Prüfung sowie die Fragwürdigkeit des Gütekriteriums der Reliabilität untersucht. In diesem Beitrag soll der Frage nach den Konstruktionskriterien von Antwortalternativen nachgegangen werden.

lenken sollen“, nennt man Distraktoren. Der Prüfling sollte die Antwortalternativen – nach Lienert (2) für „richtig, zutreffend oder annehmbar“ halten. Die Distraktoren, wir folgen im weiteren wieder Lienert, müssen neben einer gewissen Anziehungskraft plausibel, für den Nichtwissenden gleichwahrscheinlich sein. Michaelis vom IMPP sieht in der Findung von Distraktoren gar eine „Kunst und Schwierigkeit“ S. 64 (3).

Soviel Mehrdeutigkeit zeigt mit Eindeutigkeit den Mangel an Konstruktionskriterien, nicht nur in Lehrbüchern, sondern auch in der Praxis, wie eine Vielzahl von empirischen Untersuchungen gezeigt hat. Es ging in diesen Untersuchungen, einer der Autoren führte selbst eine solche durch, z. B. um die Frage der Suggestivwirkung von Alternativen, sowie um Analysen, inwieweit Plausibilitäts- und Wahrscheinlichkeitsüberlegungen bezüglich der angebotenen Distraktoren eher den Lösungsprozeß determinieren als die logische Verknüpfung von gespeichertem Wissen.

Nach unserer Meinung ist ein Teil der empirisch aufzeigbaren Defizite darauf zurückzuführen, daß m-c-Test-Konstrukteure bis dato die Antwortalternativen gemäß ihrem eigenen Einfallsreichtum sowie ohne

Multiple choice

Nennung zugrundeliegender Kriterien, am Schreibtisch generieren. Um die Bestimmung der Antwortalternativen von persönlichen Vorlieben weg und damit auf eine intersubjektive Nachprüfbarkeit hinzuwirken, fordern wir eine empirische

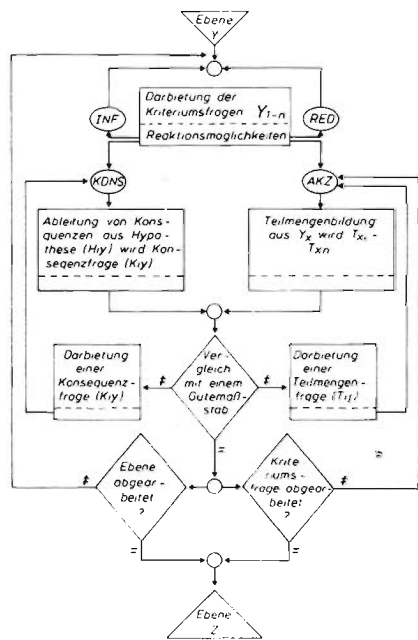
oder inhaltliche Ableittechnik, das heißt: Aus der Auftretenswahrscheinlichkeit unkorrekter Antworten, gemessen an einer repräsentativen Stichprobe werden die Distraktoren begründet und abgeleitet. Mit dieser Methode ließen sich in einer

Untersuchung von Wehner (4) zu verlässige Ergebnisse erzielen.

