

## Kinderspielplätze

kommt Karussell, Puppen, Kasperl, Kaufladen<sup>9</sup>).

Mit Recht heißt es in der Deklaration des Kinderhilfswerkes: „Eine eigenständige Lebensform des Kindes ist das Spiel. Besonders für das Kind der Industriegesellschaft hat das Spiel unverzichtbare Funktionen. Es muß helfen, Bedürfnisse zu befriedigen, die im Alltag nicht befriedigt, unterdrückt oder gar nicht geäußert werden können. Es muß ‚Lernraum‘ und ‚Experimentierraum‘ sein und zur Entfaltung von Fähigkeiten beitragen, die im alltäglichen Leben nicht oder nicht genügend zur Entwicklung gelangen. Dazu gehören Initiative, Aufgeschlossenheit, Einsatzbereitschaft, Fähigkeiten zum Problemlösen, schöpferische Fähigkeiten, viele Formen des mitmenschlichen Verhaltens und zahlreiche Fertigkeiten im Umgang mit Materialien und Geräten. Im Spiel kann das Kind über sich selbst verfügen; es kann aus eigenen Fehlern lernen; es hat eigene Erfolgserlebnisse.“

Schaffung von mehr und besseren Spielplätzen ist daher eine Hauptforderung für die Zukunft für die gesunde körperliche und seelische Entwicklung unserer Kinder und Jugendlichen. Auch die Schulhöfe sollte man mehr als öffentlich zugängliche Spielplätze ausbauen und statt der vielen staubigen und häßlichen Schulhöfe mit Asphalt- oder Kiesboden Spielhöfe mit Grünflächen und Spielgeräten schaffen. Wir brauchen aber auch mehr Spazierwege, Radfahrwege, Grünflächen, Schrebergärten, Kleinsportanlagen, Jugendherbergen, Freizeitheime, wenn wir unsere Städte mehr vermenschlichen wollen und die körperliche und seelische Gesundheit unserer Kinder und unserer Bürger nicht weiter aufs Spiel setzen wollen!

### Literatur beim Verfasser

Anschrift des Verfassers:  
Dr. med. et phil. Erich Grassl  
Allgemeinarzt, Dipl.-Psychologe  
Boschetsriederstraße 75/1  
8000 München 70

## FORUM

# Fehlerhafte Multiple-choice-Fragen im Kreuzfeuer der Kritik (III)

Zweite Fortsetzung von Heft 7/1981, Seite 302 ff., und Schluß

Rolf Buhl

### Empirischer Vortest als unabdingbarer Kontrollfilter

Sucht man in der Testliteratur nach Untersuchungen über Folgen fehlerhafter Fragen, so ist das Ergebnis mager. Neben zwei Arbeiten aus dem angelsächsischen Raum, wo in Tests bewußt Verstöße gegen MC-Regeln eingebaut wurden, um ihre Auswirkung auf die wichtigsten Testparameter zu untersuchen (Literatur angegeben bei Skannell/Tracy, a. a. O. S. 143), wird nur bei H. Freibichler (Aufgabenarten bei objektivierte Lehr- und Testverfahren, Hannover 1976, S. 25) kurz auf eine amerikanische Untersuchung Bezug genommen: „Wichtig erscheint das Ergebnis, daß schwierige Auswahlaufgaben gefährlich sein können.“

Dieses gegenüber der offensichtlichen Bedeutung des oben Ausgeführten doch auffallende Defizit hat nun aber seinen Grund in der geübten Praxis standardisierter Tests und ihrer Erstellung. Denn regelmäßig werden dort fehlerhafte Fragen nicht nur in einem sichtenden Revisionsverfahren ausgemerzt, sondern es wird zusätzlich als unabdingbarer Kontrollschritt ein empirischer Vortest durchgeführt, in dem ein weiterer Teil von Fehlern sich zu erkennen gibt und beseitigt werden kann. Nach mehreren Durchgängen liegt dann schließlich ein praktisch fehlerfreier Test vor, bei dem das Problem der Folgen von Fehlern sich nicht mehr stellt.

G. A. Lienert als „Altvater“ standardisierter Leistungstests erklärt lapidar und unmißverständlich, „daß der

Gebrauch von MC-Aufgaben bei nicht analysierten Testen einer Disqualifikation gleichkommt“, weil „nur im Rahmen der späteren Aufgabenanalyse entschieden werden kann, ob die Fragen die Grundbedingungen für ihre Testgeeignetheit erfüllen“ (Lienert, Testaufbau und Testanalyse, S. 34).

Nach Lage der Dinge ist bei den Medizinerprüfungen ein ausreichender Pool vorgetesteter Fragen erst im Wachsen und im Zuge sich schnell ändernden Wissensstandes keine Gewähr für bleibende Fehlerfreiheit. Bei der Verwendung von ungetesteten Fragen fehlt also, und darüber läßt sich durch noch so sorgfältige Revisionsarbeit am grünen Tisch nicht hinwegtäuschen, einer der wesentlichsten Kontrollfilter zur Qualitätssicherung der Fragen. Die Testliteratur ist voll von Angaben darüber, wie viele Fragen sich erst bei Vortests als fehlerhaft zu erkennen geben.

