PERFORMANCE BASED ASSESSMENT

JULY 2009 ML WEB

PSG –FAIMER

2009 fellows – Vinutha and Raghu

2008 fellows – Padma and Feroze

Report by:

Dr Feroze Kaliyadan , MD,DNB,MNAMS

Associate Professor

Department of Dermatology

Amrita Insitute of Medical Sciences, Kochi

Foreword

The July 2009, ML-Web session of PSG-FAIMER regional institute was on the important topic of 'Assessment'. Being a broad topic, we decided to concentrate on one particular aspect – 'Performance Based Assessment'. The planning for the same was done in the month of June and we were able to keep up the schedule as planned. While it goes without saying that I am indebted to the whole PSG-FAIMER group for their valuable inputs, I would like to especially thank the fellows in our group- Vinutha, Padma and Raghu, for their exemplary performance and support. I would also like to thank our faculty for the month – Dr Rita Sood and Dr Rasmi Vyas for their pearls of wisdom. It really was a valuable learning experience for me and I hope it was the same for the rest of the group too.

Sincere regards

Feroze

CONTENTS

Page no:

Introduction

The topic that was chosen for discussion during the on site session was **"Assessment".** Being a part of the educational spiral with learning objectives and T/L methods, assessment evaluates educational outcomes in the cognitive(knowledge), psychomotor(skills) and affective(communication) domain. There is no single assessment method which can evaluate all the above.

We decided to deliberate on instruments of assessment which evaluate performance/skill/ competence **(Performance based assessments)** during our month long discussion with the following objectives:

1. To define clinical competence and discuss competency based curriculum.
2. To describe Performance based assessments (PBA), their goals and relevance
3. To establish the need for alignment of PBA with curricular objectives
4. To identify the skills/competencies to be assessed by PBA
5. To gather input on the existing PBA and discuss their strengths and limitations.
6. To expand knowledge on newer, non traditional formats of PBA
7. To deliberate on the psychometric concepts of PBA (issues like reliability, validity, feasibility etc)
8. To familiarize participants with the standard setting methods in PBA (pass/fail criteria etc)
9. To exchange information on the need for training of assessors in PBA.
10. To understand the importance of constructing a proper checklist in PBA

Strategy and timeline

The timeline of the discussion was planned as below:

1.**Introduction to PBA**: What is it? What is its purpose and what are the principles of PBA and it's alignment with curricular objectives– Jul 1 – Jul 7 -moderated by Vinutha

2.**Methods of PBA**: Traditional and newer ones, their advantages and limitations – Jul 8 – Jul 15 -moderated by Padma

3.**Psychometric challenges in PBA**: Reliability, validity, feasibility - Jul 16 – Jul 22 - moderated by Feroze

4.**Standard setting in PBA**: Types, methods and pass/fail criteria – Jul 23 – Jul 30 - moderated by Raghu

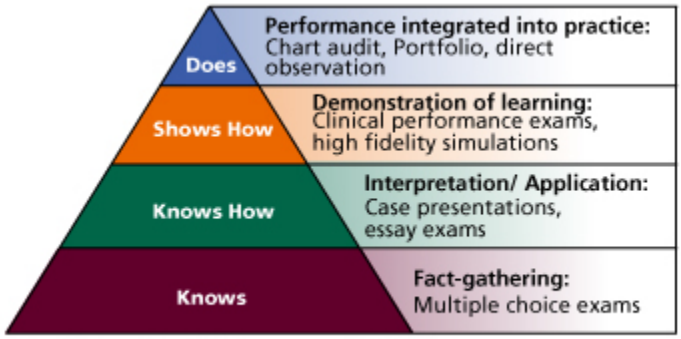
5.**Wrap up and summary**: Jul 31 by Vinutha

Summary of discussions

First week (Introduction to PBA)

The first week's discussion was centred upon the basics of assessment methods. To initiate the discussion, we requested the participants to share information on what they understood by the term "Performance based assessment". A document on "Overview of assessment" was posted which provided insight into the common terminologies used, types, principles and purpose of assessment

Assessment methods aimed at skills and performance generally fall into the 'shows how' and 'does' level and are referred to as Performance based assessment (PBA).



Adapted from Miller GE. The assessment of clinical skills/competence/performance. Acad Med. 1990; 65(9);S63-S67.

PBA are designed to measure skills required for competency in psychomotor and affective domain (behavioral skills e.g. professional behavior, communication skills).

The format of the assessment should be driven by purpose. It is important to employ methods of assessment that specifically assess students' achievement of the skills and behavior they need to learn to practice medicine. The role of Performance Based Assessment (PBA) assumes significance in this context.

Performance-based assessments may include components like oral presentations, open-ended problems, hands-on problems, real-world simulations and other authentic tasks. Such tasks are concerned with problem solving and understanding. Just like standardized achievement tests. The underlying concept is that the student should produce evidence of accomplishment of curriculum goals which can be maintained for later use as a collection of evidence to demonstrate achievement, and perhaps also the teacher's efforts to educate the Student. Performance-based assessment is sometimes characterized as assessing real life, with students assuming responsibility for self-evaluation. Testing is "done" to a student, while performance assessment is done by the student as a form of self-reflection and self-assessment. The overriding philosophy of performance-based assessment is that teachers should have access to information that can provide ways to improve achievement, demonstrate exactly what a student does or does not understand, relate learning experiences to instruction, and combine assessment with teaching.

In broad terms, there are three types of performance-based assessment: performances, portfolios, and projects.

There are several different ways to record the results of performance-based assessments (Airasian,1991 ; Stiggins,1994):

Checklist Approach: When you use this, you only have to indicate whether or not certain elements are present in the performances.

Narrative/Anecdotal Approach: When teachers use this, they will write narrative reports of what was done during each of the performances. From these reports, teachers can determine how well their students met their standards.

Rating Scale Approach: When teachers use this, they indicate to what degree the standards were met. Usually, teachers will use a numerical scale. For instance, one teacher

may rate each criterion on a scale of one to five with one meaning "skill barely present" and five meaning "skill extremely well executed."

