

The Problems in Existing Evaluation System of Clinical Subjects in Medical Colleges and Recommendations for Their Rectification

Qazi Khadim Mohyuddin,*
M.B., B.S., F.R.C.S.

Khyber Medical College,
Peshawar, Pakistan

Summary

The existing evaluation system in the medical colleges with special reference to examination in surgery has been analyzed in this paper. The present written, oral and clinical examinations and their following related problems have been discussed in detail:— potential for abuse, absence of standardisation and representativeness, scores reproductivity, marks scale, validity, usefulness and cost of the examinations. In the end, recommendations have been made to improve the existing written and clinical certifying examinations so as to include the "Objective Structure Clinical Examination" (OSCE). The evaluation should be in the form of continuing, mid-term and final certifying examinations. Thus towards the end of the three years, the student should have acquired a somewhat uniform knowledge and skills; the examination should also provide a feed-back for modifying the learning programme and introducing necessary changes in the curriculum.

Introduction

At present, examination and grading in the evaluation system of clinical subjects in a medical college is supervised by the academic council for the college comprised of senior faculty members. This council is guided by rules and regulations framed by the Pakistan Medical Council: a national organization of medical educators, health service administrators and those dealing with community health service i.e. the general practitioners.

Two examinations are held to test the competencies in surgery at the end of fifth year, on the completion of three years of clinical and theoretical teach-

*Professor of Surgery, Khyber Medical College, and Consultant Surgeons, Khyber Hospital

ing of surgery. There is an annual examination which is followed after a variable period, may be upto three months, by a supplementary examination. Those who fail in the annual examination are offered another opportunity to sit in the supplementary examination for the subject. The annual examination simply scores pass and fail: it has no feed-back for those who fail in the subject to reveal their areas of weakness.

Over a period of three years, each student undergoes clerkship in various surgical units and stays in each unit for an equal period of time. The total period of the surgical clerkship comes to about 38 weeks. There is no organized system for utilizing evaluation as a tool for assessing their knowledge and skills during the three years period of surgical training.

Direct observation of the tutor during clerkship is neither recorded in an appropriate manner nor does it carry any weight in the annual certifying examination. Hence many of the students fail to attend to their duties in the wards. Central office of the college does not have a detailed record of the progress of the students during the three years of clinical teaching. Moreover as both the faculty and the students tend to ignore concentrating on the specific objectives to be achieved in various skills, hence there is no persistent effort to design a valid clinical curriculum for the students. Lack of data on the progress of the students and the teaching process during the three years period is the most obvious end result.

By the end of three years of clinical teaching, there is an accumulation of a bunch of students who vary in gross proportions with regard to their knowledge and clinical competencies. Even if a minimum of rigid criteria are applied vigorously in the assessment, there will be a massive failure rate. Therefore examination, originally a criterion-based test, has to be transformed into a norm reference test in order to avoid massive failure rate.

The object of this paper is to analyze the existing evaluation system in the medical colleges with special reference to the examination in surgery. As the examinations in various clinical subjects are conducted on the same lines, therefore, it is assumed that the problems identified in this subject also apply to all the clinical subjects. We have tried to identify the weaknesses of the current evaluation system and recommended certain immediate changes to rectify the same. Unless and until adequate attention is given to the problems of existing evaluation system, the standard of medical education will continue to deteriorate.

Such deterioration, apart from producing inefficient medical practitioners, will also be a source of exaggerated tension in the medical colleges, creating further problems both at the college and sometimes at the provincial and national level.

Present Examination and Marking System

The rules, under which the faculty of the college operates, require that the marks be recorded for each student on the basis of an examination at the end of the completion of the syllabus designed for the academic period with a provision of preparatory leave between the end of the course and the examination. This period varies but is always round about one month. During this period, the students do not perform any clinical tasks. They just close themselves in their rooms and cram books. Some of them even try to memorize textbooks by heart.

Furthermore, separate marks are recorded for the written, oral and clinical examinations in surgery which are scored by a panel of examiners. The names of these examiners are recommended by the head of the department of surgery to the lean of faculty who in turn communicates this list of the controller of examination of the University. Marks from these three types of examinations are combined to yield a total of marks for each student. While the total marks scored in any subject determine the position of a student in that subject, the sum-total of marks scored in all the subjects determines the position of the student in the class. This rank or standing in the class is used to identify the best graduates who will be offered posts of demonstrators in basic subjects and registrars in clinical subjects or other choice position. The marks, therefore, assume tremendous importance not only in the decision as to whether or not an individual shall be promoted or graduated but also in determining the career opportunities of that individual.