Nun fehlt dieser Filter in Form eines Vortests auch in anderen Ländern, wo Medizinerprüfungen nach MC-Verfahren abgehalten werden.

Es besteht aber ein entscheidender Unterschied zwischen dem IMPP und den dortigen Instituten, der nicht deutlich genug herausgestellt werden kann. Im Anschluß an ein Examen werden alle Fragen einer insgesamt aufwendigen computergestützten sogenannten Itemanalyse unterzogen. Dabei werden für jede Frage auf Grund des vorliegenden Antwortverhaltens der Prüfungskandidaten Daten gewonnen,

die Auskunft über Qualitätsmerkmale der gestellten Fragen geben und dabei fehlerhafte Fragen erkennen lassen.

► Das amerikanische „National Board of Medical Examiners“, das als Vorbild für das IMPP gelten kann, führt diese Itemanalyse durch, aber „noch vor der Ermittlung der Prüfungsergebnisse . . . Das ist eine Art Ersatz des Vortests nach dem Test. Stellt sich dabei heraus, daß eine einzelne Testfrage die Erwartungen nicht erfüllt . . ., kann eine solche mangelhafte Frage vor der Ermittlung des Endergebnisses der Prüfung eliminiert werden. So ist es möglich, jede Prüfungsfrage eine praktische Bewährungsprobe bestehen zu lassen, bevor sie auf das Prüfungsergebnis eines Kandidaten Einfluß gewinnt“ (Hubbard, a. a. O. S. 93 f.).

► Ganz entsprechend heißt es in einer diesbezüglichen Veröffentlichung des IAE: „Nachdem die Kandidaten die Fragen eines Examinens . . . bearbeitet haben, schließt sich eine statistische Analyse an, deren Sinn und Zweck weitgehend unbekannt geblieben ist. Statistische Indices spielen eine Rolle bei der Entscheidung, ob eine Frage beim Zusammenzählen des Prüfungsergebnisses des Studenten mitgerechnet werden soll“ (H. Vorkauf in: Neue Verfahren der Mediziner-ausbildung, Sep. Druck der Schweiz. Ärztezeitung a. a. O. S. 10). Gibt sich dabei eine Frage, die trotz aller vorheriger Kontrollfilter durch die Lappen gegangen ist, als fehlerhaft zu erkennen, so gilt: „Es versteht sich von selbst, daß solch eine Frage bei der Bewertung der Leistung der Kandidaten ausgeschlossen werden muß, will man nicht gegen den Gleichheitsgrundsatz . . . verstoßen“ (Vorkauf, a. a. O. S. 13).

► Demgegenüber ist in dieser Hinsicht für das IMPP gar nichts selbstverständlich. In den Ausführungen über das Auswertungsverfahren heißt es zwar beiläufig und auffallend parallel formuliert: „Es versteht sich von selbst, daß parallel zu dieser Ergebnismitteilung und vor Be-

kanntgabe der Resultate gleichsam als zusätzliche Kontrolle erste Aufgabenanalysen durchgeführt werden“ (IMPP: Aufgaben . . . S. 212). Aber die Prüfungszeugnisse sind da ja schon geschrieben und bleiben also unbeeinflusst davon. Bis Anfang 1980 sind denn auch noch nie Fragen nachträglich herausgenommen worden.

Auch das IMPP führt die aufwendige Itemanalyse durch und entdeckt dabei durchaus immer wieder fehlerhafte Fragen: „Die folgende Aufgabe ist in der vorliegenden Form nach Kenntnis der Auswertungsergebnisse nicht geeignet“ (IMPP: Analysen S. 107). Doch in krasser Dissonanz zum Vorgehen der oben genannten Institute führt das IMPP diese Fehler aufdeckende Analyse durch, um Hinweise für eine Fragenrevision im Blick auf zukünftige Wiederverwendung zu bekommen: „Die Informationen aus den Itemanalysen bilden eine Basis für die Entscheidung, ob eine Aufgabe in künftigen Examen unverändert, modifiziert oder gar nicht mehr verwendet werden soll“ (IMPP: Aufgaben . . . S. 77).

Wird vom IMPP die Analyse ausschließlich im Blick auf künftige Examenverwendung durchgeführt, so bei den auswärtigen Instituten vor allem, um daraus resultierende Ergebnisse zunächst einmal den von der jeweiligen Prüfung Betroffenen zugute kommen zu lassen.

So selbstverständlich dies nach dem Ausgeführten jedem Leser erscheinen mag, so sehr wirft der „kleine“ deutsche Unterschied ein bezeichnendes Licht auf die Selbstüberschätzung, perfekt genug zu sein, um darauf verzichten zu können. Hier wird eine Schlüsselstelle sichtbar, warum es überhaupt zu diesen inzwischen zur Unzahl angewachsenen verwaltungsgerichtlichen Klagen kommt. Viele Studenten sehen sich angesichts fehlerhafter Fragen um eine ihrem Leistungsvermögen entsprechende Bestehens-Chance betrogen, wenn das Prüfungsergebnis nur knapp unter der Bestehensschwelle liegt. Angesichts der aufgezeigten Sachlage sehen sie nur im

Widerspruchs- und Klageweg eine Möglichkeit, vorenthaltenes Recht durchzusetzen.