Memory Approach: When teachers use this, they observe the students performing the tasks without taking any notes. They use the information from their memory to determine whether or not the students were successful. While it is a standard procedure for teachers to assess students' performances, teachers may wish to allow students to assess them themselves. Permitting students to do this provides them with the opportunity to reflect upon the quality of their work and learn from their successes and failures

A brief discussion also ensued on the present situation in the Indian context. In the present system, all the elements in the Millers' pyramid are (supposedly) assessed (but in reality its mostly knowledge), but in the Outcomes Model paradigm, only the outcome of an educational programme in terms of demonstration of competency acquired by the learner from the educational program (what s/he who is "able to do" or "perform") is assessed.

For example, If the educational program is about swimming, then in Performance Based Assessment, the assessment is about the person's ability to perform swimming . What is tested is whether he knows swimming by observing whether he can swim (performance) rather than knowing the theory behind swimming asking that person to write an essay on swimming and then assessing the essay to arrive at Pass/Fail decisions.

The curriculum in medical education now being predominantly competency based, choosing appropriate assessment methods that accurately assess these clearly defined learning outcomes is the need of the hour. In continuation with the last month’s discussion on the necessity of revising curriculum, it becomes imperative to have

assessments that are aligned to the improvised curricular objectives. Assessment should be consistent with curriculum. The ultimate goal of medical education curriculum is to produce a basic, *competent* doctor. Clinical competence is defined as the “capability to perform acceptably those duties directly related to patient care”. In 1990, George Miller, the doyen of medical education proposed a frame work of assessment methods for measuring clinical competence. The 'ultimate goal of medical education curriculum is to produce a basic, competent doctor'. To make the assessment reliable and comprehensive, there must be a clearly spelt-out check-list of 'must-know' and 'nice-to-know' skills/competencies/definition of concepts

.

Competency means set of behaviors built on the components of knowledge, skills, attitudes. Competence is the personal ability.

There are four important steps.

1) Competency identification

2) Determination of competency components and performance levels

3) Competency evaluation

4) Assessment of the process.

ACGME endorsed 6 general competencies as the foundation of all medical graduates.

Patient care

Medical knowledge

Practice based learning

Interpersonal and communication skills

Professionalism

System –based practice

While designing a curriculum lot of importance is given to these objectives which we would evaluate by appropriate assessment methods. Clear statement of the competency to be achieved will help us design assessment methods that ALIGN with the curricular objectives as we all very well know that assessment is an instrument to know if the objective has been achieved.

The underlying feature that was common in our discussion was having clear-cut objectives/skills/competencies which make PBA effective. If only we could follow **outcome based education (OBE)** where learning outcomes of the course are well defined, T/L methods suitably designed and hence assessments of performance chosen appropriately, don't you think we would have a result oriented system in place. Harden. R was the pioneer in introducing this term OBE in medical education and its basic. principles where the emphasis is on the competencies that are expected from graduate doctors.

360 degree evaluation

It is argued that Mini-CEX and 360 degree evaluation as the best method to assess the "does or performance" of the students during the training and/or working.It could be argued that the Mini-CEX and 360 degree evaluation be used during the MBBS curriculum to actually assess the "Shows How (Competence)" level and not the "Does (Performance)" level. 360-degree feedback, also known as "multi-rater feedback," "multisource feedback," or "multisource assessment," is feedback that comes from all around the assessee.

http://en.wikipedia.org/wiki/360-degree_feedback

Some relevant references and their discussion is available at: http://www.siumed.edu/resaffairs/StaffResources/360evaluation.html

PubMed:

http://www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed&cmd=Search&Term=360%20degree%20evaluation%20medical%20education&itool=QuerySuggestion

In human resources or industrial/organizational psychology, 360-degree feedback, also known as "multi-rater feedback," "multisource feedback," or "multisource assessment," is feedback that comes from all around an employee. "360" refers to the 360 degrees in a circle, with an individual figuratively in the center of the circle. Feedback is provided by subordinates, peers, and supervisors. It also includes a self-assessment and, in some cases, feedback from external sources such as customers and suppliers or other interested stakeholders. It may be contrasted with "upward feedback," where managers are given feedback by their direct reports, or a "traditional performance appraisal," where the employees are most often reviewed only by their managers.

The results from 360-degree feedback are often used by the person receiving the feedback to plan their training and development. Results are also used by some organizations in making administrative decisions, such as pay or promotion. When this is the case, the 360 assessment is for evaluation purposes, and is sometimes called a "360-degree review." However, there is a great deal of controversy as to whether 360-degree feedback should be used exclusively for development purposes, or should be used for appraisal purposes as

well (Waldman et al., 1998). There is also controversy regarding whether 360-degree feedback improves employee performance, and it has even been suggested that it may decrease shareholder value (Pfau & Kay, 2002).

In human resources or industrial/organizational psychology, 360-degree feedback, also known as "multi-rater feedback," "multisource feedback," or "multisource assessment," is feedback that comes from all around an employee. "360" refers to the 360 degrees in a circle, with an individual figuratively in the centre of the circle. Feedback is provided by subordinates, peers, and supervisors. It also includes a self-assessment and, in some cases, feedback from external sources such as customers and suppliers or other interested stakeholders".

It is understood that 360-degree assessment/ feedback model enables individuals to learn from multiple work associates as to their overall effectiveness. More links explaining about the 360 degree assessment/feedback model

1.http://www.ncbi.nlm.nih.gov/pubmed/17701628?ordinalpos=9&itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_ResultsPanel.Pubmed_DefaultReportPanel.Pubmed_RVDocSum

2.http://www.custominsight.com/360-degree-feedback/360-assessments.asp

3.http://www.cafce.ca/download.php?id=384

4.http://www.ceiainc.org/journal/current.asp

5.http://journals.lww.com/academicmedicine/Fulltext/2004/05000/Assessment_of_a_360_Degree_Instrument_to_Evaluate.17.aspx

Second week discussion (Methods of PBA)

Planning Performance-Based Measures

The following should be considered when planning performance-based assessment measures (Palomba & Banta, 1999, p. 118):

- *What skills are being examined?*
- *What tasks can appropriately demonstrate the skills?*
- *What are the criteria for evaluating performances or products?*
- *What is a reliable process for rating the performances or products?*
- *Who is most appropriate to conduct this assessment, and how can they be trained?*
- *How will the results be evaluated?*

Performance-based assessment is the process of using student activities to assess skills and knowledge.

