Fundamentals of Students Assessment and its Dimensions in Medical Education

Asaduzzaman AKM[1], Akhiruzzaman[2]

Abstract

Assessment is the measure of students' performance or it is the process of testing a student's ability. During all phases of medical education & practice, medical students and physicians are constantly being assessed for multiple purposes. In medical school, assessment should guide and promote learning by providing feedback to the student on how well progress is being made, assure the predetermined competence of the graduates produced by the school and help in developing a habit for continued learning. It is true that most teachers employ assessment as one of the most important motivators of medical students. Assessment is necessary at all the stages of a medical student's life. From selection into medical schools, yet it was only during the past three decades that assessment in medical school came to be accepted and studied as a rational and useful application in educational science. The international, national & institutional centers of medical educational science & research have played a very significant role in this process of change. Therefore, it is now the desire of the most medical teachers to hear from the enlightened teachers views about numerous technical aspects of assessment, observe them constructing higher quality test instruments and in general, spend more time attempting to improve the assessment techniques that they wish to use. As a consequence it is natural that assessment instruments that are being used in medical schools have improved considerably & medical teacher have developed greater expertise in using more and more valid and reliable assessment procedures.

This article highlights the functional requirements of all assessments, assessment scheme, evaluation and the application of assessment findings, four simple questions on quality improvement in assessment, blooms taxonomy, authentic assessment and its difference with traditional assessment, The article also indicates the domain based tests & types /assessment tools, criteria / qualities of a good test, test construction, possible barriers and clues in test items, including five vital questions needs to be considered in relation to assessment as well.

Keyword: Assessment, Medical Education, test blueprint, Validity, Reliability, Objectivity

Introduction

"Students can, with difficulty, escape from the effects of poor teaching; however, cannot escape the effects of poor assessment (if they want to graduate)".[1] "Changing the examination system without changing the curriculum had a much more profound impact upon the nature of learning than changing curriculum without altering examination system".[2]

The principle objective of designing a system of student assessment in medical schools/ college is to make the assessment procedure congruent with educational goals and instructional principles. However, the traditional examinations fail to serve this purpose adequately. Students mainly studied the prescribed examination content at the expense of other desirable areas of knowledge. Farthermore, most of the learning activity was concentrated in the period prior the examinations. Different authors, including great reformers in medical education in early part of the century, emphasized that students learning was largely influenced by examinations.[3,4] Traditional examinations have also been shown as a barrier in the development of intellectual attributes demands by medical schools.[5] Content analysis, oral and multiple choice exams in deferent medical schools shown that the abilities assessed pertained to lower intellectual levels.[6,7]

Functional requirements of all assessment in medical education are to fulfill:
1. Certification of competence and safety.
2. Provision of feedback of students and staff (Teachers).
3. Monitoring curriculum and teaching.
4. Steering of students learning.[8]

Assessment scheme should:

- Be congruent with the institute philosophy.
- Be comprehensive.
- Include formative and summative examinations and
- Consist of instruments which are valid, reliable objective and feasible.

1. Prof. Dr. AKM Asaduzzaman
   Professor, Department of Community Medicine
   Diabetic Association Medical College, Faridpur.

2. Dr. Akhiruzzaman
   Assistant Professor & Head, Department of Community Medicine
   Diabetic Association Medical College, Faridpur

Correspondence to:
Prof. Dr. AKM Asaduzzaman
Professor of Community Medicine
Diabetic Association Medical College, Faridpur.
E.mail: asaduzzamancme@gmail.com

A scheme of assessment could be developed for any educational program according to the curriculum implemented in the institution. The examinations represent a sampling of achievement of objectives delineated for the particular program at the specific level, relating to any of the areas of competencies like knowledge, practical and clinical skills, problem solving and judgment, procedural skills, self-directed learning skills and personal qualities.

Traditional examination in medical education were mainly focused on knowledge and practical and clinical skills with little emphasis on procedural skills. Abilities to problem solving and judgment, self-directed learning skills and personal qualities have rarely been considered in the assessment scheme. During the past decades a few of the innovative medical schools have included above mentioned skills in their assessment scheme.[9]

Assessment is a fact finding activity that described the condition that exist at a particular time.[10]

Evaluation and the application of assessment findings like-
1. Collecting information- qualitative or quantitative (Measurement, Assessment, Examination, and test).
2. Judging the value of information.
3. Returning some information to its source (Feedback)
4. Recommending a remedial action.[10]

In terms of assessment(and quality improvement), we need to ask oursef four simple questions:
1. How do I define, a successful (competent) learner?
2. What is the evidence that he/she meets this definition?
3. Am I satisfied with the results?
4. If not, why not and what will I do about it?