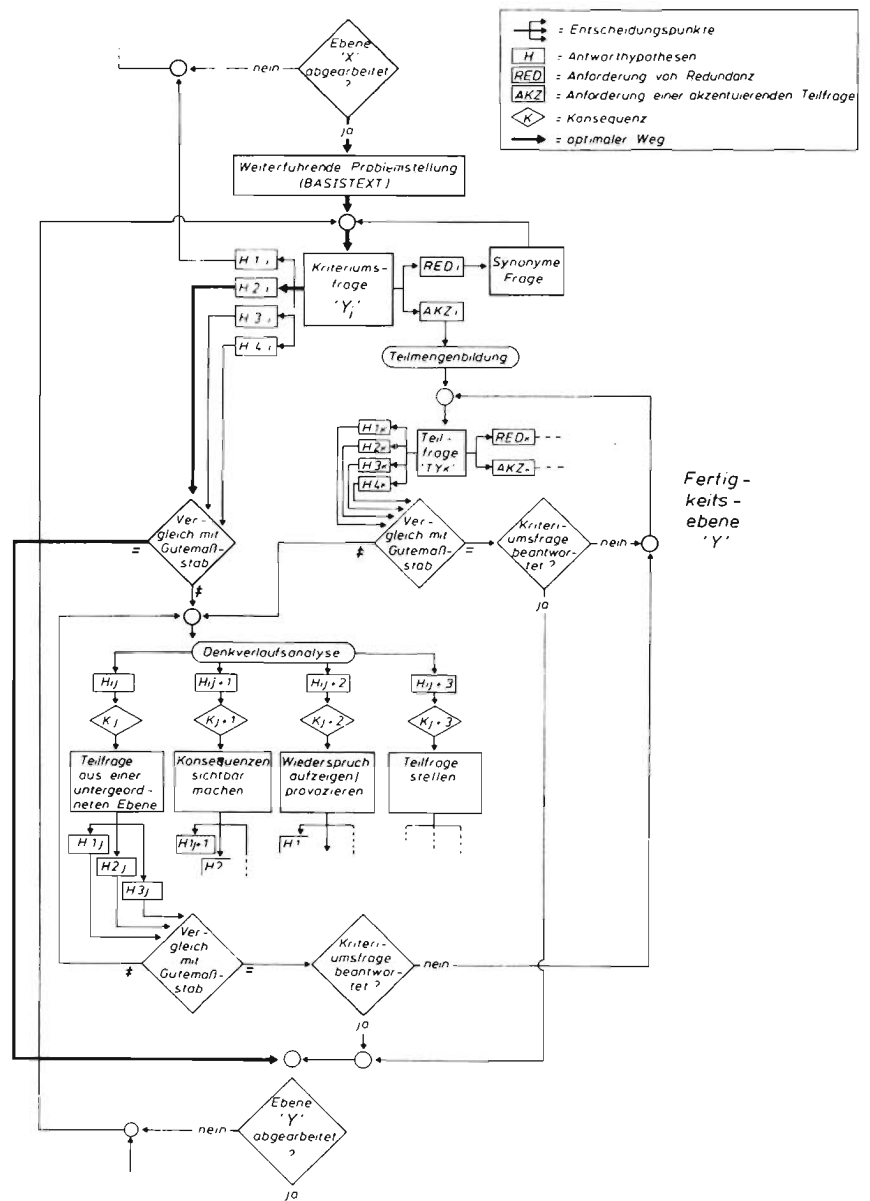
Eine Frage aus dem Bereich Geschichte der Medizin, welche sowohl im August 1974 als auch im März 1975 vom IMPP gestellt wurde,

**Abbildung: Prüfung als Denkverlaufs-analyse, Erfassung der Lernfähigkeit**

In der Abbildung wird eine Fertigkeitsebene (Diagnose) visualisiert. Sie kann nur angesteuert werden, wenn vorherige (Anamnese) Problemebenen abgearbeitet worden sind. Bei Visualisierung der Frage  $Y_1$  kann der Prüfling sowohl mit der Wahl der Hypothesen  $H_1$ - $H_n$  antworten als auch eine redundante Formulierung bzw. Merkmalsakzentuierungen vom System verlangen. Der Wahl einer Hypothese schließt sich der Vergleich mit einem Gütemaß an, der eine Analyse



AKZ = Der Prüfling fordert eine Akzentuierung von Teilspekten der gestellten Aufgabe  
 INF = Der Prüfling fordert Zusatzinformationen bz B. Untersuchungsergebnisse  
 RED = Der Prüfling fordert eine redundanzere (synonyme) Fragenformulierung  
 KDNS = Das System generiert eine der vom Prüfling gewählten Hypothese adäquate (konsequente) Folgefrage



der zugrundeliegenden Denkopoperationen zur Folge hat. Zur Frage  $Y_1$  und der Wahl  $H_3$  wird z. B. eine Widerspruch provozierende Frage  $K_3$ , als eine den Denkverlauf des Prüflings weiterführende Frage, zur Beantwortung vorgelegt. Die Zahl zu beantwortender Fragen erhöht sich in diesem Fall aufgrund individueller Handlungsstrukturen und Problemlösestrategien. Unter den Hypothesenvorgaben sind keine Distraktoren, sondern sachlogische, suchraumcharakterisierende Antwortvorgaben zu verstehen, die verschiedene Denkopoperationen aktualisieren. Bereits die Formulierung einer - kontextabhängigen, aus der gegebenen Antwort abgeleiteten - Folgefrage stellt Feedback für den in der vorangegangenen Frage induzierten Denkprozeß dar. Es handelt sich nicht um eine „Ja-Nein“-Bewertung, sondern um eine Bewertung, die Rückschluß auf die Gerichtetheit und den Verlauf des Problemlöseprozesses erlaubt

