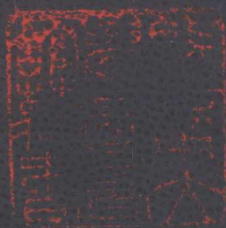


MULTIPLE-CHOICE EXAMINATIONS IN MEDICINE

A Guide for Examiner and Examinee

JOHN P. HUBBARD
WILLIAM V. CLEMANS



LEA & FEBIGER

2-44
E702 4

8591236

外文书库

HJP

MULTIPLE-CHOICE EXAMINATIONS IN MEDICINE

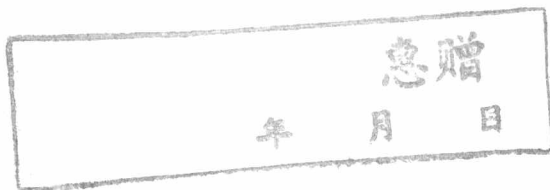
A Guide for Examiner and Examinee

John P. Hubbard, M.D.

Professor of Public Health and Preventive Medicine,
University of Pennsylvania School of Medicine;
Executive Director, National Board of Medical Examiners.

William V. Clemans, Ph.D.

Director of Testing Services, National Board of Medical Examiners
Philadelphia, Pennsylvania



LEA & FEBIGER
Philadelphia



Copyright © 1961 by Lea & Febiger

All Rights Reserved

Reprinted 1961

Reprinted 1965

HEALTH SCIENCES LIBRARY
UNIVERSITY OF MARYLAND
BALTIMORE

RC58
487
1961
v. 2

Library of Congress Catalog Card Number 61-9365

Printed in the United States of America

“Examinations are formidable even to the best prepared, for the greatest fool may ask more than the wisest man can answer.”

—Charles Kaleb Colton
Lacon, Vol. 1, #322

Preface

MULTIPLE-CHOICE examinations are being used more and more in various fields of medicine for many purposes. They are used in medical schools as departmental examinations or as comprehensive examinations at the end of a school year; they are used by several specialty boards as part of the testing procedure to determine certification for specialty practice; they are used by some state boards to determine qualification for a license to practice. Multiple-choice techniques are used by the National Board of Medical Examiners for its examinations which, when successfully completed, are recognized by state boards in lieu of further state board examinations. The Educational Council for Foreign Medical Graduates has developed multiple-choice examinations for worldwide testing of foreign physicians to determine their qualifications for hospital appointments in the United States (see Chapter 8).

In order for multiple-choice testing methods to be successful, both examiner and examinee need to be familiar with the techniques and the pitfalls that may be encountered. The manner in which the test questions are prepared and put together to form an examination, the procedures for scoring, analyzing, and reporting the results—all have a bearing upon the conclusions drawn from the performance of the individuals and groups tested.

This volume has therefore been written for both the examiner and the examinee. It describes the techniques and values of multiple-choice testing. Types of questions that are used in multiple-choice examinations in medicine, and a

sample of such questions in the form of a test with a separate key for the right answers, are included.

We have reviewed for the special interest of the examiner the methods of scoring and the detailed statistical analyses which increase the meaningfulness of the grades and provide data which allow for improving the examinations, and we have also shown the ways in which individuals' scores may be compiled in order to study group performances. The average score of the group, for example a single medical school class, may then be compared with other groups or classes. Or, in the case of National Board Examinations, which are nationwide in scope, one group or class may be compared with a nationwide sampling of student performance. Such comparisons provide factual and realistic evaluations of the caliber of the group and the excellence of the teaching program.

