PEER REVIEW AND SELF ASSESSMENT IN MEDICAL EDUCATION

דר שלמה מוניקנדס

מרשת
המחלקה לרופאות המשפחתה
מכלית שרון/أسواق
החבר לרופאות המשפחת אוניברסיטת תל אביב
Assessment methods

- **Validity** is the degree to which the inferences made about medical competence based on assessment scores are correct.
- **Reliability or generalizability** is a measure of the relative magnitude of variability in scores due to error, with the aim of achieving a desired level of measurement precision.
- **Educational effect** of assessment capitalizes on students’ motivation to do well and directs their study efforts in support of the curriculum. For example, if the goal of a particular educational intervention is increased knowledge then a written assessment will appropriately motivate students to study from books. Similarly, a goal of increased clinical skill is best supported by a clinical assessment that motivates students to interact with patients.
- **Feasibility** is the degree to which the assessment method selected is affordable and efficient for the testing purpose; assessments need to have reasonable costs.
- **Acceptability** is the extent to which stakeholders in the process (e.g., medical students and faculty, practicing physicians, patients) endorse the measure and the associated interpretation of scores.
- **Security**
Assessment

• **Mini-peer assessment tool** (mPAT)
  • the trainee nominates eight assessors from among those who are his/her supervisors and peers, including nurses and other health professionals to fill out a questionnaire concerning their technical and interpersonal skills.
  • The trainees also complete a self-assessment using the same questionnaire. The assessment forms are sent directly to the assessors from a central office.
  • Feedback is collated centrally and is presented in a way that shows the self-ratings, the mean rating for the assessor, and the national mean ratings; comments are included. These data are shared with the trainee and the educational supervisor so that there can be agreement about strengths and weaknesses and a plan can be developed for improvement.

Mini PAT

- Originally from industry
- Conducted by professional companies – psychologists
- Ramsey –landmark study 1993 ratings by 11 peer physicians needed selection of raters did not effect results
- Offers overall perception rather than structures task
- Opportunity to compare with self perception / with average peer group
- Identifies strengths and weaknesses
- Psychometric properties?
- Feedback – trainer should be trained well to give feedback -

Peer assessment

- Peer assessment appropriate formative, rather than summative, assessment method for medical students because it permits timely corrective feedback of problem behaviors and encouragement of a student’s strengths.

In contrast, there are questions about the reliability of summative peer assessments and their usefulness in high-stakes settings.
# Peer Assessment Instrument

Please answer the items below indicating the strength of your agreement or disagreement with the statements about this student’s performance in this week of PBL tutorials by circling the number on the scale. (1 = totally disagree; 5 = totally agree)

### The Student:

#### 1. Responsibility and Respect

1. Completed all assigned tasks to the appropriate level
2. Completed all assigned tasks on time
3. Participated actively in the tutorial
4. Showed behaviour and input which facilitated my learning
5. Was punctual to this PBL tutorial
6. Listened to and showed respect for the opinions of others

#### 2. Information Processing

7. Brought in new information to share with the group
8. Provided information that was relevant and helpful
9. Seemed to use a variety of resources to obtain the information

#### 3. Communication

10. Was able to communicate ideas clearly
11. Made comments and responses that did not confuse me

#### 4. Critical Analysis

12. Gave input which was focussed and relevant to the case
13. Made conclusions that can be substantiated by the evidence presented in the case
14. Gave a thorough summary of the case
15. Gave a summary of the case which showed evidence of reflection and evaluation

#### 5. Self-Awareness

16. Appeared able to assess his/her own strengths and weaknesses within PBL
17. Accepted and responded to criticism gracefully

*adapted from Das, 1998*
Peer assessment

- Self-assessment appears to be a less accurate means of evaluating student performance in PBL tutorials than peer-assessment.

- Peer-assessment offers a greater likelihood of providing accurate alternate forms of assessment within the PBL tutorial environment.

Peer-assessment provides a valuable opportunity for tutorial-based assessment. Skills gained through peer-assessment activities may transfer to self-assessment tasks and enable learners to compare their self-assessment with the assessments of others. Feedback from peers has the potential to assist learners to develop more accurate impressions of themselves and their abilities (Eva & Regehr 2005).
(Peer) Assessment (Epstein)

- Insight into trainees’ work habits, capacity for teamwork, and interpersonal.
- Multisource feedback is most effective
  when it includes narrative comments as well as statistical data,
  when the sources are recognized as credible,
  when the feedback is framed constructively,
  and when the entire process is accompanied by good mentoring and follow-up.

When trainees receive thoughtful ratings and comments by peers in a timely and confidential manner, along with support from advisers to help them reflect on the reports, they find the process powerful, insightful, and instructive.

Peer assessments have been shown to be consistent regardless of the way the raters are selected.

Such assessments are stable from year to year and predict subsequent class rankings as well as subsequent ratings by supervisors.