• Performance-based assessment allows faculty to determine student skills and abilities and to help students improve on them.

• Performance-based measures are labor intensive and meant to focus on the program rather than the student.

PBA should fulfill the requirements of curriculum, mainly it should be outcome based. This article highlights the basic views and importance, advantages and disadvantages of the PBA.

OSCE

OSCE can be used for systematic examination of skills starting from history taking and counselling as communication skills. OSCE can be done with checklist.

Features of OSCE

- Stations are short,
- Stations are numerous
- Stations are highly focused , candidates are given very specific instructions
- A pre-set structured mark scheme is used hence…
- Reduced examiner input and discretion

Emphasis on:

- What candidates can do rather than what they know
- The application of knowledge rather than the recall of knowledge

Typically

- 5 minutes most common (3-20 minutes)
- (minimum) 18-20 stations/2 hours for adequate reliability

- Written answer sheets or observer assessed using checklists
- Mix of station types/competences tested
- Examination hall is a hospital ward
- Atmosphere active and busy

Additional options

- Double or triple length stations
- Linked stations
- Preparatory stations
- “Must pass” stations
- Rest stations

CbD

Case-based Discussion is used by a trainee to facilitate a focused discussion around an actual entry in the notes. A senior doctor then assesses him/her on the basis of the discussion.

DOPS

Directly Observed Procedural Skills is used by a doctor, nurse or appropriate Allied Health Professional to assess a trainee performing a procedure, e.g. venepuncture.

mini-CEX

mini-Clinical Encounter (Evaluation) Exercise is used by a senior doctor to assess an actual clinical encounter, or part of it, with a patient by a trainee.

min-ePAT

electronic mini-Peer Assessment Tool is an online multi source feedback tool that provides peer assessment of a trainee by a variety of his/her healthcare colleagues.

SPRAT (Sheffield peer review assessment tool)

- Validated, reliable assessment methods are needed to evaluate doctors
- Multisource feedback has been explored in other countries as a way of assessing traditional and broader competencies, such as professionalism

- Multisource feedback has been evaluated quantitatively for use in the UK
- SPRAT seems to be a valid way of reliably informing the record of in-training assessment process
- With few raters needed for a robust assessment, SPRAT is a feasible way of assessing behaviours that are traditionally hard to capture

OSLER- Objective structured long examination record.

- All candidates are assessed on the 10 items of the record by the examiner over 20-30 minutes- improves reliability
- The items included the are representative of the whole process of working up of a case-- increase the validity
- Mainly more attention for the process of history taking and communication skill.
- **Out of 10 items -**
- 4 on history taking,
- 3 on physical examination
- 1 on formulation of appropriate investigations
- 1 on formulation of appropriate management
- 1 on formulation of appropriate clinical acumen

Standardized patient

A Standardized Patient is a layperson trained to replicate a clinical encounter **consistently and realistically**. Standardized Patients provide the medical student an opportunity to fine-tune professional skills, to gain self-confidence and be better able to instil confidence in patients. The student learns to become patient-oriented, more aware of patient feelings and concerns, and most of all, to become an active listener. The ability to listen and understand and communicate this understanding is a skill of significant benefit to medical students, and Standardized Patients play a central role in this process.

A simple summary-

Competency Domain	Possible Assessment Tools
Patient care	Global evaluations
	Mini-CEX
	Case-based discussions
	Evaluation of simulated experience
	CEX
	Multisource feedback
Medical knowledge	Global evaluations
	In-training and certification examinations
Practice-based learning and	Individual learning plan
improvement	Self-assessment questions (eg, PREP)
	Evaluation of a quality improvement
	project or a modified eQIPP module
Interpersonal and communication	Evaluation of videos, simulations, and/or
Skills	role-plays
	Instant feedback about critical incidents
	Multisource feedback

Professionalism	Mini-CEX for professionalism
	Instant feedback about critical incidents
	Multisource feedback
Systems-based practice	Global evaluation
	Multisource feedback from team members
	Evaluation of an advocacy activity
	Evaluation of a system error analysis

(CEX indicates clinical examination; eQIPP, Education in Quality Improvement in Pediatric Practice; PREP, Pediatrics Review and Education Program.)

ASSESSMENT FORMATS		
Format	**Nature/Purpose**	**Stage**
Baseline Assessments	Oral and written responses based on individual experience Assess prior knowledge	Baseline
Paper and Pencil Tests	Multiple choice, short answer, essay, constructed response, written reports Assess students acquisition of knowledge and concepts	Formative

Embedded Assessments	Assess an aspect of student learning in the context of the learning experience	Formative
Oral Reports	Require communication by the student that demonstrates scientific understanding	Formative
Interviews	Assess individual and group performance before, during, and after a science experience	Formative
Performance Tasks	Require students to create or take an action related to a problem, issue, or scientific concept	Formative and Summative
Checklists	Monitor and record anecdotal information	Formative and Summative
Investigative Projects	Require students to explore a problem or concern stated either by the teacher or the students	Summative
Extended or Unit Projects	Require the application of knowledge and skills in an open-ended setting	Summative
Portfolios	Assist students in the process of developing and reflecting on a purposeful collection of student-generated data	Formative and Summative

Medical student : Competency judgement, support of learning

Medical teacher : Program validation, Program improvement

Professional bodies : Certification and licensing

Medical School : Program justification, Curricular modification, Curricular improvement

Many medical curricula define objectives in terms of knowledge, skills, and attitudes. These cannot be properly assessed by a single test format. All tests should be checked to ensure that they are appropriate for the objective being tested.

A multiple-choice examination, for example, could be a more valid test of knowledge than of communication skills, which might be best assessed with an interactive test.