In response to first two questions different types of constructed (supply) response items can be used and in case next two questions long essay and then again, briefly, as part of assessment cycle or quality improvement cycle usually used.[12] Blooms taxonomy refers to a classification of the different objectives that educators set for learners. This taxonomy divides educational objectives into three domains as shown in Table-1 below:

Table 1: Educational objectives according to domains

Domain	Area	Aspect
Cognitive	Knowledge	Knowing/head
Psychomotor	Skill	Doing/hands
Affective	Attitude, behavior	Feeling/heart

In addition to these domains, within the cognitive domain, Bloom and his colleagues also identified levels of thinking (cognitive). In designing assessments concerning cognitive domain, the bloom's taxonomy of hierarchical levels of thinking is shown in the diagram below.[13]
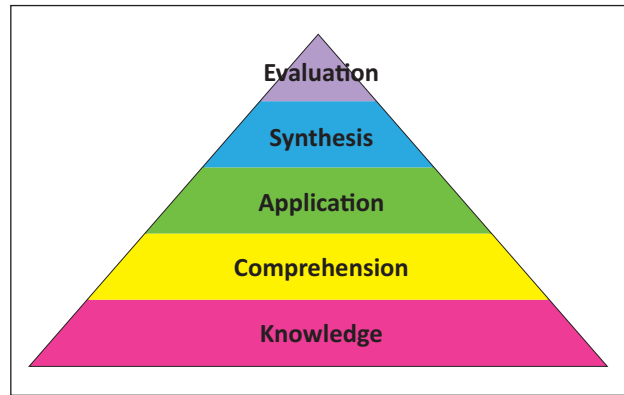

Figure 1: Bloom's taxonomy of hierarchical levels of thinking

Alongside bloom's taxonomy the physician George Miller developed is now called as Millers pyramid[14] which depicts what a learner is expected to do (at four levels of assessment) as he/she moves towards becoming a professional:
- Knows (Knowledge)
- Knows how (application of knowledge)
- Shows how (competence) and
- Does (performance)

Over the years, people have modified Millers original diagram by adding types of assessment that can measure what is expected of the learner at different levels. The diagram is presented below.
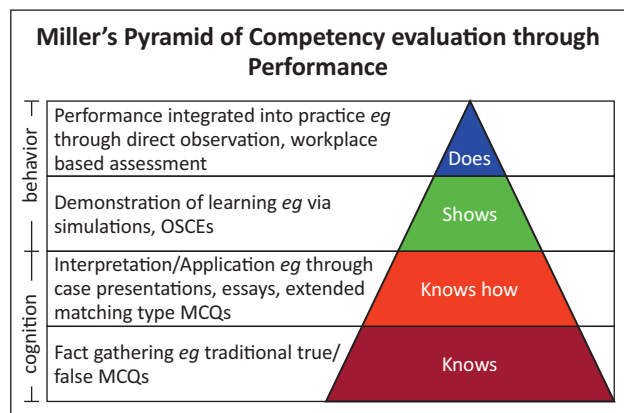

Figure 2: Millers pyramid

The pyramid is context rich. It tells us a lot of about what happens as a learner moves from a novice to a doctor or health professional in clinical practice. From the diagram it is seen that:
- The three domain of learning (knowledge, skills and attitudes) from part of professional development from novice to expert.
- In moving from novice to expert, the focus of assessment moves from knowledge based-assessment

(knows) through interpretation and application of knowledge (knows how) to demonstration of learning (shows how).

- A doctor "does" in the workplace (i.e. hospital or clinic). A mini-CEX for instance, is not at the top of the pyramid. Patient outcome might be:
  - ✓ The emphasis changes from the cognition to behavior and affect. This is where issues of professionalism become more important.
  - ✓ This does not mean that knowledge is not important. It is the foundation of being a health care professional.

Authentic Assessment

In deciding on and designing assessment we often strive to make the assessment as authentic or realistic as possible, hence the term "authentic assessment". It has four requirements: [15]

1. It samples the actual knowledge, skills and dispositions (attitude, behaviors) desired.

2. It requires the integration of multiple kinds of knowledge and skills as they are used in practice (the workplace)

3. It represents multiple sources of evidence collected over time and in diverse contexts.

4. Assessment evidence is judged by individuals with relevant expertise against criteria that matter for performance in the field or work place. The differences between traditional and authentic assessment is shown below (Table 2).