**Tabelle: Methodisch-inhaltliches Gerüst eines Multiple-choice-Prüfverfahrens**

Vorteile und Implikationen aufgrund von Objektivität und Ökonomie:	Erfassung des Basis- und Kernwissens in verbindlichem Gegenstandskatalog. (GK)	Nachteile und Konsequenzen aufgrund der Eliminierung der Subjektseite im Prüfprozeß:	Einschränkung der Freiheit der Lehre Überregionaler Konkurrenzkampf Steigerung der Anonymität Eliminierung von Spontaneität
	Trennung von Studium und Prüfung als Implikation des GK.		Notwendigkeit isolierter Examensvorbereitung Rückgang von Vorlesungsbesuchen (v. 80 a. 20%) Mangelndes Feedback für die Lehrenden
	Bedingungskonstanz: der Aufgabenstellung		Ausbildung praxisferner Beantwortungsstrategien Erfassung von latenter – momentan nicht verbalisierbarem – Wissen Favorisierung der Wiedererkennungsmethode als Reproduktionsform
	der Durchführung und Auswertung		Ausschluß von Eingriffsmöglichkeiten in den Prüfprozeß Ausschluß von Lernen in der Situation Favorisierung von resultativem vs. prozessuellem Feedback Fehlentscheidungskorrektur in Eigenverantwortung Motivationsminderung für Weiterlernen Aufhebung der Denkverlaufstransparenz Erhöhung des Streßfaktors bei Unsicherheiten Alleinige Bewertung des Denkresultats
	Bewertungstransparenz (Unabhängigkeit der Items)		Zerstückelung von Kontexten Verlust kognitiver Einbettung von Fakten Favorisierung M-c-Test prüfbarer Inhalte Gleichgewichtung der Antworten Begünstigung praxisfernen Lernverhaltens (med. Psychologie statt Anatomie!) Ausbildung inadäquater Entscheidungsstrategien (Ratemöglichkeit größer 20%!) Induktion von Fehlentscheidungen (Vorgabe schlitzohriger Distraktoren!)
Prüfung der kognitiven Fertigkeitsebene	Ausschließliche Prüfung finiter Fakten (93% Gedächtnisfragen!) Nichtbewertung psychomotorischer und affektiver Komponenten im Lernprozeß		

soll das Fehlen von Konstruktionskriterien oder die mangelhafte Beherrschung der „Kunst“ der Distraktorenableitung verdeutlichen. Die Prozentangaben beziehen sich auf die Beantwortungsfrequenz vom August 1974.

Welche Aussage trifft zu?

Unter Hömöopathie versteht man:

- a) die Behandlung durch Kurpfuscher 1 Prozent,
- b) die Behandlung mit Geheimmitteln 1 Prozent,
- c) die mediko-mechanische Behandlung 3 Prozent,
- d) die Therapie mit natürlichen Heilmitteln, insbesondere kaltem Wasser 31 Prozent,

e) die Behandlung mit kleinen Mengen von Stoffen, deren Verabreichung in großer Menge beim Gesunden die Symptome der Krankheit hervorruft 64 Prozent.

Es sei vorab erwähnt, daß die Alternative „E“ vom IMPP und der zugehörigen Sachverständigenkommission als Treffer akzeptiert wird. Es sei ferner erwähnt, daß ein Item dann zweimal verwendet wird, wenn: „Einzelaufgaben . . . in ihren formalen Kriterien den gestellten Anforderungen besonders gut entsprechen . . . und damit die Zuverlässigkeit der Examen . . . erhöhen.

Darüber hinaus können mit Hilfe wiederholt gestellter Aufgaben wichtige Informationen für den Vergleich verschiedener Examens-

populationen gewonnen werden“ Michaelis S. 83 (3).

Obwohl zwischen der zweimaligen Darbietung eines Items, laut Angaben des IMPP, eine sorgfältige Itemanalyse liegt, ist niemandem aufgefallen, daß allein schon die Länge des Distraktors „E“ einen gewissen Grad an „Unsicherheit“ nimmt, daß also ein grober Konstruktionsfehler unterlaufen ist. Wir sind darüber hinaus der Meinung, daß diese Vorgabe von Alternativen (sie ist kein Einzelfall oder gar Extrem) nichts über das medizinische Können und Wissen desjenigen aussagt, der irgendeine der Vorgaben wählte.