Peer assessments depend on trust and require scrupulous attention to confidentiality. Otherwise they can be undermining, destructive, and divisive.
<table>
<thead>
<tr>
<th>Method</th>
<th>Domain</th>
<th>Type of Use</th>
<th>Limitations</th>
<th>Strengths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Written exercises</td>
<td>Knowledge, ability to solve problems</td>
<td>Summative assessments within courses or clerkships, national in-service, licensing, and certification examinations</td>
<td>Difficult to write, especially in certain content areas; can result in cueing; can seem artificial and removed from real situations</td>
<td>Can assess many content areas in relatively little time, have high reliability, can be graded by computer</td>
</tr>
<tr>
<td>Key-feature and script-concordance questions</td>
<td>Clinical reasoning, problem-solving ability, ability to apply knowledge</td>
<td>National licensing and certification examinations</td>
<td>Not yet proven to transfer to real-life situations that require clinical reasoning</td>
<td>Assess clinical problem-solving ability, avoid cueing, can be graded by computer</td>
</tr>
<tr>
<td>Short-answer questions</td>
<td>Ability to interpret diagnostic tests, problem-solving ability, clinical reasoning skills</td>
<td>Summative and formative assessments in courses and clerkships</td>
<td>Reliability dependent on training of graders</td>
<td>Avoid cueing, assess interpretation and problem-solving ability</td>
</tr>
<tr>
<td>Structured essays</td>
<td>Synthesis of information, interpretation of medical literature</td>
<td>Preclinical courses, limited use in clerkships</td>
<td>Time-consuming to grade, must work to establish interrater reliability, long testing time required to encompass a variety of domains</td>
<td>Avoid cueing, use higher-order cognitive processes</td>
</tr>
<tr>
<td>Assessments by supervising clinicians</td>
<td>Clinical skills, communication, teamwork, presentation skills, organization, work habits</td>
<td>Global summative and sometimes formative assessments in clinical rotations, subjective feedback</td>
<td>Often based on second-hand reports and case presentations rather than on direct observation</td>
<td>Use of multiple independent raters can overcome some variability due to subjectivity</td>
</tr>
<tr>
<td>Structured direct observation with checklists for ratings (e.g., mini-clinical-evaluation exercise or video review)</td>
<td>Communication skills, clinical skills</td>
<td>Limited use in clerkships and residencies, a few board-certification examinations</td>
<td>Selective rather than habitual behaviors observed, relatively time-consuming</td>
<td>Feedback provided by credible experts</td>
</tr>
<tr>
<td>Oral examinations</td>
<td>Knowledge, clinical reasoning</td>
<td>Limited use in clerkships and comprehensive medical school assessments, some board-certification examinations</td>
<td>Subjective, sex and race bias has been reported, time-consuming, require training of examiners, summative assessments need two or more examiners</td>
<td>Feedback provided by credible experts</td>
</tr>
<tr>
<td>Clinical simulations</td>
<td>Some clinical skills, interpersonal behavior, communication skills</td>
<td>Formative and summative assessments in courses, clerkships, medical schools, national licensure examinations, board certification in Canada</td>
<td>Timing and setting may seem artificial, require suspension of disbelief, checklists may penalize examiners who use shortcuts, expensive</td>
<td>Tailored to educational goals; reliable, consistent case presentation and ratings; can be observed by faculty or standardized patients; realistic</td>
</tr>
<tr>
<td>Incognito standardized patients</td>
<td>Actual practice habits</td>
<td>Primarily used in research, some courses, clerkships, and residencies use for formative feedback</td>
<td>Requires prior consent, logistically challenging, expensive</td>
<td>Very realistic, most accurate way of assessing clinician's behavior</td>
</tr>
<tr>
<td>High-technology simulations</td>
<td>Procedural skills, teamwork, simulated clinical dilemmas</td>
<td>Formative and some summative assessment</td>
<td>Timing and setting may seem artificial, require suspension of disbelief, checklists may penalize examiners who use shortcuts, expensive</td>
<td>Tailored to educational goals, can be observed by faculty, often realistic and credible</td>
</tr>
<tr>
<td>Multisource (“360-degree”) assessments</td>
<td>Professional demeanor, work habits, interpersonal behavior, teamwork</td>
<td>Formative feedback in courses and comprehensive medical school assessments, formative assessment for board recertification</td>
<td>Confidentiality, anonymity, and trainee buy-in essential</td>
<td>Ratings encompass habitual behaviors, credible source, correlates with future academic and clinical performance</td>
</tr>
<tr>
<td>Peer assessments</td>
<td>Ability to gain patients' trust, patient satisfaction, communication skills</td>
<td>Formative and summative, board recertification, use by insurers to determine bonuses</td>
<td>Provide global impressions rather than analysis of specific behaviors, ratings generally high with little variability</td>
<td>Credible source of assessment</td>
</tr>
<tr>
<td>Patient assessments</td>
<td>Knowledge, skills, attitudes, beliefs, behaviors</td>
<td>Formative</td>
<td>Do not accurately describe actual behavior unless training and feedback provided</td>
<td>Foster reflection and development of learning plans</td>
</tr>
<tr>
<td>Self-assessments</td>
<td>All aspects of competence, especially appropriate for practice-based learning and improvement and systems-based practice</td>
<td>Formative and summative uses across curriculum and within clerkships and residency programs, used by some U.K. medical schools and specialty boards</td>
<td>Learner selects best case material, time-consuming to prepare and review</td>
<td>Display projects for review, foster reflection and development of learning plans</td>
</tr>
</tbody>
</table>
Self assessment

