WORLD HEALTH ORGANIZATION

ORGANISATION MONDIALE DE LA SANTÉ

INDEXED

ORIGINAL: ENGLISH

INTER-REGIONAL SEMINAR ON TEACHING METHODS AND TEACHING AIDS

Ankara, 9-14 September 1968

BIBLIOTHÈQUE OMS
24 AOUT 1968
WHO LIBRARY

EVALUATION AND EXAMINATION IN MEDICAL EDUCATION

by

Professor John Anderson[1]
WHO Consultant

Evaluation or testing is a key area in medical education for it enables (a) the student to assess the value of his learning, (b) the teacher to assess the value of the student's learning, and (c) the effectiveness of the educational procedure through which the student has passed in relation to the student's learning and the teaching facilities. Standard examination procedures convey very little to the student especially about the worth of his learning and the staff rarely draw any conclusions about teaching from the results. Testing, however, is an essential part of the learning process. It should go on while learning is in progress to see if what the teacher has tried to convey has in fact been learned and has been useful in modifying the student's present behaviour. Evaluation should go on continuously throughout the medical course to allow it to change so as to overcome its errors and adapt to new knowledge and circumstances.

Attitudes to testing by both teachers and students are primarily emotional for both see it as an occasion when errors are revealed. They associate mistakes with ideas of failure and undesirable personality characteristics such as weakness, poor application, and social unacceptability. In reality many of us learn best from our mistakes and new adventures arise from old errors. It is important to see testing as a means of assessing learning for both the student and teacher and the only means of improving and increasing learning.

The evaluation techniques which have been used to date have been mainly used in qualifying examinations. These are the written essay, practical and oral examinations in the laboratory and at the bedside, and more recently objective multiple choice examinations. The following references list the advantages and disadvantages of such techniques:

1. Anderson, J. & Roberts, F. J. (1965) A new look at medical education

2. Charvat, J., McGuire, C. & Parsons, V. (1968) A review of the nature and uses of examinations in medical education, Public Health Paper Series, Wld Hlth Org., Geneva

3. Miller, G. E. (1961) Teaching and learning in medical schools, Harvard University Press

4. Peterson, O. L. et al. (1956) Analytical study of North Carolina general practice, J. med. Educ., 31, 12

5. Report of the Royal Commission on Medical Education (1968)

[1] Chairman, Department of Medicine, King's College Hospital Medical School, Denmark Hill, London.

UGE/68.10

The issue of this document does not constitute formal publication. It should not be reviewed, abstracted or quoted without the agreement of the World Health Organization. Authors alone are responsible for views expressed in signed articles.

Ce document ne constitue pas une publication. Il ne doit faire l'objet d'aucun compte rendu ou résumé ni d'aucune citation sans l'autorisation de l'Organisation Mondiale de la Santé. Les opinions exprimées dans les articles signés n'engagent que leurs auteurs.

What are good examinations? This, of course, is a question that depends much on the personal viewpoint of the party concerned, and one may well get as many different answers as there are people to judge. However there are some criteria which when fulfilled do make the examination more objective and thus more valuable. These are:

1. Reliability
2. Validity
3. Discrimination
4. Objectivity

As to the reliability of examinations, this refers to the consistency or accuracy of the scores yielded. A number of reasons can be given as to why an examination can be considered unreliable. 1. There is the small sample, i.e. too few questions. This can lead to guessing and by sheer chance low or high scores can be obtained by the student. Obviously such an examination will have little meaning and the test would be of low reliability. 2. The examination may be improperly administered. 3. The students may have been ill or unduly stressed when commencing their examination.

As to the validity of an examination, this can be understood as the degree of accuracy with which the examination really measures what it is intended to measure. For instance, if you want to find out whether the student knows the symptoms of hyperthyroidism it is of little value to ask him about the biochemistry of tri-iodothyronine.

With regard to the discrimination, an item is regarded as discriminating if it is answered correctly by more of the good students (those with high scores) than by poor students.

Regarding the objectivity of an examination, this becomes a serious problem once two or more examiners disagree on the answer given. Not only do several examiners often grade quite differently, which can have serious consequences for a student who scores just below or above the passmark, but there is evidence that an examiner's grading may differ at the beginning and the end of a long examination period.