In this volume generous use has been made of the experience of the National Board of Medical Examiners. In 1952 the National Board converted its essay tests, which had been used for many years, to the multiple-choice form (see Chapter 6). At that time, the National Board benefited greatly from the assistance and co-operation of the Educational Testing Service, an agency which had achieved a leading role in the rapidly advancing science of testing. We take this occasion, therefore, to express our indebtedness to the Educational Testing Service. The testing techniques and the form of the questions used by the Educational Testing Service were adapted to the field of medicine, although the medical content of the examinations remained at all times in the hands of the examiners of the National Board. National Board examiners themselves wrote the questions and determined the level at which an individual candidate should be considered to have passed or failed the examination. In more recent years, the National Board assumed complete responsibility within its own organization for all phases of the preparation, scoring, statistical analysis,

and interpretation of the results of the examination. The methods described in the following pages are primarily those currently used by the National Board. The test questions themselves have been taken from actual tests administered by the National Board.

Since the test questions are the basic substance upon which this book depends, and since every one of these questions has been prepared by the examiners of the National Board, representing both individual effort and committee review (see Chapter 3), we wish to acknowledge the invaluable contribution made by this large body of examiners that now, owing to the rotation system of the committees, numbers well over one hundred. Their expert comprehension of the subject matter, and their prominence both in their special fields and in the broader area of medical education, are elements of prime importance in establishing the high quality of the examinations. They are the authors of all questions in the large collection now in the files of the National Board, and hence they are the authors of the examination questions included in this volume. There are too many past and present members of these committees to name them individually here, but we wish to acknowledge our obligation to them for the examination questions we have used to demonstrate method and medical content.

We also wish to express our indebtedness to the many members of the National Board's staff who have helped in the preparation of material for this volume, particularly Drs. Christian J. Lambertsen and George D. Ludwig, who work closely with the test committees in the development of the examinations and the maintenance of their high quality. Our special thanks are due to Mrs. Louise Hodgson for her assistance, especially in editing the manuscript, and to Mrs. Barbara Reis, who helped compile the bibliography.

Philadelphia, Pennsylvania

John P. Hubbard
William V. Clemans

Contents

Chapter		Page
1	Measurements of Medical Knowledge	9
2	Purposes and Types of Multiple-choice Questions	14
3	Construction of Multiple-choice Tests	30
4	Scoring Multiple-choice Tests	46
5	Analyzing Multiple-choice Tests	58
6	Multiple-choice Tests and Essay Tests	73
7	Group Comparisons	104
8	Multiple-choice Examinations for Foreign Physicians	120
Appendix		
A	Outline of Subject Matter for Individual Tests	129
B	Sample Test	141
C	Sample Answer Sheet	176
D	Key to Correct Responses	179

Chapter 1

Measurements of Medical Knowledge

EXAMINATIONS, however they may be devised and for whatever purposes, involve measurements and, therefore, like other forms of measurements, are characterized by a comparison between an unknown quantity or quality and a known standard or set of criteria. In the academic and professional field, the unknown factor may be an individual's understanding of a fact or principle which may be measured by comparing it with the accepted meaning of the fact or principle itself. Measurements much more comprehensive are made in academic areas when the unknown factor is the knowledge that the examinee has acquired by learning and experience, measured against the examiner's concept of the amount of knowledge that should be considered satisfactory or standard. Tests may also include acts or performances to be judged by the examiner as good or poor on the basis of his own judgment of good or poor performance.

When dealing with physical entities, the standards are usually well agreed upon and measuring instruments exist which can obtain almost any specified degree of accuracy. When dealing with human knowledge, however, standards may be variable or may not exist at all, and the measurements may vary with the judgment of the individual examiner. In a recent medical school examination, the following question appeared: "Discuss the effects of emotional stress." There are many different ways to approach such a question and

many difficulties in assessing the knowledge of the subject. One examinee might answer the question from the point of view of the physiological and biochemical effects of stress. A second examinee might answer the question more from a clinical point of view. Even if one examiner can score a set of such answers quite consistently, there is no assurance that another qualified reader might not judge the papers very differently.

Confronted with the problems and uncertainties of measuring knowledge, the medical profession—its medical educators and examining boards—is turning more and more to the science of educational measurement and to those skilled in testing techniques. Familiarity with these techniques and their judicious use brings to examinations in medicine a better understanding of the inherent variables, and a greater degree of reliability, comparability, and validity.

These terms, reliability, comparability, and validity, have developed important and specific meanings in the science of testing.

