
Principles of Assessment: A Primer for Medical Educators in the Clinical Years

A Vergis, K Hardy

Citation

A Vergis, K Hardy. *Principles of Assessment: A Primer for Medical Educators in the Clinical Years*. The Internet Journal of Medical Education. 2009 Volume 1 Number 1.

Abstract

Whether practicing in a rural, community, or an academic setting, physicians from all clinical specialties will participate in assessment. These assessments may be for trainees, peers, and more recently, for self-assessment. Regardless of the subject, assessors may be uncomfortable making judgments because they are unfamiliar with assessment principles. This editorial review, although a primer and aimed at the novice, will also provide information for more experienced assessors when considering assessment purpose, design, and selection. Using concrete examples, these fundamental principles are illustrated so that physicians can be confident that their evaluations are accurate, insightful and meaningful.

INTRODUCTION

In a simplistic sense, the purpose of assessment is to enhance learning. To this end, the character of assessment in medical education has been dissected, evaluated and refined for decades.^{1,2} Indeed, this interest has led to broader notions of what assessment should be doing than there had been in the past.² According to Broadfoot, "Assessment is on the agenda because change is on the agenda; because there is growing pressure in many countries for the education system to do more and different things; because it is felt that assessment is key to achieving these changes".³

If the purpose of assessment is to enhance learning, the purpose of teaching is to facilitate it. Before any particular teaching method can be widely implemented in health sciences education, however, there must be a method to assess its product. Generations of medical educators have outlined questions that guide decisions about developing the most appropriate method for assessing a learned skill.^{4,6} These considerations include: What is the purpose of the assessment? What should be assessed? What are the attributes of an effective assessment? What assessment technique should be used? How will the assessment be categorized and its results judged? Who should do the assessment? When should the assessment occur?

This paper will attempt to answer these questions. In so doing, it will provide a source of material for those involved with assessment. Although, as a primer, it is aimed at the novice, it will also contain useful information for the more

experienced assessor.

WHAT IS THE PURPOSE OF THE ASSESSMENT?

Perhaps the most important consideration is knowing the primary reason for instituting the assessment at all.⁵ In medical education, assessment is a dynamic and multi-faceted process with variable aims. These may include: providing a means by which students are graded or advanced; licensing students for practice; enabling student feedback on the quality of their learning; enabling teachers to evaluate the effectiveness of their teaching; and maintaining academic standards.⁷ By reflecting upon purpose, the educator establishes a framework in which the assessment method may be defined.

WHAT SHOULD BE ASSESSED?

Defining the purpose of an assessment shapes the important consideration of what should be assessed. In an effectively-designed curriculum, course objectives will mirror the assessment content because they both serve to facilitate the same educational product.⁸ Broadly classified, educational objectives fall into three domains: knowledge, skills, and attitudes. As illustrated by Harden, knowledge objectives are those that address cognitive measures.⁹ These range on a continuum from being able to recall factual events to integrating processes for problem solving. Skills objectives involve psychomotor aspects that are needed to be an efficient clinician. Attitude objectives relate to personal

qualities of the learner and their approach to medicine, patients and their peers. By harmonizing course objectives with assessment content, educators ensure a unified curriculum.

WHAT ARE THE ATTRIBUTES OF AN EFFECTIVE ASSESSMENT?

Next, it is important to consider the attributes desirable for an effective assessment tool. This consideration requires an understanding of the fundamental concepts of validity and reliability. Also, as outlined by Turnbull, an ideal assessment tool would also possess the following features (Table 1): accountability, flexibility, comprehensiveness, feasibility, timeliness and relevance to both the examiner and examinee.

6

Figure 1

Table 1. Qualities of an ideal assessment tool

Quality	Brief explanation
Validity	The degree to which an assessment measures what it purports to
Reliability	The reproducibility of an assessments results
Accountability	To relevant stakeholders (students, staff, institution, community)
Flexibility	The versatility of the assessment tool
Comprehensiveness	The overall scope of the assessment
Feasibility	The practicality of using the assessment
Timeliness	The elapsed time between the target behavior and assessment
Relevance	The perceived significance of the assessment

Adapted from: *Teaching the Surgical Craft: From selection to Certification*⁴

VALIDITY

In assessment, the fundamental property of any testing method is that ““it measures what it purports to measure””¹⁰. This, in essence, is an assessment’s validity. While an outwardly simple concept, validity testing often requires the availability of other frameworks to which the results of the index assessment can be compared.¹¹ A test then, may have multiple aspects of validity.^{5,11} These aspects may be compiled to establish the overall validity of a particular assessment method. To rationalise its multiple facets, several standards have been developed in the educational literature to appraise the validity of an assessment instrument. These standards include face, content, construct, and criterion validity.