A natural corollary is that those who control the scoring of these marks i.e. the examiners, assume tremendous importance in the life and thoughts of the students. Hence the students would prefer to learn what the teacher knows and would all the time depend on him, rather than acquire and develop a self-learning drive.

Nature of Examination

1. WRITTEN EXAMINATION in the syllabus of surgery is comprised of 12 essay (traditional) type questions. The student has a choice to attempt any 8 questions. As he can omit any four questions, therefore, from the very start a part of syllabus is omitted deliberately. These questions cannot sample ade-

quately the content of knowledge to be tested. Moreover each of these essay questions require a lengthy answer. The student prefers that the questions be framed in a way in which he has crammed from the book. If a question is designed so that it requires the student to rearrange his knowledge of facts and do some analysis to answer it, then it is very likely to result in an uproar amongst the students.

The fault is not all that of the student. He has been permitted to learn in that particular way and hence would like to answer in his own way and would prefer that he is questioned in the same manner.

2. ORAL EXAMINATION in surgery consists of 3-4 rapid-fire questions, composed on the spot by the examiner on gross and microscopic surgical pathology, operative surgery, X-rays and surgical instruments. These questions differ from one examiner to another examiner and from one student to another student, both with regard to their contents and level of difficulty.

3. CLINICAL EXAMINATION generally consists of certain tasks. These comprise one long and two short cases. This is a face-to-face interaction between the student and the examiner. The long cases allotted to the students are different for every student. This difference applies to :-

1. Nature of the clinical problems;
2. Language which the patient speaks and
3. Cooperation of the patient.

Clinical problems vary from a simple case of hernia to difficult problems of chronic pancreatitis. It may not matter if a good student is examined on difficult clinical problems but it may result in a great difference if an average or below average student gets difficult clinical problems.

On the long case, the student discusses all the problems with the examiner. He may spend about 2 to 3 hours with the patient but the examination itself lasts for not more than 15 minutes. The examiner does not observe the student while he takes the history. Moreover he may or may not go through the history written by the student and may or may not observe the student elicit a sign on the patient.

On short cases, the student is to take a quick history and ask a few rapid-fire questions to reach a diagnosis. He may have to demonstrate some of the signs on the patient. On short cases, the student is tested as to whether he can gene-

rate an hypothesis on sight of the clinical problem and whether he can limit himself to only relevant and pertinent questions and can go on to quick clinical assessment and reach a plausible hypothesis. This type of examination is to test the degree of his long term memory and hence the extent of clinical experience that he has acquired in three years.

Preparation, administration and marking of all the examinations is the responsibility of senior faculty members. The questions for WRITTEN EXAMINATION are not reviewed by any body of experts. In fact there is no national panel of experts in any discipline to assure comparability of standards from one faculty to another.

In practice the written examination questions are composed by four examiners. Half of the questions are composed by two external examiners which are subject to moderation by the internal examiners. These questions are typed, duplicated, packed and sealed before one of the internal examiners. The question papers under sealed covers are taken to the examination hall five minutes before the beginning of the examination and opened by the supervisor of the examination hall who checks the intactness of the seal.

The whole examination, the compilation and the declaration of the results is done under the supervision of the controller of examination of the concerned University. The task of marking the examination papers is undertaken by the senior members of the faculty.

There is no advanced discussion on model answer or of number of marks to be deducted for various types of deviations from such a model. Half of the answers are marked by a senior member of the same discipline from an outside faculty who has been responsible for designing the questions.

The ORAL AND CLINICAL EXAMINATIONS are administered by senior members of the faculty and the external examiners mentioned above. Each examiner determines independently the questions that he will ask from any student and the way in which his/her response will be scored.

There is no attempt to create a uniformity in the questions asked from the student or the way they would be scored. There is little attempt to train the examiners in constructing, administrating and marking of examinations.