Die Einbeziehung eines nachgezogenen Vortests nach dem Muster auswärtiger Institute in den Verfahrensablauf der Medizinerprüfungen ist meines Erachtens als entschiedene Forderung zu stellen. Mit einem Schläge würde ein erheblicher Anteil der Klagen überflüssig. Auch Lehrmeinungsunterschiede, die Einfluß auf die Beantwortbarkeit haben, können erfaßt werden, wie das Schweizer Beispiel zeigt, da auffällige regionale Antwortunterschiede einen durchaus sensiblen Hinweis auf diesbezügliche Sachverhalte geben (dazu auch IMPP: Analysen S. 108). Es ist mir unverständlich, daß dies nicht schon von Anfang an in den Verfahrensablauf integriert worden ist in Anlehnung an Institute mit Erfahrung, die ihre Begrenzung beim Wunsch nach Perfektion wohl besser einzuschätzen wissen, als es sich im Vorgehen des IMPP erweist. Schwierigkeiten bei der Feststellung der aktuellen Bestehensgrenze, wenn fehlerhafte Fragen eliminiert sind, können kein schlagender Grund dagegen sein und werden offensichtlich im Ausland gemeistert.

Nach Fertigstellung des Manuskriptes haben die Frühjahrsprüfungen 1980 stattgefunden. Offenbar unter dem Eindruck des scharfen Protestes vom Herbst sind nicht nur auffallend weniger fehlerhafte Fragen gestellt worden, es ist diesmal zum ersten Mal vom IMPP eine Liste herausgegeben worden, auf der in bescheidenem Umfang für einige Fragen zusätzliche oder gar alle Antwortalternativen als richtig gewertet angegeben wurden. Man kann gespannt sein, wie vom IMPP diese „Neuerung“, die in Wirklichkeit Rückkehr zu allgemein geübter Testpraxis ist, nach nun sechsjähriger Erfahrung begründet wird. Da, wo es besonders nötig gewesen wäre, im Herbst '79, ist es entgegen den Angaben in der Spiegel-Veröffentlichung jedenfalls nicht zu einer solchen nachträglichen Anerkennung gekommen. ▷

## Multiple-choice-Fragen

### Auswirkungen fehlerhafter Fragen auf die Mißerfolgsquote

Wie viele Studenten bei den Herbstprüfungen im Umfeld der Bestehensgrenze durch Auswirkung fehlerhafter Fragen das Examen nicht bestanden haben, läßt sich durch folgende Überlegung etwa abschätzen: Wer im Rahmen seiner Leistungsmöglichkeiten auf die Chance angewiesen war, auch die fehlerhaft gestellten Fragen richtig beantworten zu können, hat ja per definitionem etwa 60 Prozent der fehlerfreien Fragen richtig beantwortet. Wären die zu beanstandenden Fragen ebenfalls fehlerfrei gewesen, so wäre auch bei diesen mit einer etwa 60 Prozent richtigen Beantwortung zu rechnen gewesen. Ohne auf die Problematik einzugehen, wie hoch man die Ratewahrscheinlichkeit ansetzen muß (vgl. dazu z. B. Möglichkeiten und Grenzen der Testanwendung, hrsg. K. H. Ingenkamp, Weinheim 3. Aufl. 1976, S. 243 ff.), die insbesondere bei fehlerhaft gestellten Fragen eher gering anzusetzen ist, seien für diese Abschätzung 20 Prozent angenommen. Das heißt, daß statistisch gesehen alle Kandidaten, die bis zu 40 Prozent (als Differenz von 20 Prozent auf 60 Prozent) der fehlerhaft gestellten Fragen unterhalb der Bestehensgrenze das Examen nicht bestanden hatten, noch eine reelle Bestehens-Chance gehabt haben.

Bedenkt man, daß für den II. Prüfungsabschnitt im Herbst 1979 im Intervall bis zehn Fragen unter der Bestehensgrenze etwa 160 Prüfungskandidaten lagen, im III. Teil angesichts der insgesamt geringeren Fragenzahl bis acht Fragen unter der Bestehensgrenze 300 Studenten, so kann jeder leicht abschätzen, wie beträchtlich sich fehlerhafte Fragen auf die Mißerfolgsquote ausgewirkt haben. Die hohe Zahl der verwaltungsgerichtlichen Klagen rückt hier in ein verständliches Licht.

Zugleich machen diese Zahlen folgendes deutlich: Mit Einführung der 60-Prozent-Regelung liegen im Umfeld der Bestehensgrenze schon ei-

ne erhebliche Anzahl von Prüfungskandidaten. Sie alle haben einen Anspruch darauf, daß im Testverfahren alle Möglichkeiten ausgeschöpft werden, angesichts des scharfen Bestehensschnittes von einem Punktwert auf den anderen die Bandbreite der Testunschärfe so gering wie möglich zu halten. Bedenkt man zusätzlich, daß in einer weitverbreiteten Tabelle der sogenannte Standardmeßfehler bei 150 gestellten Fragen schon bei plus/minus sieben Punkten liegt (dazu: Möglichkeiten und Grenzen der Testanwendung, S. 87), so addieren sich diese beiden Unschärfen schon gewaltig und betreffen im III. Teil weit über die Hälfte derer, die das Examen nicht bestanden haben!