However, because of the ***complexity of clinical competence, many different tests should probably be used.***

Third week discussion **(Psychometric concepts – related to PERFORMANCE BASED ASSESSMENT methods)**

The aim of the week was to deal with the basics of psychometric concepts in the context of performance based assessment methods. The week started of with a discussion on basic definitions of psychometric concepts in general. One of main topics of discussion was on the importance of construct validity. The discussion went on to discuss the reliability and validity of existing methods. This include methods with an established high level of reliability like OSCEs, to methods like 360 degree evaluation and DOPS, which probably require more studies to establish reliability and validity. A brief mention was made of the importance of checklists in the context of assessments like the OSCEs.

Psychometrics is the field of study concerned with the theory and technique of educational and psychological measurement, which includes the measurement of knowledge, abilities, attitudes, and personality traits. The field is primarily concerned with the study of measurement instruments such as questionnaires and tests.

Some definitions:

Reliability: a measure of whether the assessment (or test) is consistent and accurate; examines the extent to which factors such as examiners, questions, occasions affect the marks (or scores) awarded

Validity: a measure of the extent to which the test actually measures what it is intended to measure. For example, the question 'Discuss the roles of insulin and glucagon in glucose homeostasis' is not a valid assessment of a candidate's ability to manage diabetes

Face Validity: the acceptability of the assessment to the examiners and candidates, ie, does the assessment appear relevant, is the wording appropriate?

Construct Validity: "evidence of validity gained by showing the relationship(s) between a theoretical construct and tests that propose to measure the construct."

Construct validity is the degree to which a score can be interpreted as representing the intended underlying construct. It also states that Validity has traditionally been separated into 3 distinct types, namely, content, criterion, and construct validity and elaborates about each in detail .

The term face validity is sloppy at best and fraudulent and misleading at worst" says Steve Downing, one of the assessment gurus

The key traditional concepts in classical test theory are reliability and validity. A reliable measure is measuring something consistently, while a valid measure is measuring what it is supposed to measure. A reliable measure may be consistent without necessarily being valid, e.g., a measurement instrument like a broken ruler may always under-measure a quantity by

the same amount each time (consistently), but the resulting quantity is still wrong, that is, invalid.

Validity may be assessed by correlating measures with a criterion measure known to be valid. When the criterion measure is collected at the same time as the measure being validated the goal is to establish concurrent validity; when the criterion is collected later the goal is to establish predictive validity. A measure has construct validity if it is related to other variables as required by theory. Content validity is simply a demonstration that the items of a test are drawn from the domain being measured.

The considerations of validity and reliability typically are viewed as essential elements for determining the quality of any test.

Testing standards

In this field, the Standards for Educational and Psychological Testing place standards about validity and reliability, along with errors of measurement and related considerations under the general topic of test construction, evaluation and documentation. The second major topic covers standards related to fairness in testing, including fairness in testing and test use, the rights and responsibilities of test takers, testing individuals of diverse linguistic backgrounds, and testing individuals with disabilities. The third and final major topic covers standards related to testing applications, including the responsibilities of test users, psychological testing and assessment, educational testing and assessment, testing in employment and credentialing, plus testing in program evaluation and public policy.

Evaluation standards

In the field of evaluation, and in particular educational evaluation, the Joint Committee on Standards for Educational Evaluation has published three sets of standards for evaluations. The Personnel Evaluation Standards was published in 1988, The Program Evaluation Standards (2nd edition) was published in 1994, and The Student Evaluation Standards was published in 2003.

Each publication presents and elaborates a set of standards for use in a variety of educational settings. The standards provide guidelines for designing, implementing, assessing and improving the identified form of evaluation. Each of the standards has been placed in one of four fundamental categories to promote educational evaluations that are proper, useful, feasible, and accurate. In these sets of standards, validity and reliability considerations are covered under the accuracy topic. For example, the student accuracy standards help ensure that student evaluations will provide sound, accurate, and credible information about student learning and performance.

Assessments are not valid or invalid; rather, the scores or outcomes of assessments have more or less evidence to support (or refute) a specific interpretation (such as passing or failing a course). Validity is approached as hypothesis and uses theory, logic and the scientific method to collect and assemble data to support or fail to support the proposed score interpretations, at a given point in time. Data and logic are assembled into arguments--pro and con--for some specific interpretation of assessment data. Examples of types of validity evidence, data and information from each source are discussed in the context of a high-stakes written and performance examination in medical education.

Reliability estimates the amount of random measurement error in assessments. All reliability analyses are concerned with some type of consistency of measurement. For written tests, the internal test consistency is generally most important, estimated by reliability indices such as Cronbach's alpha or the Kuder-Richardson formula 20. The internal consistency coefficients are all derived from the test-retest design and approximate the results of such test-retest experiments."

"In order to improve the reliability of assessments, one should maximise the number of questions or prompts, aim for middle difficulty questions, and make certain that all assessment questions are unambiguous and clearly written and are, if possible, critiqued by content-expert reviewers. Pretesting, item tryout and item banking are recommended as means of improving the reliability of assessments in medical education, wherever possible."

1. **Reliability of one hour OSCE and case presentation is same**. In other words, it suggests that reliability of OSCE is not dependent on its check lists and structure. If it were so, it would have been more reliable than a case presentation.

2. **The longer time you give to the student, the better is the reliability**. Figures for 8 hour viva, case presentation and OSCE are almost same.

3. **The higher time allotted denotes including more areas and competencies in the assessment process.** The major source of unreliability in assessment (even more than

erratic examiners) is the content specificity. A student who does well on a CNS case is not necessary equally good on a heart case.

4. There is nothing like 'the reliable' method.

Almost any method can be reliable if you put enough time and effort into it."

Reliability table:(Dr Tejinder Singh, CMC-FAIMER)

Instrument	**1 hour**	**2 hours**	**4 hours**	**8 hours**
MCQ	0.62	0.76	0.93	0.93
Orals	0.5	0.69	0.82	0.9
Long case	0.6	0.75	0.86	0.9
OSCE	0.54	0.69	0.82	0.9
Min CEX	0.73	0.84	0.92	0.96
PMPs	0.36	0.53	0.69	0.82

Fourth week discussion (Standard setting)

In medical education literature, the **standard is defined as the end point of assessment.** we have to set the appropriate standard .eg., minimum competence in advance. This standard is a special score that serves as a boundary between those who perform well and those who don't. Standard setting is a systematic way of gathering value judgement, reaching consensus and that consensus as a single score on a test. As this involves judgement, the credibility of the standard would vary according to who sets the standard. The two that are well known are:

1. Norm referenced standard
2. Criterion referenced standard.

In norm referenced standard, other students' performance is taken into account while deciding on the pass or fail grading of the given student. This is based on the assumption that scores are distributed normally ans score of given student is compared with the scores of other students.