The response of external examiners to participate in the examination is also very uncertain. Usually at the eleventh hour, by a simple telephonic message.

the external examiner excuses himself from attending the examination. Under such circumstances, the head of the department of the respective subject has to make local arrangements in haste and may have to bring in examiners who have no previous experience of conducting an examination at all. The usual cause for the external examiners not to participate in examination is the length of examination-time and hence their inability to leave their lucrative practice for the sake of their educational responsibilities.

Feed-back to the students and the faculty on any examination is usually limited to the marks scored. The scores of the student who fails in the examination, are neither discussed in the faculty nor given to the student. Moreover these scores are such that they do not reveal specific weaknesses, hence their feed-back value is nil. On the whole the feed-back value of these examinations to the student as well as to the faculty to modify the curriculum is not there.

Problems in the Present System

There are about 300 students to be tested in a certifying examination. With this large number of students, the investment of faculty time in the administration and marking of examinations is exorbitant.

The interesting feature is that the marking and administration of these examinations is carried out by the senior members of the faculty who are also occupied with patient-care and private practice. The junior members of the faculty, who will be taking over the responsibilities of tomorrow, are totally deprived of having any experience of the examination. Moreover the time spent in the administration and marking the examination is, therefore, not available for instruction to the students. This enormous investment yields a single out-come which is ranking of the students for the purpose of subsequent appointments.

The data collected from examination is so vague that its feed-back value is almost nothing for evaluating the curriculum or realizing weaknesses of the students and directing their further learning. Indeed given the nature of examinations and the way in which they are scored, it would be extremely difficult to extract diagnostic information from the results that would be helpful in these ways, nor does the system yield information which is of any practical value to the faculty in evaluating curriculum and in modifying programmes to make them more effective. Detailed data on what the students know and what they do not know, or what they can and cannot do at a satisfactory level are simply not available from this system.

The natural outcome is that each faculty member is vaguely aware of the weakness of the graduate or undergraduate due to lack of reliability and validity

of examinations; he does not have either the necessary data, adequate know-how or an organizational capacity to rectify the problem. Furthermore serious questions may be raised about the validity and the reliability of whatever little data is generated in the process of the existing examinations because :-

WRITTEN EXAMINATIONS (1) do not sample the whole spectrum of knowledge to be tested, (2) measure only the recalls of limited amount of information and (3) are neither reliable nor valid in the present format.

In the present format, written examinations do not measure the competence of the student to gather data, interpret data, reach a probable hypothesis and advise appropriate therapy and hence the written examinations lack validity. A valid evaluation must assess the above mentioned competencies which are the main objectives for a competent medical graduate.

As there are no standard answers prepared for the questions before-hand for reference, therefore, reproducibility of scores is impossible.

ORAL EXAMINATIONS in their present format as currently employed, measure but a single aspect of the competence namely the student's ability to recall and reproduce under stress, the detailed fragments of factual information. Such tests lack both in validity and reliability. Orals and clinicals are never preceded by a pre-test meeting to devise the questions to be asked from the student in the light of the objectives set forth, if any, by the faculty. The questions differ from one examiner to another examiner and from one student to another student both with regard to their contents and level of difficulty.

The questions asked are never structured and designed to be valid. As the answers are marked subjectively, therefore, temperament of the examiners assumes tremendous importance. Yet temperament of the examiners varies from day to day and from the beginning of the examination to the end of day. The degree of anxiety in the students is so high at times, that he would have performed much better in a different environment.

In view of subjective influences in the scoring of answers, variability of questions asked from different students and high level of anxiety in the students, the reliability of the orals is highly questionable.

CLINICAL EXAMINATIONS fail to evaluate skills like :-

1. History taking :
 - a. Interview technique,
 - b. Communication skills and
 - c. Interpersonnel skills;

2. Total physical examination techniques: as on a long case, only a component of such skills is demonstrated;
3. Identification of the patient's problem from the information obtained and differential diagnosis: as the students on the long case may have had feeding from interneers;
4. Technical skill with diagnostic tools;
5. Interpreting the results of investigations;
6. Use of non-patient resources;
7. Patient counselling and
8. Recommending and undertaking appropriate management including patient education.

Many of these abilities are to a greater or lesser extent ignored in the clinicals in their present format. Ultimately the confrontation between the examiner and the examinee in clinicals, in their present format, degenerates into a theoretical discussion.