Reliability is the degree to which the measuring device can in fact be relied upon to provide consistent and reproducible results. To use a simple illustration, a ruler is reliable if the gradations are accurately laid out and if it is easily read so that when measurements of an object are made by one individual and then repeated by the same individual or by another individual, the second observations agree with the first. In like manner, a test of medical knowledge is reliable to the degree that it can produce consistent scores for individuals or groups of individuals tested on repeated occasions.

Comparability is a concept of higher order. First, it depends upon the degree of reliability; one test cannot be accurately compared with another unless each is reliable. Second, it requires certain points in one test which can be equated with comparable points in another test, such as the mean score. Third, it requires equal units of measure for the measuring

device concerned. Given two comparable tests, scores on one may be compared with scores on the other throughout the scale of measurement. For example, average scores of 85 obtained by two different classes in different years would imply the same level of performance if and only if the tests are comparable.

The *validity* of a measuring instrument is determined by the relationship which exists between a set of measurements obtained by it and other measurements obtained independently of it. To illustrate this point, let us assume that the problem is to determine which tables can go through a door that, according to specifications, has been made precisely 3 feet wide. A ruler is judged to be a valid measuring instrument, at least for this particular problem, if all tables less than 3 feet wide, as measured by the ruler in question, can go through the door, while tables measuring more than 3 feet wide cannot.

The validity of the measuring instrument always involves an understanding of the purpose of the measurement. The ruler, which proved to be valid in the measurement of tables, would obviously have no validity or usefulness in determining the difference between the diameter of an erythrocyte and that of a leukocyte. Similarly, a test of medical knowledge is valid to the degree that it distinguishes between those who, by accepted standards, are well qualified and those who are not. But then the question immediately arises: qualified for what? To be valid, a test must be constructed to serve its intended purpose. A test designed to determine a physician's knowledge of medicine may have a high degree of validity for this purpose, but may have little value in predicting his skill at giving an intravenous medication.

Increasing recognition of the fact that a test of medical knowledge should itself be subjected to critical evaluation of its reliability and validity has resulted in increasing use of the newer techniques of objectively scorable multiple-choice examinations. This is not to say that the time-honored essay

type of examination should now be discarded. The examination question quoted above, "Discuss the effects of emotional stress," has certain values in the measurement of student knowledge that can be derived only by the traditional essay response. Both types of examination questions have their advantages and their disadvantages; these will be considered in Chapter 6. The fact remains, however, that multiple-choice tests can be analyzed and studied in ways not possible for essay examinations. This is true because answers are not equivocal; they are either right or wrong. For this reason multiple-choice examinations are being used more and more in medical education and in tests of qualification, and the characteristics and techniques of these examinations must be well understood.

It would seem evident, therefore, that both those who use the techniques and those to whom they are applied should be familiar with the methods, the values, and also the limitations inherent in this type of examination. This familiarity, unhappily, does not always exist. The examiner, accustomed to phrasing an essay question, may think that the test composed of multiple-choice questions can be prepared as readily as an essay examination, and with no more expenditure of time and effort. Such an approach may result in measurements which have little or no relationship to those qualities of the student or the physician which the examiner is seeking to measure. A review of the various types of multiple-choice questions and the purposes for which they have been devised, as described in Chapter 2, should serve to impress upon the examiner the fact that the preparation of this type of test requires not only knowledge of the method, but also thoughtful diligence in its use.

It is equally important that the examinee be familiar with the type of question that he is to encounter. Assuming that the purpose of the test is to measure knowledge of the subject matter and not the examinee's ability to handle the techniques, the method of the examinations should be thoroughly

understood ahead of time. If the examinee fails to understand what is expected of him and for this reason makes a wrong response to a question, he gives an erroneous impression of his knowledge, and the test is a false and misleading instrument. A clear understanding of the objectives and the methods of the examination are therefore essential for both examiner and examinee as the one measures and the other is measured.