Wie sehr Meßfehler und eingeschränkte Zuverlässigkeit sich auf die Prüfungsergebnisse ausgewirkt haben, wird in einer sehr gründlichen feinstatistischen Analyse früherer Examensrunden in einer mehrteiligen Serie der Zeitschrift des Marburger Bundes „Der Arzt im Krankenhaus“ 1980, Heft 4 ff., aufgezeigt. Die Autoren Prof. Dr. med. H. Kuni und Rechtsanwalt P. Becker kommen dabei zu dem Ergebnis, daß weitaus die meisten der nicht bestandenen Examina nicht mit der zu fordernden Sicherheit unterhalb der gesetzten Bestehensgrenze liegen (a. a. O. S. 408)!

Da vom Gesetzgeber mit Recht für Berufszulassungsprüfungen gefordert wird, daß die Versagung der Berufszulassung nur mit auf den Einzelfall bezogenen Gründen ausgesprochen werden darf, ist die Frage wohl berechtigt, ob dieser Forderung bei der aufgezeigten Unschärfe in ausreichendem Maße Rechnung getragen ist. Zumindest ist zu erwägen, inwieweit diese Unschärfe sich noch wesentlich von dem unterscheidet, was in den früheren mündlichen Prüfungen an Trennschärfe zwischen „bestanden“ und „nicht bestanden“ erreichbar gewesen ist.

Auf jeden Fall dürfte deutlich geworden sein, daß das Auflösungsvermögen der MC-Prüfung wesentlich geringer ist, als es der scharfe Beste-

hensschnitt erwarten läßt. Bei näherem Hinsehen ist die Objektivität der Bestehensentscheidung erheblich eingeschränkt.

In noch einer weiteren wesentlichen Hinsicht hat die so als Vorzug gepriesene Objektivität der MC-Prüfungen ihre Feuertaufe zu bestehen – oder besteht sie eben nicht, nämlich in der Frage, inwieweit die mit dem Anspruch der Objektivität getroffene Prüfungsentscheidung einer dementsprechenden objektiven gerichtlichen Überprüfung zugänglich ist.

### Objektivität der MC-Prüfungen und gerichtliche Überprüfbarkeit

Im Anschluß an die Herbstprüfungen haben weit über hundert betroffene Studenten Widerspruch gegen ihren Prüfungsbescheid eingelegt und den Klageweg beschritten. Wie aber erklärt sich dabei die vom IMPP so betonte fast 100prozentige Erfolglosigkeit solcher Klagen, wo doch fehlerhaft gestellte Fragen außer Zweifel stehen? Man wird nicht umhinkommen, hinter die Kulissen zu schauen und sich einen Einblick in die Rechtspraxis zu verschaffen, die sich im deutschen Verwaltungsrecht für die Überprüfung von Prüfungsentscheidungen herausgebildet hat.

Herausgebildet hat sich die geübte Rechtspraxis an den allgemein bekannten früheren Prüfungsformen wie der mündlichen bzw. schriftlichen Prüfung. Rechtlicher Schlüsselbegriff ist dabei der dem Prüfungsgremium eingeräumte gerichtsfreie subjektive Beurteilungsspielraum für fachlich-wissenschaftliche Bewertungen der Prüfungsleistung, der ausschließlich hierbei und bei Beamteneignungsfragen zugestanden ist (BVerwG DVBl 66, 571; H. Hummel, Gerichtsschutz gegen Prüfungsbewertungen. Schriften zum Öffentlichen Recht Bd. 105, Berlin 1969, S. 51).

Wegen seiner Konkurrenz zur Rechtsschutzgarantie (Art. 19 IV GG) war dieser Beurteilungsspielraum seit jeher umstritten (dazu N. Nie-

hues, Schul- und Prüfungsrecht, München 1976, S. 57 f), so daß T. Schramm (Beurteilungen im Prüfungs-, Schul- und Beamtenrecht, Köln 1975) feststellt: „Es gibt dankenswerterweise nicht viele Gebiete, in denen die Judikatur des höchsten Verwaltungsgerichts derart wechselhaft wie in sich widersprüchlich ist“ (S. 19).

Grundgedanke dabei ist, daß in einer Prüfung erbrachte Leistungen vom Prüfer nicht voll objektiviert werden können, sondern ein nicht unerheblicher Rest an subjektiver Beurteilung bleibt, der sich darin widerspiegelt, daß erfahrungsgemäß unterschiedliche Prüfer zu unterschiedlichen Ergebnissen kommen können (BVerwGE 8, 272 ff 1959), ohne daß die ja theoretisch vorauszusetzende einzig „richtige“ Beurteilung objektiv festgestellt werden könnte (BVerwGE 5, 153 1957). „Eine volle gerichtliche Überprüfung führe nur dazu, daß das Gericht seine nicht weniger subjektive Wertvorstellung an die Stelle ... setze“ (Hummel, S. 16). Dementsprechend heißt es in einem Urteil des Oberverwaltungsgerichts Lüneburg vom 27. 2. 76 (II OVG A 20/76, zitiert nach: Entscheidungen der Oberverwaltungsgerichte Bd. 31, Münster 1979, S. 317), daß „das Fachurteil der Prüfer nicht durch ein abweichendes Urteil anderer Prüfer oder anderer Stellen einschließlich der Gerichte ersetzt werden kann“.