A further problem is the choice of the patients. These are frequently atypical of the problems likely to be encountered in practice. They usually represent cold rather than acute or emergency situations and some of the sub-specialities are under-represented.

Because in their present form, the clinicals have very little productive validity, therefore, the best graduate may not necessarily turn out to be the best doctor with regard to his future performance.

RELIABILITY OF CLINICALS

In their present format, the clinicals are not an objective assessment of skills. The score in most cases is not consistent or even accurate. Marks awarded by one examiner often vary considerably from those awarded by another examiner often for the same performance. Moreover the same examiner may vary in the scores he awards on different occasions.

Another source of unreliability is that different students are examined on different patients. The patients vary with regard to complexities in their problems. In addition, one may come across a patient who plays havoc with the examination by helping some students and obstructing others. Each student is asked to answer a different question which again varies both with regard to its

complexity and harshness. In a reliable clinical assessment, the variability in the patient and the examiner must be reduced to a minimum.

In addition to the usual sources of error in marking essays, orals and clinicals, there is a special problem of language. Instructions and examinations are conducted in English which is a foreign language and facility with that language varies significantly amongst the students. In many cases the student's command of English is so poor as to constitute a serious obstacle to his/her understanding exactly what question is being asked. Even when a question has been clearly understood, some students have such a poor command of English that it is almost impossible to determine what they are trying to say in their answers. Thus extreme subjectivity and serious bias are constant hazards in the marking process.

Summary of Problems in Existing Evaluation System

1. *Potential for Abuse :*
 - a. In the form of feeding from internee (house surgeon);
 - b. Patient may not co-operate with a student;
 - c. Resident registrar in the ward may assign difficult cases to a particular student whom he may not like, creating a state of high tension amongst the students prior to the actual examination.
2. *Absence of standardization of the examination :* No student ever faces exactly the same examination. Different examiners ask different questions and even the same examiner asks different questions from different students, or different students get different patients.
3. *Representativeness of the examination :* In oral and clinical examination, the discussion is limited to one or two jars, or one or two x-rays. Clearly one cannot generalize about the student's knowledge of the field from his/her discussion of one or two topics. Written examination samples only a fraction of the content of knowledge.
4. *Reproducibility of Scores :* In this type of examination, there are basically three sources of error that jeopardize reliability of the marks :—
 - a. First the examination is not standardized, different students get different questions.
 - b. The examination has too narrow a base for generalization.
 - c. The examiners differ greatly amongst themselves: each has a unique set of interests, standards and requirements.

5. *Scale used in assigning marks in the examination (0-100)*: It is very difficult to make distinction between 50-55-56 marks. Such difference in marks is crucial to the students as these marks go with the student throughout his carrier. The person with 56 marks will always have a preference over the one with 55 marks.

6. *Validity of the examination*: There is no evidence that in written, oral or clinical examination, in their present format, the students are being tested on any of the observational, interpretational or technical skills essential in the competent practice of medicine.

7. *Usefulness of the examination results*: The data gathered from evaluation should be a gold mine of useful information. But the present evaluation system only yields marks which in themselves are highly unreliable and lack validity. The results in the present format do not have any feed-back value for the students or the faculty.

8. *Cost of the examination*: Cost both in terms of stress and time are exorbitant. In the present system of evaluation the student is under stress, the examiners are under stress and the hospital staff is under stress without any useful information being derived from the system. Moreover the tension, that is built up during the process of examination, reaches its height when results are declared: a point when disturbances usually develop.

Recommendations

In considering possible modifications in the system, it is important to recognize that the present faculty/student ratio is likely to continue for a long time. Moreover the rules under which the faculty operates to construct and conduct examinations are framed by Pakistan Medical Council and the respective Universities, therefore, the faculty itself cannot be expected to bring about drastic changes to modify the system of examinations unless concurrence on basic changes in the rules can be obtained.

Nevertheless it is important that in the subject of surgery, specific learning objectives must be clearly identified alongwith specific skills which each student must master during the course of instructions for each year during the three years of clinical teaching. Attempts should be made to define the details of curriculum both in breadth and depth in the light of those goals for each academic year. Very briefly the objectives in clinical achievements would be

In the 3rd year —

1. The student should have acquired the knowledge of common vocabulary in use while addressing the patients to gather data;
2. He should have mastered the art of data gathering and
3. He should develop mastery of interview technique, communication skills and interpersonal skills. He should be continuously developing the skill of integrating patho-physiological concepts with the clinical phenomena.