Chapter 2

Purposes and Types of Multiple-choice Questions

OBJECTIVE testing has progressed far beyond the relatively simple "true-false" type of question. Multiple-choice questions, with their present and potential variety of types, are far more searching and enable the examiner to select that type of question best suited to measure the particular point or function he has in mind. Certain types of questions test not only an individual's knowledge, but also the more subtle qualities of discrimination, judgment, and reasoning. Different types of questions test for the recognition of the similarity or dissimilarity of diseases, drugs, and physiologic or pathologic processes. Other questions evaluate understanding or judgment as to cause and effect, or the lack of causal relationships. Anatomical drawings, charts, diagrams, roentgenograms, electrocardiograms, or photographs may be reproduced to illustrate the situation about which many questions may be asked. Case histories may be used to simulate the experience of a physician confronted with a diagnostic problem; series of questions can then follow to determine the knowledge of related aspects of the case, such as associated laboratory findings, treatment, complications, and prognosis.

It is essential that the techniques used in these examinations be thoroughly familiar to both examiner and

PURPOSES AND TYPES OF MULTIPLE-CHOICE QUESTIONS 15

examinee: the examiner so that he may use these measuring devices to measure what he is really seeking to measure, and the examinee so that he can fairly display his knowledge of the subject matter without being bothered by technique.

Five-choice Completion (Type a)

The simplest and most widely used multiple-choice form consists of a question or incomplete statement followed by several (usually four or five) suggested answers or completions. Examinees are directed to select *the one best answer*. In the broad field of medicine, contrasts are not always sharply defined as black and white but are apt to be varying shades of gray. In answering these questions, therefore, the examinee is expected to look for the best or most appropriate choice and discard others which may appear plausible but definitely less applicable.

Here is a rather straightforward example of this so-called "completion type":

Item 1.* Of the following, the most effective prophylactic agent for the prevention of recurrences of rheumatic fever is

- (A) acetylsalicylic acid
- (B) para-aminobenzoic acid
- (C) adrenocorticotrophic hormone
- (D) cortisone
- (E) sulfadiazine

In the language of multiple-choice tests, reference is frequently made to "items" rather than "questions." In the above example, the leading clause, which is defined as the stem, is not a question but rather an incomplete statement followed by the five suggested completions. The stem might have been written as a complete question: "Which of the following is the most effective prophylactic agent for the prevention of recurrences of rheumatic fever?" Although there

* All items are numbered consecutively throughout the text. Correct responses are underlined or given in parentheses.



16 PURPOSES AND TYPES OF MULTIPLE-CHOICE QUESTIONS

is little advantage of one of these forms over the other, the incomplete statement is apt to be preferred since it can usually be expressed in a simpler manner with fewer words.

The suggested responses other than the correct one are called distractors. In the above example, E is the correct response and those designated as A, B, C, and D are incorrect responses or distractors. In the preparation of multiple-choice questions or items, the development of good distractors is one of the most difficult parts of the examiner's task. Each distractor should be a plausible answer and should fit into the context of the problem at hand. Silly wrong answers should be strictly avoided. Any distractor that is obviously wrong weakens the test. If, for example, two out of five choices are so clearly wrong as to present no problem to any of the examinees, the correct response becomes a one-out-of-three choice instead of the intended one-out-of-five.

One of the principles of multiple-choice testing is that the choice should indeed remain multiple. The main point of the above question could be written in the true-false form: "Sulfadiazine is an effective prophylactic agent for the prevention of recurrences of rheumatic fever." The examinee is then given one of two choices: true or false. With the true-false technique, an individual knowing nothing about the subject matter has a 50 per cent chance of guessing the right answer. The guessing factor is not eliminated in the multiple-choice item, but it is diminished. With one correct response out of five choices, random guessing would give only a 20 per cent chance of a correct response. Further consideration of the factor of random guessing appears in Chapter 4, which deals with scoring procedures.

One of the criticisms often made of the multiple-choice type of test arises from having the correct response in the question, even though mixed in with a number of distractors. This would seem to place a premium on recall memory at the expense of such qualities as critical judgment. It is not, how-