Whereas criterion referenced standard is based on predefined test goals in performance during exams, where a certain level of skill has been determined as required for passing. This method is preferred in performance based assessments to make pass/fail cut off points because this is made independently of other candidate's performance.

The absolute standard/cut score can be used to determine whether the examinee attained the requirement to be certified competent. A number of methods for standard setting are described in literatures. All of them are judgmental with subjectivity and imprecision. There is no perfect method to determine cut score on a test3 and none is agreed upon as the

best method. All these methods have their advantages and disadvantages depending on the specific application. There is no scientific way of choosing a standard setting method from the group. Though different standard setting methods are recommended for different nature and format of tests, they often produce different results

Angoff and Nedelsky methods of standard setting are commonly used methods for performance assessment and certification. The Angoff method is most preferred

as it provides reasonable standard.Setting the standard is tiresome but it is important. It is imperative to identify the cut score to differentiate the competent from the non-competent. Percentage of must know survival knowledge and skills in the curriculum may be the basis of test items and hence the standard/cut score for pass/fail decision. Test items on need-to-know and nice-to-know areas of curriculum may be added to make up the full marks of 100%, which will not be decisive of their basic competence of pass and fail; these may improve their grades if included. Practice analysis may help to identify the curricular content specially the must-know areas. Test items should reflect the representative samples of the learning objectives. When 70% of the learning objectives in the curriculum are of must-know category then it may be appropriate to set the pass/fail marks at 70%. This may give the credibility, validity and defensibility of the standard.

The best experience

The whole session was a wonderful learning experience for me as I was introduced to a bevy of new terms and concepts.If I had to pick out a single area I would choose the discussion on the 360 degress assessment concept.If possible this would be one of most ideal methods of assessing students in our setting, though it is fraught with umpteen practical problems in the present scenario.

What could have been better?

The discussion tended to lose a bit of steam towards the end .Some aspect like ‘Checklists’ which probably needed a more in-depth study were touched only superficially.The problem of ‘silent spectators’ naturally , still unfortunately continue to affect us.

Take home messages

Week 1:

Assessment methods aimed at skills and performance generally fall into the 'shows how' and ‘does’ level and are referred to as Performance based assessment (PBA).

PBA are designed to measure skills required for competency in psychomotor and affective domain (behavioural skills e.g. professional behaviour, communication skills).

The format of the assessment should be driven by purpose. It is important to employ methods of assessment that specifically assess students’ achievement of the skills and behaviour they need to learn to practice medicine. The format of the assessment should be driven by purpose. It is important to employ methods of assessment that specifically assess students’ achievement of the skills and behaviour they need to learn to practice medicine. The role of Performance Based Assessment (PBA) assumes significance in this context. PBA are designed to measure skills required for competency in psychomotor and affective domain (behavioural skills e.g. professional behaviour, communication skills). The curriculum in medical education now being predominantly competency based, choosing appropriate assessment methods that accurately assess these clearly defined learning outcomes is the need of the hour. The curriculum in medical education now being predominantly competency based, choosing appropriate assessment methods that accurately assess these clearly defined learning outcomes is the need of the hour. While designing a curriculum lot of importance is given to these objectives, which we would evaluate, by appropriate assessment methods. Clear statement of the competency to be achieved will help us design assessment methods that ALIGN with the curricular

objectives, as we all very well know that assessment is an instrument to know if the objective has been achieved.

The underlying feature that was common in our discussion was having clear-cut objectives/skills/competencies which make PBA effective. If only we could follow **Outcome Based Education (OBE)** where learning outcomes of the course are well defined, T/L methods suitably designed and hence assessments of performance chosen appropriately, don't you think we would have a result-oriented system in place. Harden. R was the pioneer in introducing this term OBE in medical education and its basic

Week 2:

There a number of different ways to go about PBA.These range from time-tested method like the OSCE/OSPE/Modified OSCEs to newer relatively newer concepts like Standardized /Simulated Patients (SP) ,DOPS,mini-CEX .

Methods of performance based methods need to be used in alignment with the curricular needs.Different tools used include

360-Degree Evaluation Instrument

Chart Stimulated Recall Oral Examination (CSR

Checklist Evaluation of Live or Recorded Performance

Global Rating of Live or Recorded Performance

Objective Structured Clinical Examination (OSCE)

Procedure, Operative, or Case Logs

Patient Surveys

Portfolios

Record Review

Simulations and Models

Standardized Oral Examination

Standardized Patient Examination (SP)

CbD Case-based Discussion

DOPS Directly Observed Procedural Skills

mini-CEX mini-Clinical Encounter (Evaluation) Exercise

min-ePAT electronic mini-Peer Assessment

Competency Domain	Possible Assessment Tools
Patient care	Global evaluations
	Mini-CEX
	Case-based discussions
	Evaluation of simulated experience
	CEX
	Multisource feedback
Medical knowledge	Global evaluations
	In-training and certification examinations
Practice-based learning and	Individual learning plan
improvement	Self-assessment questions (eg, PREP)
	Evaluation of a quality improvement

	project or a modified eQIPP module
Interpersonal and communication Skills	Evaluation of videos, simulations, and/or role-plays
	Instant feedback about critical incidents
	Multisource feedback
Professionalism	Mini-CEX for professionalism
	Instant feedback about critical incidents
	Multisource feedback
Systems-based practice	Global evaluation
	Multisource feedback from team members
	Evaluation of an advocacy activity
	Evaluation of a system error analysis

Week 3:

Psychometrics is the field of study concerned with the theory and technique of educational and psychological measurement, which includes the measurement of knowledge, abilities, attitudes, and personality traits. The field is primarily concerned with the study of measurement instruments such as questionnaires and tests. The key traditional concepts in classical test theory are reliability and validity. A reliable measure is measuring something consistently, while a valid measure is measuring what it is supposed to measure.