Eine solche Distraktorenvorgabe mag den Prüfungsstreß abbauen oder zur Aufheiterung der Prüflinge

## Multiple choice

beitragen, ein Beitrag zu qualifiziertem Diagnostizieren ist allerdings nicht geleistet. Eine solche Vorgabe erhöht zudem die Rateeffizienz erheblich: Mit dem oft zitierten „gesunden Menschenverstand“ lassen sich zumindest die Distraktoren A, B, C ausklammern. Die Ratewahrscheinlichkeit für die verbleibenden Alternativen erhöht sich – bei totaler Unwissenheit – von ursprünglich 20 auf 50 Prozent.

Unsere Forderung nach einer empirischen oder inhaltsbezogenen Ableitetechnik zielt zwar auch auf eine Verbesserung von m-c-Items, fordert jedoch darüber hinaus eine Modifikation des dem m-c-Verfahren zugrundeliegenden Denkmodells.

Es ist unter Beachtung pädagogisch-psychologischer Erkenntnisse unverantwortlich, durch die Alternativenvorgabe Hypothesen anzubieten, die den Problemlösungsraum beliebig verzerren (wer käme schon auf die Idee, daß homöopathische Anwendungen eine Behandlung mit Geheimmitteln darstellen könnte?).

Es werden ferner sowohl motivationale als auch affektive Bedürfnisse des Prüflings ignoriert, wenn nach der Wahl einer angebotenen Hypothese keinerlei Rückmeldung, erfolgt. Guthke (5) konnte zeigen, daß durch bloßes „Richtig-Falsch“-Feedback die Prüfungsergebnisse im Vergleich mit einer Parallelgruppe signifikant besser ausfielen.

Rückmeldung wird von uns nicht einfach als positive oder negative Sanktion aufgefaßt, sondern als notwendige Bedingung für die Verbesserung oder Aufrechterhaltung der Anpassung von individuellem Verhalten an objektive Forderungen oder Besonderheiten der durchzuführenden Handlungen. Durch extreme Anonymität und unqualifiziertes Feedback bei einer m-c-Testbearbeitung wird eine die Tätigkeit (das Lösen einer Aufgabe verstehen wir als Tätigkeit) abschließende bzw. sie weiterführende Handlungsbeurteilung und Handlungsplanung ausgeschlossen. Dies bedeutet eine unnötige Belastung im Sinne der

Speicherung „unerledigter Aufgaben“ (Zeigarnik-Effekt).

Der Prüfling, der z. B. der Meinung war, Chromosomen bestehen im wesentlichen aus RNA und Protein, kann dies nur dann korrigieren, wenn ihm ein ungutes Gefühl bleibt und er nach der Testbearbeitung in einem Lehrbuch nachschlägt. Vom IMPP wird keine Korrektur dieser Hypothesenwahl eingeleitet; es wird lediglich mit einem Punktabzug reagiert, was absolut nicht zur Löschung der zugrundeliegenden Operatoren oder Denkprozesse beiträgt. Falls sich bei der Bearbeitung von 300 m-c-Fragen öfters ein solch ungutes Gefühl einstellt, dann gerät die Beantwortung von nachfolgenden Fragen in eine unkontrollierbare und wohl auch nicht intendierte Abhängigkeit.

In einem solchen Fall werden nicht mehr nur kognitive Fertigkeiten gemessen, es werden vielmehr psychische Belastbarkeit, Frustrations- oder gar Aggressionstoleranz überprüft; es wird zusätzlicher und unnötiger Streß erzeugt und damit eine Leistungsminderung provoziert. Streß ist hierbei gekennzeichnet als eine überflüssige Belastung oder die Wegnahme von Gedankenkapazität und vermindert dadurch die eigentliche Handlungskompetenz.

### Die Gesetzmäßigkeiten des Lehrens und Lernens anwenden

Das dem m-c-Verfahren zugrundeliegende Denkmodell revidieren heißt: Die Gesetzmäßigkeiten des Lehrens und Lernens beachten und zur Anwendung bringen. Dies bedeutet für die Definition eines Prüfmodells, daß es eine Umkehrung des Lehr-/Lernverlaufes darstellen muß. Der Unterschied zu einem Lehr-/Lernsystem besteht dann darin, daß ein Prüfsystem keinen Lehrstoff vermittelt, sondern diesen voraussetzt. Das Prüfsystem muß jedoch den zu reproduzierenden Gegenstand in der sachlogischen Struktur belassen, in der er vermittelt und erworben wurde; es bietet statt Gegenstandsinformation Rückinformation

bezüglich der vom Prüfling vorgenommenen Entscheidungen an.