Deswegen gilt die Entscheidung des Prüfungsgremiums bei „vermuteter Fachkompetenz“ als „höchstpersönliches Urteil“ (BVerwGE 11, 165 1960), das „unvertretbar“ sei (BVerwGE 8, 272 ff 1959) und dem auch gegenüber den Gerichten die Befugnis der Letzterkenntnis zukomme (P. Guhl: Prüfungen im Rechtsstaat. Rechtsstaatliche Anforderungen an Prüfungsverfahren, Bad Honnef 1978, S. 34 f.).

Gegenüber diesem gerichtsfrei gehaltenen Beurteilungsspielraum kann das Gericht zur Wahrung der Rechtsschutzgarantie überprüfen, ob dieser Beurteilungsspielraum nicht durch Willkür überschritten

und mißbraucht ist. Das Gericht tut dies, indem es überprüft,

- ▶ ob bei der Prüfung Verfahrensfehler unterlaufen sind,
- ▶ ob die Prüfer von falschen Tatsachen ausgegangen sind,
- ▶ ob sachfremde Erwägungen angestellt worden sind,
- ▶ ob (allgemein gültige) Bewertungsgrundsätze nicht beachtet worden sind (BVerwGE 8 272 ff 1959).

Was der in dieser Weise als „gesicherte“ Rechtsprechung geltende Beurteilungsspielraum für jemanden bedeutet, der den Klageweg einschlägt, erhellt beispielhaft aus einem Urteil des Bundesverwaltungsgerichtes vom 6. 1. 1977 (BVerwG VII B 182, 76), wo gegen eine Physiknote im Abitur geklagt wurde und das Gericht die Anhörung eines Sachverständigen zur Wahrung des Beurteilungsspielraumes rechtskräftig ablehnte:

„Der Kläger meint, die Prüfer, die seine Prüfungsarbeit in Physik beurteilt hätten, seien von einem falschen Sachverhalt ausgegangen, hätten also einen Fehler begangen, der – trotz der nur beschränkten Überprüfbarkeit von pädagogisch-wissenschaftlichen Bewertungen durch die Gerichte – der verwaltungsgerichtlichen Nachprüfung unterliege; der Umstand, daß die Prüfer von einem falschen Sachverhalt ausgegangen seien, hätte durch den Sachverständigen nachgewiesen werden können. Dem kann jedoch nicht gefolgt werden ... Was ihnen der Kläger vorwirft, ist vielmehr, daß sie seine Arbeit und die in ihr enthaltenen fachlichen Aussagen falsch bewertet hätten, indem sie von einer unrichtigen fachlichen Auffassung ausgegangen seien ... Diese nach Meinung des Klägers unrichtige Auffassung der Prüfer zu bestimmten fachlichen Fragen bedeutet aber nicht, daß die Prüfer von falschen Tatsachen ausgegangen seien ... Vielmehr handelt es sich dabei um eine bestimmte wissenschaftliche Auffassung, die der pädagogisch-wissenschaftlichen Beurteilung zugrunde liegt und – ebenso wie diese Beurteilung – der gerichtlichen Nachprüfung nicht zugänglich ist. Was der Kläger als falschen Sachverhalt bezeichnet hat, ist mithin in Wahrheit eine vom Kläger für falsch gehaltene, insoweit aber vom Gericht nicht überprüfbare Meinung zu pädagogisch-wissenschaftlichen Fachfragen ...“ (zitiert nach Schulrecht. Ergänzbares Sammlungs SchuSPE 41 III D I.)

### Subjektiver Beurteilungsspielraum und MC-Prüfungen

In diese „Rechtslandschaft“ hinein sind die Multiple-choice-Prüfungen geboren worden. Ihre Einführung wurde seinerzeit neben der unbestrittenen Ökonomisierung des Prüfungsverfahrens bei sonst kaum zu bewältigenden Studentenzahlen vor allem damit begründet, daß MC-Prüfungen als objektivierbare Leistungsmessung gegenüber der oft unerfreulichen Subjektivität früherer Prüfungsentscheidungen ein Höchstmaß an Objektivität mit sich brächten. „Der Vorteil solcher Prüfungen besteht ... darin, sie sind objektiv, transparent, nachprüfbar“ (IMPP: Aufgaben, S. 18).