-There is a nothing like THE reliable method of assessment

-Almost any method can be reliable if you put enough time and effort into it

-Construct validity is the most important aspect of validity

- Improved reliability of assessments can be obtained by: increasing number of questions, raters or performance cases.

- The main validity threats are CU (construct under representation) and CIV (construct irrelevance validity).

-The 'checklist' forms an important part of assessments like the OSCE and the reliability of the same needs to established.

Instrument	**1 hour**	**2 hours**	**4 hours**	**8 hours**
MCQ	0.62	0.76	0.93	0.93
Orals	0.5	0.69	0.82	0.9
Long case	0.6	0.75	0.86	0.9
OSCE	0.54	0.69	0.82	0.9
Min CEX	0.73	0.84	0.92	0.96
PMPs	0.36	0.53	0.69	0.82

Week 4:

In medical education literature, the **standard is defined as the end point of assessment.**

we have to set the appropriate standard .eg., minimum competence in advance. This

standard is a special score that serves as a boundary between those who perform well and those who don't.

The two that are well known are:

1. Norm referenced standard

2. Criterion referenced standard.

In norm referenced standard, other students' performance is taken into account while deciding on the pass or fail grading of the given student. This is based on the assumption that scores are distributed normally ans score of given student is compared with the scores of other students.

Whereas criterion referenced standard is based on predefined test goals in performance during exams, where a certain level of skill has been determined as required for passing.

References:

1. Miller GE. The assessment of clinical skills/competence/performance. Acad Med. 1990;65:63-7.
2. Brualdi A. Implementing performance assessment in the classroom. Practical Assessment, Research & Evaluation, 6(2). http://PAREonline.net.Accessed August 3, 2009 .
3. Catherine A. Palomba, Trudy W. Banta . Assessment Essentials: Planning, Implementing, and Improving Assessment in Higher Education 1999 Jossey-Bass Publishers San Francisco
4. Epstein RM. Assessment in medical education. NEJM. 2007;356:387-396
5. TOOLBOX OF ASSESSMENT METHODS© A Product of the Joint Initiative ACGME Outcomes Project Accreditation Council for Graduate Medical Education American Board of Medical Specialties (ABMS) version 1.1 September 2000. Available at:www.acgme.org/outcome/assess/toolbox.pdf accessed July 5 2009
6. Archer J, Norcini J, Southgate L, Heard S, Davies H. mini-PAT (Peer Assessment Tool): a valid component of a national assessment programme in the UK? Adv Health Sci Educ Theory Pract. 2008;13:181-92.
7. Archer JC, Norcini J, Davies HA. Use of SPRAT for peer review of paediatricians in training. BMJ. 2005 28;330:1251-3.

8. Munger, BS. Oral examinations. In Mancall EL, Bashook PG. (editors) Recertification: new evaluation methods and strategies. Evanston, Illinois: American Board of Medical Specialties, 1995: 39-42.
9. Noel G, Herbers JE, Caplow M et al. How well do Internal Medicine faculty members evaluate the clinical skills of residents? Ann Int Med. 1992; 117: 757-65.
10. Winckel CP, Reznick RK, Cohen R, Taylor B. Reliability and construct validity of a structured technical skills assessment form. Am J Surg. 1994; 167: 423-27.
11. Nebraska Department of Education (State Government) Assessment-Glossary Available at: http://www.nde.state.ne.us/read/framework/glossary/assessment.html . Accessed : 2009 Jul 14.
12. Colorado State University. Glossary of Key Terms . Available from: http://writing.colostate.edu/guides/research/glossary.Accessed: 2009 Jul 14.
13. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. Am J Med. 2006 Feb;119(2):166.e7-16
14. Downing SM.Face validity of assessments: faith-based interpretations or evidence-based science? Med Educ. 2006 Jan;40(1):7-8.
15. Downing SM, Haladyna TM. Validity threats: overcoming interference with proposed interpretations of assessment data. Med Educ. 2004 Mar;38(3):327-33

16. Downing SM. Reliability: on the reproducibility of assessment data. Med Educ. 2004 Sep;38(9):1006-12.
17. Downing SM. Threats to the validity of clinical teaching assessments: what about rater error? Med Educ. 2005 Apr;39(4):353-5
18. Sloan DA, Donnelly MB, Schwartz RW, Strodel WE. The Objective Structured Clinical Examination. The new gold standard for evaluating postgraduate clinical performance. Ann Surg. 1995 Dec;222(6):735-42.
19. Wood EJ. What are Extended Matching Sets Questions?http://bio.ltsn.ac.uk/journal/vol1/beej-1-2.htm
20. van der Vleuten C. Validity of final examinations in undergraduate medical training. BMJ. 2000 Nov 11;321(7270):1217-9
21. Kaufman DM, Mann KV, Muijtjens AM, van der Vleuten CP. A comparison of standard-setting procedures for an OSCE in undergraduate medical education. Acad Med. 2000 Mar;75(3):267-71
22. Wilkinson TJ, Frampton CM, Thompson-Fawcett M, Egan T. Objectivity in objective structured clinical examinations: checklists are no substitute for examiner commitment. Acad Med. 2003 Feb;78(2):219-23.
23. Tudiver F, Rose D, Banks B, Pfortmiller D. Reliability and validity testing of an evidence-based medicine OSCE station. Fam Med. 2009 Feb;41(2):89-91
24. Singer PA, Robb A, Cohen R, Norman G, Turnbull J. Performance-based assessment of clinical ethics using an objective structured clinical examination. Acad Med. 1996 May;71(5):495-8