FACE VALIDITY

Face validity, the most subjective form of validity, relates to an item or theory making common sense and seeming correct to the expert reader. While its simplicity is attractive, its vague nature and subjectivity create difficulties. For example, one expert may feel that an endoscopic skills examination based on popular arcade games lacks face validity because the target behaviour is different, while another feels it is high, since arcade performance illustrates

transferable endoscopic skills. Although face validity may lend aesthetically to an assessment, its limitations are preclusive as a sole measure of validity.

CONTENT VALIDITY

Content validity measures the extent to which assessment items reflect the overall domain of knowledge and skills required for mastery of the subject. For example, for an assessment designed to specifically to measure chest tube insertion technique, the instrument must ensure that it is not measuring something related, but different, such as indications for chest tube insertion. Content validity is also highly subjective in that it relies on the opinions of content experts about the relevance of items used.¹⁰

CONSTRUCT VALIDITY

Construct validity examines the extent to which an assessment measures a non-observable trait that explains behaviour. This allows an assessor to infer a psychological construct from test scores. For example, one may theorise that resident performance in the intensive care setting relates to a sophisticated understanding of animal physiology. Although it would be difficult to assess every aspect of animal physiology (i.e. from single celled organisms to humans), a validated assessment in general animal physiology could be administered and correlated with established intensive care test scores. If correlation is high, the relationship is demonstrated. In this example, measuring construct validity could be useful in course design by identifying requisites for achievement.

CRITERION VALIDITY

Criterion validity examines the degree to which a tests correlates to other measures of performance. Within this category are two subtypes: Concurrent and predictive validity.

Concurrent validity is the degree of agreement between the scores of the index test and an established one (i.e., the ““Gold Standard””) when administered at the same time. An example would be a medical school that wishes to improve time and cost efficiency by introducing computer simulators to test gross anatomy knowledge. To assess concurrent validity, the school would administer both the computer simulation and the traditional cadaver examination. If correlation with the examination scores is high, the simulator demonstrates the validity of the established assessment. With this agreement, the school may more comfortably adopt the new assessment and utilise its advantages.

Predictive validity. While each of the illustrated types of validity is important, predictive validity is the most likely to provide a clinically useful assessment. It is defined by the extent to which test scores relate to actual ability. An example of a test with high predictive ability would be a surgical skills simulator assessment whose scores correlate directly to performance in the operating room.

For more information on validity please see Gallagher et al.¹⁰ and Downing.¹¹

RELIABILITY

Reliability relates to the precision, stability or reproducibility of an assessment tools results. In basic mathematical terms, reliability is estimated as:

$$R_x = V_t/V_x$$

Where: R_x is the reliability in the observed (test) score, X ; V_t and V_x are the variability in 'true' (i.e., candidate's innate performance) and measured test scores respectively.

Simply stated, reliability is a term that covers the dependability of an assessment and measures the extent to which a test will yield the same result after multiple administrations under the same conditions.¹⁰ Reliability is recorded as a coefficient on a scale from 0 to 1. A test with a reliability coefficient of 0 is completely unreliable. That is, the variability in test results are independent of candidate ability. A test with a coefficient of 1 indicates complete reliability and is rarely achieved. There is general agreement that if important decisions are going to be based on the results of a test, a reliability of ≥ 0.8 is required.⁵

In general, the reliability of an assessment is easier to determine than validity. Like validity, there are a number of methods to establish a test's reliability. Important methods include internal consistency, test-retest, equivalent forms, and inter-rater reliability.

INTERNAL CONSISTENCY RELIABILITY

Internal consistency is a reliability estimation where a single test is administered to a single group on one occasion to determine the test's internal consistency. While there are many types of internal consistency, split-half reliability nicely illustrates this concept. In this method, items that purport to measure the same construct are randomly divided into two sets. The entire test is administered and the total score is calculated for each random half. The split-half reliability estimate is the correlation between these two

scores. In more sophisticated but similar method, the internal consistency of a test is measured with Cronbach's alpha coefficient. In essence this method correlates the performance of pupils using all possible random split-halves.

TEST-RETEST RELIABILITY

The test-retest method measures the degree to which test results are consistent over time. This reliability coefficient is calculated by comparing results of the same test administered to the same testing population on two separate occasions. There are several problems with this method. First, it is often impractical to administer the same assessment on multiple occasions. Second, if the tests are given too closely together, the students will remember their answers from the initial sitting (thus artificially increasing test-retest reliability). Finally, it is difficult to control for information learned by the student between administrations, especially if the interval is long.

EQUIVALENT FORMS RELIABILITY

When two forms of the same assessment exist, equivalent forms reliability may be determined. In this method, the first form of the test is given followed soon after by the second. Correlations are then calculated between the results. Although the specific items of the two forms may be dissimilar, the two tests should be the same length, structure, and level of difficulty. Also, they must measure the same objectives. The main difficulty with this method is the practicality of designing two essentially equivalent assessments that measure the same construct. Although similar, equivalent forms reliability differs from split-half reliability in that equivalent forms of a test are constructed that can be used independently from one another.