Neben der Durchführungsobjektivität und der Auswertungsobjektivität ist hier vor allem die Interpretationsobjektivität, das heißt die Unabhängigkeit vom Prüfer, von Interesse. Erst jüngst hat Prof. Gebert als Abteilungsleiter des IMPP diese Objektivität im Einklang mit der Testliteratur erneut klar definiert: „Objektivität meint, daß das Prüfungsergebnis nicht von den subjektiven Auffassungen des Prüfers beeinflusst werden kann“ (Gebert in: Der Internist 1980, S. 136). „Die Überlegenheit der MC-Prüfung in bezug auf die Objektivität, also die Ausschaltung subjektiver Momente aus der Leistungsbewertung, ergibt sich aus der Methodik selbst“ (a. a. O. S. 139). Dem entspricht die vom IMPP gesehene Notwendigkeit, „Examensfragen zu erarbeiten, die nicht nur eine klare und eindeutige Antwort ermöglichen, sondern auch einer späteren gerichtlichen Nachprüfung standhalten“ (Kraemer, DEUTSCHES ARZTEBLATT 1978, S. 188).

Nach der oben geschilderten Rechtspraxis, die sich herausgebildet hat, wird jedem deutlich sein, warum das IMPP feststellen muß, „daß die von der Rechtsprechung und Literatur entwickelten Grundsätze zur gerichtlichen Kontrolle von Prüfungsentscheidungen unter Umständen nicht ohne weiteres auf das Antwort-Wahl-Verfahren übertragen werden können“! (IMPP Aufgaben, S. 47). ▶

## Multiple-choice-Fragen

Bei Anwendung der obigen Rechtspraxis hätte die Objektivität des MC-Verfahrens keine Entsprechung in einer adäquaten gerichtlichen Überprüfbarkeit.

Man ist nun aber aufs äußerste überrascht, wenn man sich daraufhin Gerichtsurteile von Klagen ansieht, die Medizinstudenten gegen fehlerhafte Fragen angestrengt haben. In einem Urteil des Oberverwaltungsgerichts Münster vom 23. 11. 79 heißt es, daß das IMPP als Beklagte „beantragt, die Klage abzuweisen. Es hat vorgebracht, die Frage, ob bei einer wissenschaftlichen Prüfung eine Antwort als die einzig richtige oder als die beste von mehreren möglichen Antworten anzusehen sei, falle in den pädagogisch-wissenschaftlichen Beurteilungsspielraum der Prüfer, der der richterlichen Nachprüfung nicht unterliegt. Es wiederholt seine Auffassung, daß der dem Prüfungsausschuß auch im standardisierten Verfahren zustehende Beurteilungsspielraum die Hinzuziehung eines Gutachters verbiete.“ Das Gericht entspricht diesem Antrag in Anlehnung an die übliche Rechtspraxis: „Dem Gericht ist die inhaltliche Auseinandersetzung mit dem größten Teil der . . . erhobenen Rügen verwehrt. Auch für eine Beweiserhebung durch Einholung eines Sachverständigengutachtens ist kein Raum.“

An diesem Urteil wird in besonderer Weise deutlich, daß nicht nur die Richter sich auch bei MC-Prüfungen an die ihnen vorliegende Rechtsprechungspraxis halten, sondern daß vor allem das IMPP selbst diese für seine „Verteidigung“ in Anspruch nimmt. Ganz abgesehen davon, daß das IMPP damit in unglaublicher Weise doppelzünftig von der Nachprüfbarkeit von MC-Fragen und deren „Gerichtsfestigkeit“ spricht, wird dabei die vielgepriesene Objektivität des MC-Verfahrens unterlaufen und praktisch aufgehoben.

Das IMPP zieht sich zur Festlegung der jeweils einen richtigen Lösung gewissermaßen ins Konklave des gerichtsfreien subjektiven Beurteilungsspielraumes zurück. In dieser

unüberprüfbaren subjektiven Lösungsfestsetzung kann ich keinen großen Unterschied zu der vielbelagten Subjektivität früherer mündlicher Prüfungsformen sehen. Daran vermag auch der Hinweis auf eine große Sachverständigenkommission des IMPP nichts zu ändern, wie die mit vielen Gutachten belegten fehlerhaften Fragen erweisen.

Wer aber nicht hinter die Kulissen der Gerichtsverhandlungen schaut (und wer unternimmt diese unerfreuliche Arbeit schon?), der wiegt sich in dem Glauben an den Vorzug der Objektivität von MC-Prüfungen. Wenn das IMPP ausdrücklich von der Nachprüfbarkeit der Fragen spricht, so muß das angesichts des obigen Gerichtsverfahrens als glatte Irreführung bezeichnet werden, der viele Medizinstudenten erlegen waren, als sie gemäß dem Objektivitätsanspruch den Klageweg gegen fehlerhafte Fragen einschlugen.