25. Stark R, Korenstein D, Karani R. Impact of a 360-degree professionalism assessment on faculty comfort and skills in feedback delivery. J Gen Intern Med. 2008 Jul;23(7):969-72
26. Musick DW, McDowell SM, Clark N, Salcido R. Pilot study of a 360-degree assessment instrument for physical medicine & rehabilitation residency programs. Am J Phys Med Rehabil. 2003 May;82(5):394-402
27. Massagli TL, Carline JD. Reliability of a 360-degree evaluation to assess resident competence. Am J Phys Med Rehabil. 2007 Oct;86(10):845-52
28. Barman A.Standard Setting in Student Assessment: Is a Defensible Method Yet to Come?.Ann Acad Med Singapore 2008;37:957-63
29. Norcini JJ. Setting standards on educational tests. Med Educ 2003;37:464-9.
30. Jaeger RM. Establishing standards for teacher certification tests. Educ Meas 1990;9:15

SURVEY RESULTS ON METHODS OF PERFORMANCE BASED ASSESSMENT

Compiled by Dr.K.M.Padmavathy, 2008 Fellow
22.07.09

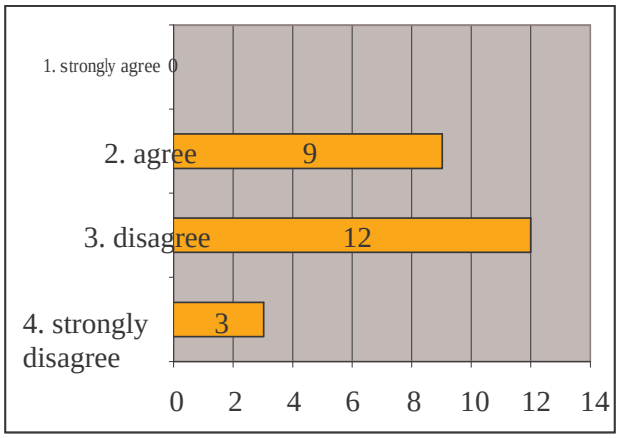
1. The probable reasons for failure of current assessment methods are (order of importance)

	very important	Important	Not so important	Response (n= 24)
1. increased student intake	30.0% (3)	20.0% (2)	50.0% (5)	10
2. increased no. of medical colleges	15.4% (2)	23.1% (3)	61.5% (8)	13
3. shortage of faculty	25.0% (4)	68.8% (11)	6.3% (1)	16
4. non alignment of assessments with curriculum	73.7% (14)	21.1% (4)	5.3% (1)	19
5. Others	1. Faculty not aware of psychometric properties of the assessment methods			2

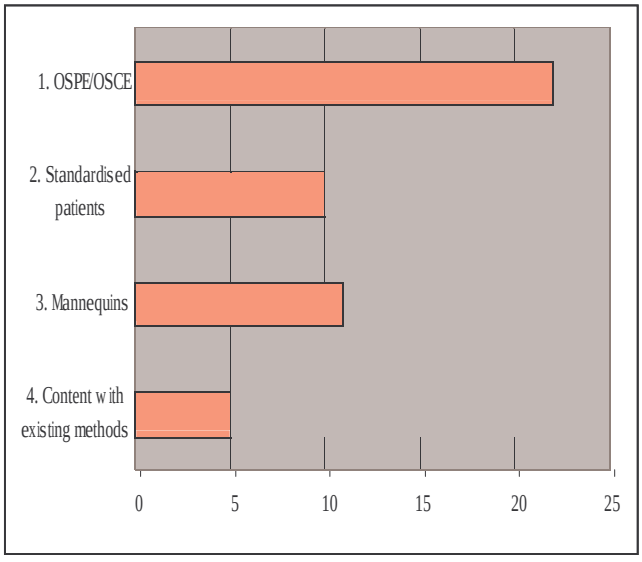
2. The Physical examination of skills can be best assessed by

					Response (n=24)
1. OSPE/OSCE	70.0% (14)	25.0% (5)	5.0% (1)	0.0% (0)	20
2. DOPS	42.9% (6)	35.7% (5)	14.3% (2)	7.1% (1)	14
3. Long case	23.1% (3)	23.1% (3)	23.1% (3)	30.8% (4)	13
4. Simulators	31.3% (5)	43.8% (7)	12.5% (2)	12.5% (2)	16

3. The existing methods of assessment measure learning outcomes effectively

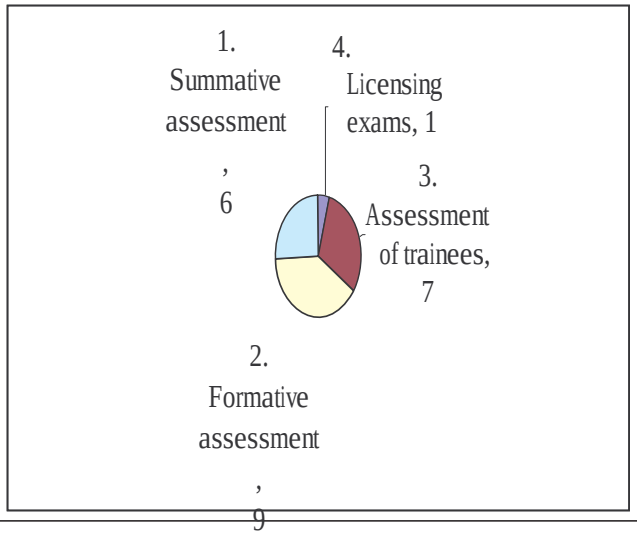
		Response (n=24)
1.strongly agree		0
2. agree		9
3. disagree		12
4. strongly disagree		3

4. Which method would you like to adopt that is feasible in your institution to assess performance? (multiple answers)

		Response (n=24)
1. OSPE/OSCE		22
2. Standardised patients		10
3. Mannequins		11
4. Content with existing methods		5

5. Performance based assessments are best suited for

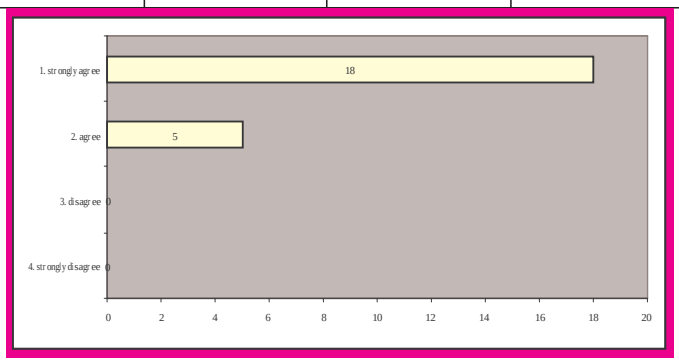
					Response (n=23)
1. Summative assessment					6
2. Formative assessment					9
3. Assessment of trainees					7
4. Licensing exams					1