Das von uns vorgeschlagene Prüfsystem – eine detaillierte Ableitung kann bei den Autoren angefordert werden – unterscheidet sich in folgenden weiteren Komponenten vom m-c-Verfahren. Das System erzwingt keine einfache Reaktion, es versucht nicht nur das Denkresultat zu quantifizieren, sondern macht auch die Denkverlaufsstruktur sichtbar bzw. der Messung und Leistungsbewertung zugänglich. Eine Denkverlaufsanalyse wird eingeleitet, indem zu einem Problembereich ein Netzwerk von abhängigen, das heißt in einem logischen Verband stehende Fragen festgelegt wird.

Jede der vom Prüfsystem vorgegebenen Antwortalternativen ist direkt mit einer Folgefrage verknüpft, die eine logische Konsequenz der gegebenen Antwort darstellt.

Die Abbildung und Beschreibung der hierarchischen Gegenstandsstruktur z. B. einer Fallbeschreibung und ein prozeßbegleitendes Feedback sind das Kernstück dieses Prüfsystems. Es wird dadurch neben reinem Faktenwissen die persönlichkeitskonstituierende Qualität der Lernpotenz sichtbar gemacht, indem aus jeder Entscheidung des Prüflings eine denkverlaufsadäquate Konsequenz gezogen wird. Nur dadurch wird dem Subjekt der Tätigkeit die Möglichkeit der Selbstbewertung und Korrektur von Entscheidungsprozessen gegeben. Es wird ferner ein Weiterlernen angeregt sowie motivationale Grundlage für eine unmittelbare Weiterbearbeitung eingeleitet.

Auf eine Frage „Y“ folgt nicht mehr unabhängig von der jeweiligen Beantwortung eine Frage „Z“, sondern eine aus der Antwort von Y abgeleitete Frage  $Y_1$ . Bakterien haben nämlich nicht deshalb keinen von einer Membran umgebenen Zellkern, weil die Mainzer m-c-Auswerteschablone für diese Antwortalternative kein „Loch“ gelassen hat, sondern weil der Prüfling falsche – jedoch analysierbare und korrigierbare – Unter-

scheidungsoperatoren der Beantwortung und damit einer Entscheidung zugrunde legte. Diese Denkleistung muß modifiziert werden, indem sie beispielsweise in Widerspruch mit gegebenen Antworten gesetzt wird. Die Abbildung versucht eine graphische Umsetzung des Gesagten zu leisten.

Wir werden in nächster Zeit ein interaktiv arbeitendes Terminal programmieren und mit Fragen aus dem Bereich der Physiologie für Psychologen bestücken. Ob sich ein solches Prüfsystem auch als Papier- und Bleistift-Test realisieren läßt, kann erst nach dieser Erprobung beantwortet werden. Die Frage, wie die gemessenen Qualitäten (Lernpotenz und Denkverlauf) quantifiziert, das heißt einer mathematischen Auswertung zugeführt und damit zur Prognose herangezogen werden können, wird nur angedeutet.

Die Informationstheorie bietet eine Reihe ausgefeilter Verfahren an, mit denen ein lückenloses, empirisch gewonnenes Beantwortungsprofil ausgewertet werden kann. Diese Meßverfahren stellen nicht nur die Anzahl richtig oder falsch gelöster Aufgaben fest, sondern beantworten etwa folgende Fragen: Wie hoch war der Neuigkeitswert des Begriffs Y? Wie hoch war die Informationsmenge, mit der die Frage X beantwortet wurde? Wieviel Redundanz mußte erzeugt werden, bis der Begriff X erkannt wurde?

Ein zusammenfassendes Maß stellt zum Beispiel der Grad der Abweichung des empirischen Profils von einem – zu definierenden – optimalen oder pessimalen Profil dar.