Nun mag jemand nach Aufdecken dieses Schwindels sagen, daß es nach geltender Rechtsauffassung keinen Rechtsanspruch auf Objektivität des angewandten Prüfungsverfahrens gäbe, da jemand, der mündlich geprüft wird, ja nicht fordern kann, statt dessen einem objektivierten Prüfungsverfahren unterzogen zu werden. Ganz abgesehen davon, daß die im Grundgesetz verankerte Rechtsschutzgarantie die Nachprüfbarkeit und damit Objektivität überall da bindend fordert, wo sie den Gegebenheiten entsprechend irgend möglich ist (vgl. Hummel, S. 101 f.), muß insbesondere für das MC-Verfahren ein weiterer wesentlicher Gesichtspunkt berücksichtigt werden. Die geradlinige Übertragung des gerichtsfreien Beurteilungsspielraums auch auf die Überprüfbarkeit fehlerhafter MC-Fragen hat in Gerichtsurteilen zu folgender Absteckung dieses Beurteilungsspielraums geführt, wie das schon zitierte Urteil des OVG Münster vom November 1979 zeigt:

„Ein Verstoß gegen allgemein gültige Bewertungsgrundsätze liegt noch nicht vor, wenn die Prüfer von einer umstrittenen, möglicherweise nicht widerspruchsfreien Bewertung eines wissen-

schaftlichen Tatbestandes ausgehen (BVerwG, Beschluß vom 31. Oktober 1978 7 B 39.77 – Buchholz 421. O Nr. 99). Vielmehr ist die Grenze des Beurteilungsspielraumes weiter zu ziehen. Nicht mehr durch ihn gedeckt sind die vom Prüfungsamt bzw. den von ihm beauftragten Wissenschaftlern zugrunde gelegten Kriterien und wissenschaftlichen Aussagen erst dann, wenn sie, wie der Senat mehrfach entschieden hat, unter keinem erdenklichen wissenschaftlichen Gesichtswinkel mehr gerechtfertigt sein können.“

Damit wird die Objektivität und zugleich das Funktionieren der MC-Prüfung quasi aus den Angeln gehoben. Denn wenn der Beurteilungsspielraum Fragen, die auf Grund von Lehrmeinungsunterschieden und ähnlichem für den Prüfungskandidaten nicht durch Wissen im Sinne des IMPP zu entscheiden sind, vor der juristischen Überprüfung schützt und dies nicht als Verletzung allgemein gültiger Bewertungsgrundsätze gilt, dann ist damit ein jedem einsichtiger Grundpfeiler des Funktionierens von MC-Prüfungen aufgegeben, und hinreichend zuverlässige Prüfungsergebnisse sind in Frage gestellt. Es sei denn, man schließt sich der Meinung des Gerichts und des IMPP an, daß bei der großen Anzahl der Fragen einzelne „spezielle“ Fragen (Prof. Gebert) oder „Unzulänglichkeiten der Fragestellung“ (Gerichtsurteil OVG Münster) nicht ins Gewicht fallen.

Die Testliteratur warnt ausdrücklich: „Besondere Vorsicht ist geboten, wenn zum Beispiel nach der besten Lösung gefragt wird. Bei der Auswahl und Formulierung der Wahlantworten dieser Art bedarf es besonderer Sorgfalt, denn der Übergang von der objektiven Richtigkeit zur subjektiv ‚richtigen‘ Meinung ist nicht genau faßbar. Man sollte versuchen, sich bei der Zusammenstellung der Wahlantworten in genügendem Abstand von dieser unsicheren Zone zu bewegen. Maßgebend ist im Zweifelsfalle der Standpunkt des Schülers und nicht der des Lehrers“ (F. Barrabas in: Das Bochumer Testverfahren, Stuttgart 1974, S. 76).

Wo hier die Perspektive des Prüfungskandidaten für maßgeblich gehalten wird, ist im Gerichtsverfahren

noch nicht einmal Waffengleichheit gegeben. Gilt die vom IMPP festgesetzte Lösung erst dann nicht mehr als gerechtfertigt, wenn sie „unter keinem erdenklichen wissenschaftlichen Gesichtswinkel mehr gerechtfertigt“ ist, so gilt noch lange nicht, daß die vom Prüfungskandidaten angekreuzte Lösung erst dann als falsch gewertet werden darf, wenn sie unter keinem wissenschaftlichen Gesichtswinkel mehr gerechtfertigt ist.

Diese offensichtliche Rechtsungleichheit, die sich aus dem Urteil des OVG Münster ergibt, hat ein Urteil des OVG Koblenz unabhängig davon zur gleichen Zeit als nicht zulässig erklärt:

„Mangelnde Eindeutigkeit darf nicht zu Lasten des Prüflings gehen, so daß, wenn eine zur Lösung gestellte Aufgabe ausnahmsweise doch einmal zwei Antworten zuläßt und beide Antworten in dem Prüfungsbogen angeführt sind, nicht nur eine bestimmte Antwort richtig sein kann. Gleichgültig welche der beiden Antworten der Prüfling ankreuzt, in jedem Fall muß die Antwort als richtig gewertet werden“ (Beschuß des OVG Koblenz 2. Senat 20. 12. 79).

Da dies aber oft erst in Sachverständigengutachten aufdeckbar ist, wird man gespannt sein, ob diese in Zukunft nun doch zugelassen werden. Erste Konsequenz dieses Beschlusses dürfte es gewesen sein, daß das IMPP, das ja im Zuständigkeitsbereich dieses OVG liegt, zum Frühjahr erstmals nachträglich zusätzliche Lösungen anerkannt hat in Form der oben schon erwähnten Liste zum Lösungsschlüssel.

Es ist beschämend, daß zu dieser jedem einsichtigen Maßnahme erst der Druck eines Gerichtsurteils nach sechs Jahren geführt hat.