Table 2: Differences between traditional and authentic assessment

Traditional assessment	Authentic assessment
Selecting a response	Performing a task
Recognition and recall	Application of knowledge, construction of new knowledge
Indirect evidence of learning	Direct evidence of learning
Contrived	Real life

Test & types/Assessment tools

A test is an instrument or systematic procedure for measuring a sample of behavior.

Types of tests: [16]

1. ***Informal versus Standardized tests*:** Informal test are those instructed by classroom teachers, whereas those designed by the test specialists and administered, scored and interpreted under standard conditions are called standardized tests.

2. ***Individual versus Group tests*:** Some tests are administered on a one-to-one basis using careful oral questioning (e.g. individual intelligence test), whereas others can be administered to a group of individuals.

3. ***Mastery versus Survey tests*:** Some achievement tests measures the degree of mastery of a limited set of specific learning outcomes, whereas others measure a pupil's general level of achievement over a broad range of outcomes. Mastery test typically use criterion reference interpretations and survey test tend to emphasize norm-referenced, interpretations but some criterion-referenced interpretation also are possible with carefully prepared survey tests.

4. ***Supply versus Selection tests*:** Some test requires examinees to supply the answer (eg- eassy test) whereas others require them to select the correct response from a given set of alternatives (eg. multiple-choice test).

5. ***Objective versus Subjective tests*:** An objective test is one on which equally competent scorers will obtain the same scores (eg- multiple choice test), whereas a subjective test is one where the scores are influenced by the opinion or judgment of the person doing the scoring (e.g- eassy test).

6. ***Verbal versus Performance tests:*** some tests require examinees to make only verbal responses (eg- vocabulary test) whereas others require some type of motor or manual response (e.g. typing, solving a maze)

7. ***Speed versus Power tests:*** A Speed test is designed to measure the number of items an individual can complete in a given time, whereas a power test is designed to measure level of performance under ample time conditions. Power test usually have the items arranged in order of increasing difficulty.

Criteria/qualities of a good test [17]

A. Validity

B. Reliability

C. Objectivity

D. Feasibility/ Practibality/ Usability

A. Validity:

It refers to the ability of the instrument (test) to measure what is supposed to measure. A careful examination of test items will indicate whether the test appears to measure the subject matter content and the mental functions that the teacher is interested in testing. However any of the following factors can prevent the test items from functioning as intended and thereby lower the validity of the interpretations from the test scores.

1. Unclear directions

2. Reading vocabulary and sentence structure too difficult

3. Inappropriate level of difficulty of the test items

4. Poorly constructed test items

5. Ambiguity

6. Test items inappropriate for the outcomes being measured

7. Inadequate item limits

8. Test too short

9. Improper arrangement of items

10. Identifiable pattern of answers

B. Reliability:[18]

It refers to the consistency of measurement that is how consistent test scores (results) are from one measurement to another. It indicated how much confidence can place in our results. The longer the spread of scores is the higher the estimate of reliability will be how high should reliability be?[18] The degree of reliability demands in our educational measures depends largely on the decision to be made. A number of factors have been shown to affect the conventional measures of reliability and those are as follows-

i. **Length of test:** In general the longer the test is the higher its reliability will be. This is because a longer test will provide a more adequate sample of behaviour being measured and the scores are apt to be less distorted by chance factors such as guessing.

ii. **Difficulty of test:** Norm reference tests that are too easy or too difficult for the group members taking it will tend to produce scores of low reliability. This is because both easy and difficult test results in a restricted spread of scores. For the easy test the scores are close together at the top end of the scale. For the difficult test scores are grouped together at the bottom end of the scale. For both however the differences among individuals are

small and tend to be unreliable. A norm reference test of ideal difficulty will permit the scores to spread out over the full range of the scale.

C. Objectivity:

It refers of to the degree to which equally competent scorers obtain the same results. The test items are of objective type (e.g. MCQ) and the resulting scores are not influenced by the assessor judgment or opinion. Tests constructed by teachers however, objectivity may play an important role in obtaining reliable measures of achievement. In essay testing the result depend on a large extent on the person doing the score. Different persons get different results and even same person may get different results at different time. The solution is not use only objective tests and to abandon all subjective tests, as this would have an adverse effect on validity and it is the most important quality of test results. A better solution is to select the assessment procedure most appropriate for the behaviour being assessed and then to make the assessment procedure as objective as possible. In the use of eassy test, for example objectivity can be increased by careful phrasing of the questions and by a standard set of rules for scoring. Such increased objectivity will contribute to greater reliability without sacrificing validity.