It is with the above in mind that objective examinations need to be implemented. This, however, depends to a large extent on the proper construction of examination questions. This is by no means a simple task and careful preparation of the staff is often needed to arrive at a number of useful questions. Such an "item bank" of examination questions is a valuable possession for each faculty and forms the basis of objective, valid and reliable examinations. The scope and content of the "item bank" must be decided upon by a committee, i.e. by all staff who were involved in the teaching of the subjects.

Further, there must be agreement on whether the question is supposed to measure simple recall, manual skills, understanding, or other more complex forms of knowledge such as problem solving or analysis of a given situation.

There are various ways of orienting the faculty in the preparation of examination questions and of making sure that there is a fair distribution of questions covering the whole field supposed to be known by the student.

The attached and modified chart is an example used by the International Association for Research in Medical Education (Arbeitsgruppe für Ausbildungsforschung), Munich, during a recent meeting of a medical faculty. The meeting convened in order to become familiar with item construction, i.e. the design of examination questions. When considering the columns of the chart an item stem such as endocrinology (i.e. diabetes) is not enough and the method of diagnostic procedures as well as the co-operation needed between the diabetic patient, the nurse and the physician (patient management) might be incorporated into the question. Further, the level of comprehension of the student, which is reflected in the taxonomy, i.e. classification of educational objectives, has to be assessed. The question must also

be relevant to the educational objectives of the faculty. Detailed examinations about rare diseases hardly seen in the community in which the majority of the young graduates are going to work are of lesser value than examination of knowledge and understanding of diseases which are widespread in the country.

The difficulties in evaluating the examinations can be reduced if certain rules are set for the task. Some of these rules are:

1. Number of correct answers per student as calculated for the whole examination or for parts thereof.

2. Percentage of correct answers per question.

3. Percentage of wrong answers per question.

4. Frequency distribution of correct answers for the whole examination and parts thereof.

5. Standard deviation as a measure of the variation of scores among individuals.

6. Standard errors of the difference of the means obtained on two measures of the group.

7. Reliability co-efficient of the examination as an estimate of the extent to which an examination is consistent in its measurement.

The importance of objective evaluation for the general goals of instruction can only be stressed briefly, and some points are mentioned below:

1. Proper evaluation of academic achievement changes the learning pattern of students so as to encourage their initiative and responsibility.

2. The tensions during the study period are decreased, i.e. the teacher is no longer mentor and examiner and the student knows that his academic achievement is judged objectively by evaluation committees and not by his teacher.

3. The development of checks during any stage of the student's education show him where he stands and what remains to be achieved.

4. The objective evaluation of academic achievements of the group or class informs (and evaluates) the staff about their success or failure as mentors or teachers.

There is no doubt that academic achievements can be measured scientifically, and medical schools are increasingly adopting a positive attitude towards making examinations a more reliable instrument to assess their educational objectives.

BLUEPRINT TO CONSTRUCT ITEMS

Each column I-III of the roster has to be considered for the item

	I. 1. Heart	2. Cardiovascular	3. Respiration	4. Digestive tract	5. Lymphatic system and spleen	6. Kidney and urinary tract	7. Nervous system	8. Muscular-skeletal system	9. Skin and connective tissue	10. Special senses	11. Reproductive organs	12. Endocrines		
II													1. Diagnosis without instruments and apparatus	
													2. Diagnosis with instruments and apparatus	
													3. General therapeutic measurement (patient management)	
													4. Pharmacotherapy	
													5. Radiotherapy	
													6. Obstetrics	
													7. Surgical performances	
													8. Prophylaxis	
													9. Prophylaxis	
													10. Relevance to anat., physiol., biochem. and pharmacology	For preclinical subjects
													11. Relevance to clinical cases	For preclinical subjects
													12. Relevance to pathology and microbiology	For preclinical subjects
III													1. Knowledge (recall)	Taxonomy (Various levels of comprehension of the subject)
													2. Understanding	Taxonomy (Various levels of comprehension of the subject)
													3. Problem solving	Taxonomy (Various levels of comprehension of the subject)
													4. Skills	Taxonomy (Various levels of comprehension of the subject)
													5. Analysis	Taxonomy (Various levels of comprehension of the subject)