Ablösen sollen dieses System zumindest die sogenannten „kleinen-m-c-Tests“, die aus methodologischer Sicht oft große Mängel aufweisen und eher ein Ritual des Zeitalters der Technik darstellen, als daß sie irgendeine Prognosefunktion hätten. Das System könnte ferner, da es durch die integrierten Feedback-Prozesse Erfahrungsbildung und Lernen erlaubt, in der ärztlichen Fortbildung benutzt werden. Für

den Bereich medizinischer Weiterbildung, der der Freiwilligkeit unterliegt, ist ein rigides Selektionsinstrument ohnehin fehlgegesetzt. Von Arnold (6) wird der m-c-Test deshalb als höchst unzulänglich bewertet, obwohl kein Alternativmodell zur Verfügung steht.

Für das Gebiet der Fortbildung könnte man unserer Meinung nach den Versuch wagen, psychologische Erkenntnisse, die zur Zeit vorrangig archiviert und verwaltet werden, einzusetzen und bei der Ist-Wert-Analyse statt einer Richtig-Falsch-Bewertung die jedem Denkprozeß zugrundeliegenden Operatoren aufzeigen und eventuell sofort korrigieren.

Für diesen Bereich muß die Behauptung, Prüfung sei kein Lernprozeß und beinhalte auch keinen Ausdruck situativer Lernfähigkeit, aufgegeben werden. Natürlich hoffen wir – und werden gezielt daran arbeiten –, daß dieses Prüfsystem den m-c-Wissenstest auch dort ablöst, wo es um mehr als die reine Überprüfung von Fakten oder Gedächtnisleistungen gehen sollte.

Diese Ablösung kann allerdings nur eingeleitet werden über konkrete Erfahrungen, die an einem alternativen Modell von Betroffenen gewonnen wurden. Diese Erfahrungen werden dann *nicht mehr nur* an intellektuell aufzeigbaren Mängeln des gegenwärtig gehandhabten Prüfungsmodells ansetzen, sondern *auch* subjektive und emotional erfahrene Kritikpunkte artikulierbar machen.

### Zusammenfassung

In unserer Arbeit geht es vornehmlich darum, einen Beitrag zur Artikulation des allgemeinen Unbehagens an den zur Zeit praktizierten Multiple-choice-Prüfungen zu leisten und die entstandene Diskussion um den Versuch der Neukonzeptualisierung einer alternativen, psychologisch-pädagogisches Wissen integrierenden, Prüfungsform zu erweitern.

Die von uns am m-c-Test geübte Kritik analysiert und systematisiert die

aus methodologischer und psychologischer Sicht aufzeigbaren Mängel insbesondere in bezug auf ihre inhaltlichen Konsequenzen und Implikationen. Die enge Verflechtung zwischen Studium, Prüfung und Berufspraxis läßt dabei verantwortlicherweise eine losgelöste Betrachtung nicht zu.

Als Konsequenz aus der Feststellung, daß es sich beim m-c-Prüfverfahren um die objektivere, bei der mündlichen Prüfung jedoch um die gültigere, der Struktur medizinischen Wissens gemäßere Prüfungsform handelt, favorisieren wir ein Prüfsystem auf der Basis eines interaktiv bedienbaren Computer-Terminals, das sowohl die Einhaltung konstanter Bedingungen als auch Flexibilität und inhaltliche Relevanz gewährleisten könnte.

Das Kernstück des Modells besteht aus einer Denkverlaufsanalyse gegebener Antworten und der adäquaten Ableitung von Konsequenz-Fragen; es handelt sich um die Integration von qualifiziertem Feedback in den Prüfprozeß.

Die Behauptung, Prüfung sei kein Lernprozeß, wird von uns aufgegeben.

### Literatur

- Flörkemeier, V., Kerschbaum, T.: Objektivierte Leistungskontrollen in der medizinischen und zahnmedizinischen Ausbildung, DEUTSCHES ARZTEBLATT 27, 1:74, S. 2133–2139 – Lienert, G. A.: Testaufbau und Testanalyse, Weinheim 1968 – Michaelis, J.: Die schriftlichen Prüfungen nach dem Antwort-Wahl-Verfahren, in: IMPP Aufgaben – Entwicklung – Analysen, Mainz 1977, S. 61–206 – Wehner, T.: Empirische Studie über M-C-Aufgaben, Münster 1978 (unveröffentl. Manuskript) – Guthke, J.: Zur Diagnostik der intellektuellen Lernfähigkeit, Berlin (DDR) 1972 – Arnold, M.: Erkenntnisse – aber nicht für die ärztliche Fortbildung, DEUTSCHES ARZTEBLATT 8, 1976, S. 519–521.

Anschrift der Verfasser:  
Theo Wehner  
Peter Kruse  
Wasserweg 59  
4400 Münster