D. Feasibility/Practibality/Usability:

The time involved in developing, administrating, costing, scoring, interpreting and reporting a test is justified.

Test construction

Construction a table of specification or test blue print:
There are three steps involved in this procedure

1. Formulation of a test matrix

2. Selection of test items

3. Editing and reviewing the final products

Content	No. of Items	Wt	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluations
Concepts of health & Disease	2	10%	✓	✓				
Epidemiology	4	20%	✓✓	✓	✓			
Communicable disease	4	20%	✓✓	✓				✓
MCH & FP	3	15%	✓	✓	✓			
Environmental health	4	20%	✓✓	✓	✓			
Nutrition	3	15%	✓✓		✓			
Total	20	100%	50%	25%	20%			5%

Figure 3: Example of a test blue print

General suggestion for writing test items: [19,20,21]

In preparing a set of test items for a test, there are some general rules of item writing that apply to all test item types:

1. Use your test specification/test blue print as a guide to item writing

2. Write more test items than needed

3. Write test items will in advance of the testing date

4. First think the answer then write the appropriate question

5. Write each test items, so that the task to be performed is clearly defined

6. Write each test item at an appropriate reading level

7. Write each test item so that it does not provide help answering other items of the test.

8. Write each test item so that it is at the proper level of difficulty.

9. Write each test item so that the answer is one that would be agreed upon by experts.

10. Whenever a test item is revised, recheck its relevance

Possible barriers in test items:

- Ambiguous statements
- Excessive wordiness
- Difficult vocabulary
- Complex sentence structure
- Unclear instructions
- Unclear illustrative material
- Race, ethnic and sex bias.

Common clues in test items:

- Grammatical inconsistence
- Verbal associations
- Specific determiners (e.g. Always)
- Phrasing of correct responses
- Length of correct responses
- Location of correct responses[19]

One of the most important responsibilities is to find out how much students have learned and assessment is the process to find out that.[19] Assessment of students is done to check whether students will be able to do the job reasonably well. Assessment is an important part of any education and it is much more important how this assessment system is being used. Assessment is a matter and it has a powerful influence over learning.[20]

A good assessment has certain characteristics & several important issues. It encourages the students to work harder. Assessment can be used to guide students about which topics & skills they need to learn more. Assessment can also guide teacher about which part of the course has been successfully & which part need to be improved.[19]

A medical teacher should answer 05 (five) questions in relation to assessment the question are:

1. What should be assessed?

2. Why it should be assessed?

3. When it should be assessed?

4. How it should be assessed?

5. Who should carry out the assessment?

When an examiner is going to assess student he should think about above mentioned questions in mind. A number of factors affecting learning, i.e. the quality of teaching, approaches to learning, availability and access to resources etc. The most powerful single influence on learning is probably the assessment system, which is used.[22,23]

Assessment is not only a measure of performance but also provides an indications of the effectiveness of teaching situation and also the effectiveness of teaching situation and also the appropriateness of content input. However, evaluation- Systematic collection, analyzing and interpreting information to determine the extent to which students are achieving instructional objectives. Evaluation implies finding out how students & others see teaching performance. It comprises collection of information, interpretation of result for future action to improve teaching.[24]

In answering question no-1: what should be assessed?

The answer is simply according to objective of the instruction which is divided into three educational domain eg- cognitive-intellectual process, psychomotor-skills and affective- attitudinal.
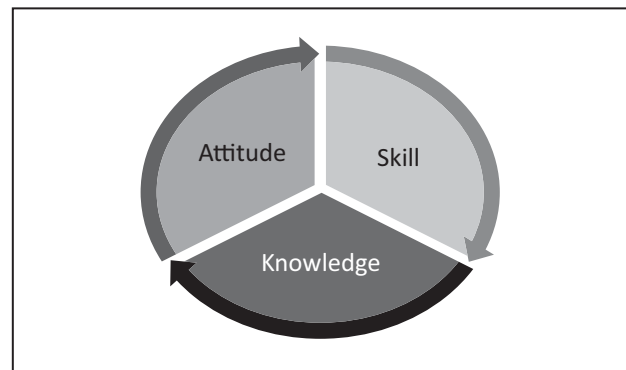
Assess What?

Figure 5: Educational domains

1. Knowledge

2. Attitude

3. Skill

Cognitive domain has six stages which is known as Bloom's Taxonomy-

- Knowledge

- Comprehension
- Application
- Analysis
- Synthesis
- Evaluation

Major categories in psychomotor domain's-
Perception, Set, Guided response, Mechanism, Complex overt response, Adaptation, Origination.