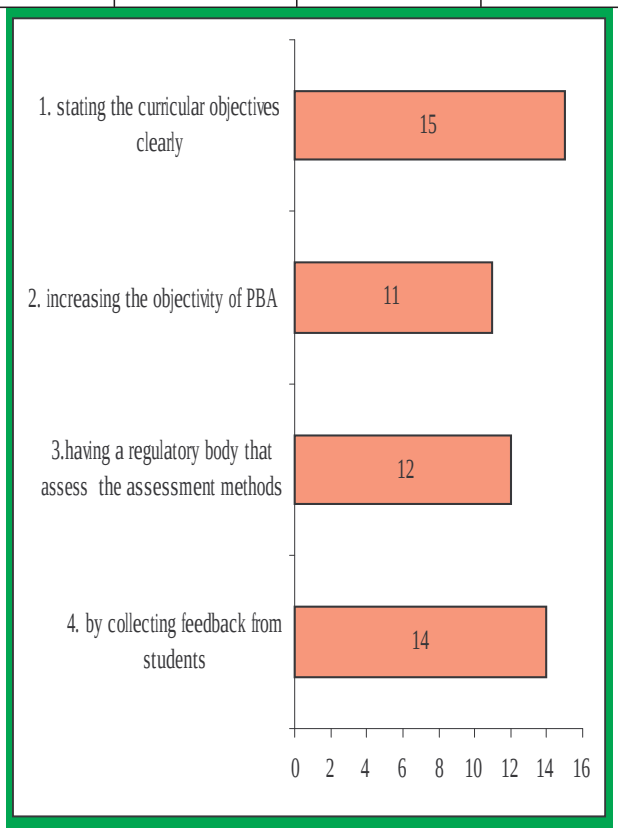
6. The effectiveness of a novel instrument of assessment depends on

					Response (n=24)
1. resources available	60.9% (14)	21.7% (5)	8.7% (2)	8.7% (2)	23
2. trained faculty	68.4% (13)	26.3% (5)	5.3% (1)	0.0% (0)	19
3. encouragement from the institution	55.0% (11)	30.0% (6)	15.0% (3)	0.0% (0)	20
4. regulatory body making it compulsory	57.9% (11)	10.5% (2)	15.8% (3)	15.8% (3)	19

7. There is a definite need to train the assessors to assess

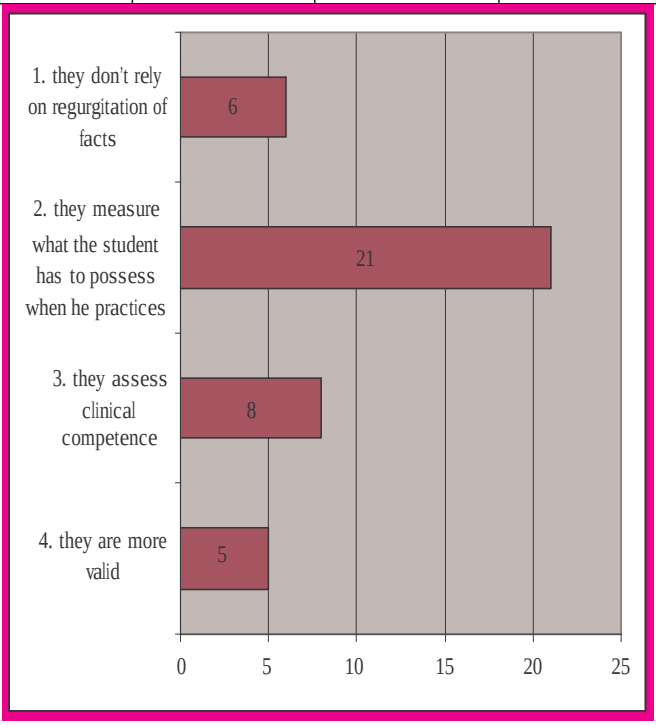
		Response (n=23)
1.Agree strongly		18
2. agree		5
3. disagree		0
4. strongly disagree		0

8. Assessment methods can be improved by

		Response (n=24)
1. stating the curricular objectives clearly		15
2. increasing the objectivity of PBA		11
3.having a regulatory body that assess the assessment methods		12
4. by collecting feedback from students		14

9. Performance based assessments are authentic because

	Response (n=24)
1. they don't rely on regurgitation of facts	6
2. they measure what the student has to possess when he practices	21
3. they assess clinical competence	8
4. they are more valid	5



10. Regulatory bodies (like MCI) has to take the onus of revising assessment methods as and when curriculum is revised. Comment

Answered	19 (24)

IMPORTANT POINTS

It is good and essential for the regulatory body for continuous scrutiny of the correlation of the curriculum objectives and the assessment methods

The regulatory bodies should take the responsibility to improve the existing assessment methods and it can be uniformly followed in all the medical and paramedical institutions.

Though the onus lies on MCI to make sure that the assessment is uniform.

The process can be initiated by individuals and institutions.

somebody should be responsible for all institutions to uniformly adapt

newer , more objective and reliable methods. revision of curricula becomes futile without revising assessment methods

The curriculum has been revised many a times but we still follow the same method of assessment.
Must know ...nice to know areas in our curriculum, whether assessment methods is aligned to this, almost in every examinations we see quite a number of questions from nice to know areas…students and faculty cannot rely on the curriculum given by university. To compensate, faculty try to cover all the topic without paying attention to directive of university thereby utilizing precious time set aside for tutorials and small group discussions.

We must have the learning objectives for each and every teaching session (bedside clinic, lecture, practicals, tutorials etc) and align assessment to learning objectives.

Take home message.

Mandatory to revise the curriculum within specified period

Accordingly assessment methods are also to be revised and uniformly followed in all universities.

There are many good methods but we must know what to assess.