By the end of the 4th year —

- 1 He should have developed mastery of interpreting clinical data;
2. He should be able to utilize various resources to obtain required data;
3. He should recognize the required data and
4. He should formulate a hypothesis.

By the end of the 5th year —

- 1 He should have mastered problem-solving;
2. He should be able to formulate plausible scheme of therapy;
3. He should be able to detect complications and recognize urgency and seriousness of the situation;
4. He should be able to recognize primary disorder;
5. He should be able to recognize underlying and associated disorders and
6. He should recognize the possibility of other disorders and perform check-up to rule out such disorders.

IMMEDIATE NEXT STEPS

— IN WRITTEN EXAMINATION

With respect to written examination, it is recommended that the brief unstructured essay questions requiring lengthy responses should be gradually abolished. An increasing number of specific highly structured essay questions with MCQs and other objective forms of problem-solving exercises be introduced in the non-university examinations which should be held at the end of 3rd and 4th years.

As both the faculty and students gain experience with these newer forms of exercises, the proportion of written examinations in objective format be gradually expanded and ultimately may take the place of essays in the final certifying

examination. This will enable adequate sampling, and realize validity and objectivity in written examinations. To make it more reliable and objective, for each essay (short or long), detailed model answers be prepared in advance of the examination to guide the persons marking the examination.

With these model answers to guide them, the responsibility for marking examination be increasingly shifted to junior faculty members who, now by virtue of being involved in examination, will continue to gain experience and may introduce new ideas. Prior to constructing questions for written examination, assistance of individuals nominated by the faculty must be solicited. These individuals who have the requisite experience should help in modifying, rejecting or selecting questions to tailor the test for evaluating the requisite skills objectively.

— IN CLINICAL EXAMINATION

Each department moves as rapidly as possible to identify the specific skills which each student should master in the course of instructions in the department: mastery in these skills be evaluated by a process of continuous assessment. To implement this procedure, each student should be provided with a skills — record book, naming each of the requisite skills. Each student, as he or she progresses through the course, has a responsibility for notifying the professor of the unit of his/her readiness to be tested in the skill. Upon such notification, the professor authorizes associate/assistant professor to observe the student perform the task and at his/her satisfactory completion, the associate/assistant professor records by initialling the appropriate line in the student's skills-book, the fact that the student has demonstrated satisfactory proficiency. By the end of the course, each student is expected to have demonstrated proficiency in all the requisite skills, and the associate/assistant professor should be responsible for documenting this fact. This continuing evaluation of clinical skills should be utilized in strengthening the weakness of the students and the curriculum. Quality control should be maintained by holding the associate/assistant professor accountable for any student whose proficiency he has documented. No student, who has not mastered all the skills listed in the objectives, should be permitted to sit in the annual examination or the final certifying examination. It is recommended that separate orals should be eliminated as rapidly as possible. Those aspects of oral examination, designed to assess recall, simple interpretation and problem-solving, should be incorporated in the new types of objective-type questions in the written examination. All other parts of oral examination should be integrated with clinicals.

— IN CLINICALS IN CERTIFYING EXAMINATION

Clinicals in certifying examination should consist of :-

1. Examination on long case;
2. Examination on two short cases and
3. An OSCE as devised in Dundee.

Examination on long case should provide evaluation in problem-solving skills in continuity. Factors which are the source of undermining the reliability and validity outlined above have to be reduced to a minimum by adapting strict disciplinary measures and a good deal of effort in the planning and execution of the examination, so that it tests the requisite objectives reliably. In this regard problems presented to the examinee for his evaluation should be uniform with regard to their nature and complexities upto maximal possible extent. Such a uniformity can be achieved by resorting to the use of simulated patient. Validity can be enhanced by representing problems like acute cases and emergencies on simulated patients. Moreover sub-specialities should be fairly represented.

Examination on short case is the best way of evaluating the clinical experience that the student has acquired during the three years period. It tests his long-term memory in clinical experience. This type of test also evaluates his ability to avoid unnecessary questions and his ability to reach a hypothesis with minimum relevant data.