• Fundamental cognitive limitations in the ability of humans to know themselves as others see them restrict the usefulness of self-assessment

• .... rating oneself on prior clinical performance may not achieve another important goal of self-assessment: the ability to monitor oneself from moment to moment during clinical practice.

  A physician must possess this ability in order to meet patients’ changing needs, to recognize the limits of his or her own competence, and to manage unexpected situations
SELF ASSESSMENT
Table of Contents

Foundations in Continuing Education

1 Toward a common understanding of self-assessment
   Joan Sargeant

5 Self-monitoring in clinical practice: A challenge for medical educators
   Ronald M. Epstein
   Daniel J. Siegel
   Jordan Silberman

14 “I’ll Never Play Professional Football” and Other Fallacies of Self-Assessment
   Kevin W. Eva
   Glenn Regehr

20 Making Self-Assessment More Effective
   Robert M. Galbraith
   Richard E. Hawkins
   Eric S. Holmboe

25 Self-Assessment and Continuing Professional Development: The Canadian Perspective
   Ivan Silver
   Craig Campbell
   Bernard Marlow
   Joan Sargeant

32 Self-Assessment in CPD: Lessons From the UK Undergraduate and Postgraduate Education Domains
   Tim Doman

38 Self-Assessment of Practice Performance: Development of the ABIM Practice Improvement Module (PIM™)
Self assessment

Distinguish from:

• Self-directed Assessment seeking
  (mcq test)
• Reflection

• Self monitoring
  (looking something up)

“I'll Never Play Professional Football” and Other Fallacies of Self-Assessment
KEVIN W. EVA, PHD; GLENN REGEHR J CONT EDUC HEALTH PROFESSIONS, 28(1):14–19, 2008
Self assessment

- Self efficacy and Self concept
- Cognitive and Metacognitive Theory
- Social Cognition

Self-Assessment in the Health Professions: A Reformulation and Research Agenda
Eva and Regehr. Acad Med 80, 10 Oct 2005 Suppl
Self assessment

• ..... the route to self improvement is not through becoming a more accurate self-assessor, but through seeking out feedback from reliable and valid external sources (experts, self administered tests etc.), and then, according to the self-reflection literature, making a special effort to take the resulting feedback seriously rather than discounting it: to reflect rather than ruminate.

• Indeed Boud has suggested that the phrase self-assessment should not imply an isolated or individualistic activity; it should commonly involve peers, teachers, and other sources of information.

Self-Assessment in the Health Professions: A Reformulation and Research Agenda. Eva and Regehr. Acad Med. 80, 10 Oct 2005 Suppl
Self assessment

• Finally, we must be aware that the purpose of self-assessment is more complicated than simply “finding gaps and learning more.” As the self-efficacy literature suggests, at the very least there are times when accurate self-assessment is not always consonant with improved performance.

• But, perhaps most importantly for daily practice, ……, it seems likely that most of the value that self-assessment provides for the practitioner does not happen at the level where the individual is consciously reflecting on her performance or ability at a time that is remote from the performance or use of that ability. While this “reflection-on practice” may be important to self-directed learning and continuous professional development, there is an important way in which it may not be vital to self-regulation and safe practice.

  Safe practice in a health professional's day-to-day performance requires an awareness of when one lacks the specific knowledge or skill to make a good decision regarding a particular patient (i.e., when more information and/or a consultation is required).

  This.

• decision making in context is importantly different from being able to accurately rate one’s own strengths and weaknesses in an a contextual manner

Self-Assessment in the Health Professions: A Reformulation and Research Agenda
Eva and Regehr. Acad Med. 80, 10 Oct 2005 Suppl
Self assessment

- The self

Most studies in the health professions explore the personal assessment of performance as an ability to identify one's own strengths and weaknesses in relation to other peoples' views. However, with this approach self-assessment is described as poor, and as shaped by culture and gender rather than as representative of a shared reality. There is no evidence for the effectiveness of self-assessment using this approach. It is methodologically flawed.

Social psychologists view self-assessment as flawed because our behaviour and performance are informed by our unconscious minds. This unconscious self is focused on self-preservation. It helps explain why feedback that threatens self-esteem or contains nothing more than unconditional praise may be less effective. Similarly, recipients of negative feedback blame external factors and reject personal responsibility; this is known as fundamental attribution error.