Dabei ist nach allem Ausgeführten deutlich, daß es bei Fehlen einer Vortestung alles andere als eine Schande ist, wenn sich nachträglich Fragen als fehlerhaft zu erkennen geben. Diese aber in einem verbissenen Kampf mit allen juristischen Tricks zu halten versuchen, ist meines Erachtens sehr wohl eine Schande.

In einem schon vor Jahren gefällten Urteil des Münchener Verwaltungsgerichtshofes wird dem MC-System in sachlicher Konsequenz noch mehr Rechnung getragen.

Ohne den subjektiven Beurteilungsspielraum als „Tabu“ zu tangieren, wird festgestellt, daß entgegen den Ausführungen des IMPP bei einer Frage mehr als eine Lösung richtig war und damit gegen allgemeine Bewertungsgrundsätze verstoßen worden sei, die sehr wohl der gerichtlichen Nachprüfung unterliegen. Selbst wenn der Kläger seine Klage nicht in diese Richtung formuliert hat, muß das Verwaltungsgericht in diesem Sinne tätig werden.

„Im Verwaltungsprozeß gilt prinzipiell die Untersuchungsmaxime, das heißt, das Gericht erforscht den Sachverhalt von Amts wegen (§ 86 Abs. 1 Vw GO). Stellt es dabei . . . Umstände fest, welche die Klage rechtfertigen, so hat es sie seiner Entscheidung zugrunde zu legen, auch wenn der Kläger sie – weil sie ihm möglicherweise nicht bekannt waren – selbst nicht vorträgt“ (Niehues, Schul- und Prüfungsrecht, S. 57). In diesem Sinne ist der Münchener Verwaltungsgerichtshof tätig geworden und führt aus:

„Dem allgemeinen Bewertungsgrundsatz, daß Prüfungsfragen klar und zweifelsfrei gestellt werden müssen, kommt bei der vorliegenden Art der schriftlichen Prüfung besondere Bedeutung zu, weil dem Prüfer ein Nachfassen mit ergänzenden und klärenden Prüfungsfragen nicht möglich ist und dem Prüfling außer dem Ankreuzen einer Antwortalternative zusätzliche Erläuterungen nicht gestattet sind. . . . Als Ergebnis bleibt festzuhalten, daß in den der Frage . . . beigegebenen fünf Antwortmöglichkeiten nach allgemeinen Bewertungsgrundsätzen nicht mit hinreichender Zuverlässigkeit nur eine als zutreffend erachtet werden kann.

Die Frage . . . war deshalb nicht systemgerecht gestellt und konnte dem entsprechend auch nicht systemgerecht beantwortet werden. Sie war somit entgegen § 8 Abs. 2 AAppO [Anm.: es handelt sich um die Klage eines Pharmaziestudenten] nicht geeignet, ein zuverlässiges Prüfungsergebnis zu ermöglichen . . . Daß der Kläger keine der beiden zutreffenden Antworten angekreuzt hat, kann ihm nicht als unrichtige Antwort angelastet werden, weil die Frage systemwidrig gestellt war.

Es ist nicht auszuschließen, daß er bei rechtlich fehlerfreier Formulierung der Frage . . . eine zutreffende Antwortalternative angekreuzt und damit die Prüfung bestanden hätte . . . Der Kläger ist deshalb berechtigt, die Prüfung . . . erneut erstmalig, das heißt nicht als Wiederholungsprüfung abzulegen“ (Urteil vom 25. 7. 1977, Nr. 153 VII 77, zitiert nach SchuSPE 41 III F III 21).

In diesem Sinne ist das jetzt begonnene Vorgehen des IMPP dringend ergänzungsbedürftig: Nicht nur zusätzliche richtige Lösungen müssen nachträglich als richtig anerkannt werden, sondern zu systemwidrig gestellten Fragen falsch gegebene Antworten dürfen nicht als falsch gewertet werden.

Die hier gemachten Ausführungen sind gewissermaßen der heute praktizierten MC-Prüfung entlanggegangen: Vom Entstehen der MC-Fragen mit ihren typischen Konstruktionsklippen und ihrer großen Fehleranfälligkeit, den Auswirkungen fehlerhafter Fragen auf die Leistungsfähigkeit in der Prüfung und auf die Bestehenschancen bis hin zur derzeit geübten Rechtspraxis bei der gerichtlichen Überprüfung. Auf allen Ebenen sind erhebliche Mängel deutlich geworden, die allesamt zu Lasten der betroffenen Medizinstudenten gehen. Ungeachtet der grundsätzlichen Frage nach der Eignetheit von MC-Prüfungen hat sich gezeigt, daß das IMPP sehr wohl einen beträchtlichen Teil dieser Mängel zu beseitigen in der Lage sein sollte.

Wenn dieser Beitrag dazu Wegmarkierungen aufzeigen und möglichst viele davon überzeugen konnte, daß es neben der so leichten Globalkritik im Interesse der Medizinstudenten zur Zeit vor allem auch um konstruktive Kritik im Detail geht, dann hat er das von mir angestrebte Ziel erreicht.

Anschrift des Verfassers:  
cand. med. Rolf Buhl  
Kronenstraße 2  
5600 Wuppertal 2