Efforts should be made to make the tests reliable by arranging a pre-test meeting in the morning to create uniformity in questioning and scoring. I cannot provide a data, but I think examination on short cases should be highly valid for evaluating and discriminating those who have acquired clinical skills from those who have not.

OBJECTIVE STRUCTURE CLINICAL EXAMINATION

Objective Structure Clinical Examination (OSCE) is a method of assessing student's clinical competence which is objective rather than subjective and in which the areas to be tested are carefully planned by the examiners. The clinical competence to be tested is broken down into its various components e.g. taking history, interpretation of an X-ray or coming to a conclusion on the basis of the findings. Each component is assessed in turn and it is the objective of one station in the examination. The student during the examination rotates around 20 stations. For each component of a skill, there are two stations. On one station he performs an examination and at the subsequent station, he answers questions

which are designed on MCQ pattern on the basis of information that he has acquired on the preceding station.

OSCE provides a more valid examination than the traditional clinical examination. The examiners can decide in advance what is to be tested and can then design the examination to test these competencies. The examiner can control not only the contents but also the complexities of the examination: more straightforward cases for junior and more advanced cases for senior students. The emphasis can be moved away from testing the factual knowledge to testing a wide range of skills including history taking. Minor specialities can be included. The examination is also more reliable. The variables of the examiner and the patient are to a large extent removed. The use of a check list on the examination station and the use of MCQs on the answering station result in a more objective examination. A further advantage of OSCE is that with the number of stations, a larger sample of the student's skills is tested.

Finally the OSCE has the advantage that it can be used with a large number of students. With the OSCE, the criteria for a pass can be specified in advance and following the examination, feed-back can be given to the staff and the student.

Number and types of evaluations recommended

Following types of evaluations should be incorporated :-

1. Continuing evaluation.
2. Mid-term evaluation.
3. Final certifying evaluation.

1. *CONTINUING EVALUATION.* This type of evaluation has been ignored upto now. Continuing evaluation should be utilized to detect the weaknesses and strengths of the student and the curriculum, and help in modifying the learning techniques and identifying priorities on the part of tutor and student. This should be achieved in two ways :-

a. Direct observation by the tutor: such observation should be recorded in an objective manner during the clerkship on specially devised check-list and the various clinical competencies rated. These check-lists should be designed for each of 3 years of clinical teaching so that only those competencies are evaluated which would be the objective for that year. The person responsible for checking the competence should provide comments to support his rating and also explain under general comments the methods he recommends to correct the deficiencies. Direct observations should concentrate mainly to evaluate the prerogative achievements.

b. Skills-record book: satisfactory completion of this book guarantees that the student has mastered all the skills listed in the objectives for each year.

2. *MID-TERM EVALUATION.*

a. At the end of the 3rd year, a paper and pencil examination, MCQs, structured essay and data gathering question should be utilized to evaluate objectives set for the period of one year i.e. 3rd year.

b. At the end of 4th year, a paper and pencil examination, MCQs and structured essays to evaluate the objectives already set for the two years i.e. 3rd and 4th years.

c. An objective structured clinical examination should be utilized to test the clinical skill acquired during these two years such as :-

1. History-taking.
2. Clinical examination.
3. Interpretation of the clinical data.
4. Gathering laboratory and radiological data.
5. Interpretation of the acquired data.

3. *FINAL CERTIFYING EVALUATION.* To be structured on lines outlined above at the end of the final year.

End Result of Evaluation

At the end of each evaluation, the results should be analyzed by the faculty members. In the light of these results, the curriculum should be evaluated and necessary changes introduced. The use of skills-card will guarantee that each student has performed and mastered all the skills outlined in the objectives for each year. Results of direct observation by the tutor should be discussed with the student and his weakness explained to him. During his clerkship in the next ward, special emphasis should be laid on correcting these weaknesses. In this way the efforts of tutor and student can be observed objectively towards the process of rectifying the weaknesses of the whole teaching and evaluating programme. By utilizing continuous and mid-term evaluations towards formative objectives, it is hoped that towards the end of three years, the student would have acquired a somewhat uniform knowledge and skills.

In this situation, it would be easy to identify criteria and modify such criteria if necessary for certifying examination according to the performance record during the three years of learning. This will enable even this examination to provide a feed-back for modifying learning programme and rectifying the weaknesses of those who fail to qualify.