Major categories in affective domain's-
Receiving, Responding, Valuing, Organization, Characterization by a value or a value complex.

In answering the question no- 2: Why assess?
- To classify student achievement
- To motive student learning
- To provide feedback to student
- To provide feedback to teacher

In answering the question no-3: when it should be assessed?
Assessment can be classified in a variety of different ways and many of them are reasonable. One useful classification in assessment types is as diagnostic, formative and summative.[25]

Diagnostic assessment:
An assessment of trainees relevant to and before exposed to a particular educational intervention is generally referred to as diagnostic assessment. Analogas to medicine, the purpose of these assessments is to determine the trainee's educational needs with the goal of optimizing learning. Often they produce a profile which identifies the area of strength and weakness. Example of this type of test is most common in continuing medical education.

Formative assessment:
Assessment that take place for the participants during the course of an instruction.
Functions of formative assessment:
- To aid the learning of course material
- To provide continuous feedback to students and teacher concerning learning success or failures.
- It is used to monitor learning progress during instruction.
- To influence student learning.
- To encourage interaction between teachers & students.
- To diagnose student weakness.

Summative assessment:
An assessment typically comes in at the end of term or course. Its purpose is to determine the extent to which the instructional objectives have been achieved and rank of the students, awards etc. by judging essential skills, knowledge and attitude.[20]

A score is a letter or number that reflects how well or individual performs on an assessment. When a test is being develop, one of the first decisions to be made is how the scores will be interpreted. According to score interpretation type of assessments are:[26]
- Norm reference assessment
- Criterion reference assessment

Norm reference assessment:
Here the purpose is to distinguished between individuals to rank those sitting the examination into some sort of order. So, a test is designed to provide a measure of performance that is interpretable in terms of an individuals relative standing in some known group.

Criterion reference assessment:
Is to determine whether the students have achieved specific standard of competence or not- the criterion. ie the aim is to separate the sheep from the goats. So a test designed to provide a measure of performance that is interpretable in term of a clearly defined and delinited domain of learning tasks. It is the construction and use of the tests where difference can be noted. A key feature in constructing norm-referenced tests is the selection of items of average difficulty and elimination of items that all pupils' are likely to answer correctly. This procedure provides a wide spread of scores so that discrimination among pupils' at various levels of achievement can be made. This is usefully for decision based on relative achievement such as selection, grouping relative grading. In contrast, a key feature in constructing criterion referenced tests in the selection of items that are directly relevant to the learning outcomes to be measured, without regard to the power of item to discriminate among pupils. If learning tasks are easy then the test items will be easy. No attempt to be eliminate items alter their difficulty. Here the main purpose is to obtain a description of the specific knowledge and skills each pupil demonstrate.

The next question no. 4 is how it should be assessed?
It means methods of assessment.
- Written examination
- Oral examination
- Practical/clinical examination

Choice of examination method:[9]
There are mainly three methods- Written, Oral, Practical/Observational methods. The written or pen and paper methods includes essay questions, short essay questions, short answer questions, modified essay questions and MCQs. The observational method is used in

clinical and practical examination; the oral examination (viva voce) is traditional and integral part of assessment in pre, para and clinical phases. Although it involve personal, contact between the examiner and students, though it suffers from a lot of limitations, it should be used to test qualities that cannot be tested by other methods.

The Choice of examination method depends on the purpose of evaluation, the domains of educational objectives are to be tested and feasibility, validity and objectivity of the method as briefed earlier.

Next question is no. 5 who should carry out the assessment?[27]

Teachers: They have a vested interest as they are directly involved with the learners. If they are informed of areas of deficiency in their discipline or area of expertise in terms of the cohort or individual learners, remediation can be arranged.

Head of department: Again, the head of a department has a vested interest in how Students tackled the questions in this discipline or specialty. A proactive head will address any issues that arise in terms of the content or the teachers.
Assessment committee: Most medical schools will have some sort of committee or board which reviews summative assessment results. It will probably be the same group of individuals involved in reviewing questions and Standard-setting. They require feedback on how the questions were handled by the Candidates. They will record the item analyses for improving questions.

Dean: Dean is responsible for the overall assessment of medical schools through professional exams by involving internal and external examiners.

Licensing body: For high stakes final exit examinations, there might be a licensing body involved. In some countries, medical schools have individual examinations and then report the results to the professional or regulatory body. In other cases, all graduating medical students write a national examination.

Conflict of interest: No

Conclusion

No particular method of assessment is perfect. Therefore, no single method of performance assessment/evaluation can be recommended, the teacher should use a variety of methods based on domains to be tested as per previously set objectives during teaching learning session. Emphasis should be placed on testing abilities of knowledge & skills, principles of clinical problem solving and practice of medical care both in primary, secondary and tertiary care settings. It is therefore important that our assessment procedures that it should be allowed to test what our students are expected to learn in order to be competent member of the profession. The choice of methods would be

to assess wide range of knowledge skill and attitude that our students must know and develop in order to carry their jobs well in different settings. If the assessment put into right directions to meet the criteria of the tests performed in respect of objectivity, reliability, validity & feasibility certainly there will be the improvement of teaching learning process towards improving the health care delivery system of the country in order to meet the health care needs of the population at large.

References

1. Boud, D. (1995). Enhancing Learning Through Self-Assessment. London. Routledge Falmer.

2. Seldin P, Miller JE, Seldin CA. The teaching portfolio: A practical guide to improved performance and promotion/tenure decisions. John Wiley & Sons; 2010 Jul 30.

3. Osler W. Examinations, Examiners and Examinees. Lancet, 1913;2:1047-1059.

4. Flexner A. Medical Education: A comparative study, New york, MCMillion, 1925.

5. Newble DI, Extwistle NJ. Learning styles and approaches: implications for medical education.

6. Jayawick ramarajah PT. Oral examination in medical education. Medical Education, 1985;19:290-293.

7. Jayawick ramarajah PT. An inquiry into medical school curricula, doctoral thesis, University of Groningen 1991.

8. Godfery RC. Undergraduate examinations: a confineeing tyranny. British Medical Journal, 1995;345:765-767.

9. Srinivasa DK. Kumar S. Assessment of students achivement, Essential of medical Education, 1st edu.1996:111-113.

10. Ebel RI Frisble DA. Essentials of educational management, 5th edition, new Delhi: Prentice Hall of India private limited, 1991.

11. Heywood J Assessment in higher education: Student learning. Teaching programmes and institutions (higher Education Policy). Jessica kingsly publishers, London 2000.

12. Ebel RL. Measuring educational achivement. Prentoce Hall, Engelwoodcliffs, New Jersey. 1965.

13. http://Professor boker. wordpress.com/2011/03/06/blooms-taxonorney-study glestions-looking for catarina/

14. Miller GE. The assessment of clinical skills/competence/performance. Academic Medicine, 1990;65:63-67.

15. Hammond DL. Synden J. Authentic assessment of teaching in context. Teaching and teacher education, 2000;16:529-545.

16. Cronbach LI. Essentials of psychological testing 4th New york: harpar and Row, 1984 chapter 2; Verities of tests and tesf interpretations" describe tests of maximum performance & features of psychometric & impressionistic setting.

17. Van Der, Linden WJ., Hambleton RK. Editor: Handbook of modern item response theory, Newyork, 1997. Springer Ven Lag

18. Roid G & Haladyna T. A technology for test-them writing. New York: Academic press.1981.Reviews and describes a number of new approaches for developing criterion-referenced achievement tests.

19. Abbat F.R. Teaching for better learning, world health organization, Geneva.

20. Sood R. Assessment in medical education, trends and tools (Ed), SEARO WHO project WR/IND HRH 001/LCS, New Delhi-110029, India 1995.

21. Millman J. & Greenej. "The specification & development of tests of achivement and ability" In RL Linn. Educational measurment 3rd ed New York: Macmillion.1989 chapter 8. A comprehensive & advanced treatment of test development.

22. Crooks TJ. The impact of classroom evaluation practices on student review of educational research, 1988;58:438-81.

23. Gibbs G. Improving the quality of students learning, technical and education services ltd. Bristol 1992.

24. Ramsden P, 1992 London & New York, Routledge ISBN 0-415-06414-7 (hbk); 0-415-06415-5 (pbk), 290 pp. np. Studies in higher education. 1993 Jan 1;18(1):105-11.

25. Hanauer DI., Hatfull GF.,Jacobs- Sera D: Active assessment: Assessing scientific inquiry, Newyork 2009, Springer

26. Glasser R. Instructional Technology and Measurement of Learning Outcome: Some Question American Psychologist, 1963, 18: 519- 581

27. McLean M., Gale R. Health professional education: Accreditation and assessment. FAIMER, Module-1