INTER-RATER RELIABILITY

In many situations, an institution may want to disseminate a testing tool for use by multiple examiners on different occasions or for ≥ 2 examiners that are assessing the performance of a single examinee. In these situations, it is useful to determine the inter-rater reliability of the assessment. This measure assesses the degree to which test scores are dependant on the candidate's performance rather than on the particular examiner administering the test. For example, to estimate the consistency of two individuals examining a subject using categorical items (such as item demonstrated or not demonstrated), percent agreement could be calculated. If the two individuals check the same category for 6 out of 10 items, the inter-rater reliability would be 60%.

For more information on reliability, please see Gallagher et al.¹⁰

ACCOUNTABILITY

Any assessment mechanism must be accountable to all ‘stakeholders’ involved. This is a fundamental principle from which the other characteristics of the ideal assessment tool should arise. In academic medicine, these stakeholders include students, clinical educators, the program and institution, licensing bodies and ultimately the community that the clinician will serve. To facilitate accountability an assessment must be defensible and able to provide a logical analysis or explanation of results.¹²

FLEXIBILITY

Clinical medicine is practiced in a diverse and sometimes unpredictable environment. Therefore, the chosen assessment method must be flexible and allow the examiner to evaluate the complete clinical spectrum of the content domain in question multiple times and in multiple settings (e.g., elective and emergency surgeries).

COMPREHENSIVENESS

To be effective overall, an assessment will evaluate all pertinent objectives and document corresponding examinee performance for the course it was designed to evaluate. The CanMeds competency framework, which defines seven essential physician competencies (Medical Expert, Professional, Communicator, Collaborator, Manager, Health Advocate, and Scholar), illustrates the robustness of clinical practice.¹³ With this view of practice, the clinical educator can better understand the importance of comprehensive mechanisms to assess trainees.

FEASIBILITY

To facilitate acceptance by all stakeholders, an assessment should be portable, cost-effective, practical, and limit physical and human demands. This is important, as an assessment that taxes limited resources, is unlikely to find mainstream acceptance. However, under some circumstances, these considerations may be tempered. For example, for licensure or other high-stakes examinations, a more labour intensive and expensive assessment may be used if it is proven to be superior to other available tests. Examinations such as the Objective Structured Clinical Examination or Objective Structured Assessment of Technical Skills (discussed later) are examples of resource intensive but valuable assessment tools.

TIMELINESS

To maximise its function, assessment should be administered as close to the target behavior as possible. Undue delay allows for recall of target events to degrade and thus increases the subjectivity of assessment. Also, the results of the assessment should be communicated to the examinee (and other relevant parties) promptly. Failure to do so deprives stakeholders of the assessments full utility (e.g., feedback or curriculum planning functions). Indeed, if documentation is delayed, assessment is less effective as a learning tool, more subject to bias, and less defensible.¹⁴

RELEVANCE

To be effective, the importance of the assessment must be apparent to all involved stakeholders. The results of assessment, favorable or not, must be used to facilitate learning and influence promotion and curriculum planning decisions. An assessment that is viewed as irrelevant cannot fulfill these functions because its results appear meaningless and unusable.

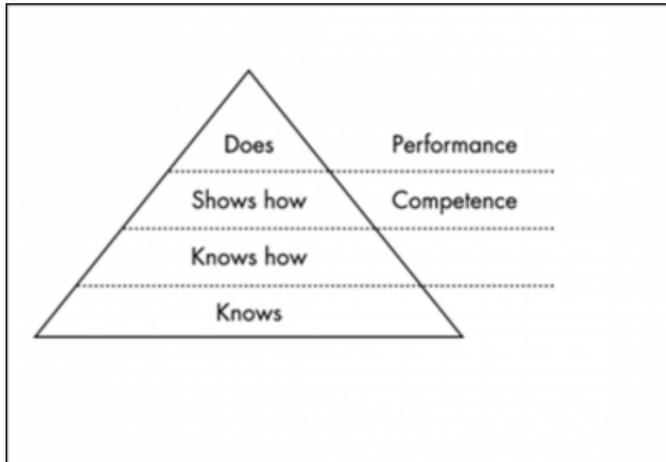
WHAT ASSESSMENT TECHNIQUE SHOULD BE USED?

To date, a range of assessment techniques has been described and utilised in all areas of medical education. Although too numerous to describe, each method has its own inherent advantages and disadvantages. When choosing a method, it is important that the assessment technique be closely related to what one is trying to examine.⁴ This concept is reflected in Miller’s triangle model which attempts to stage clinical competence.¹⁵ In this model, the cognitive and behavioral progression a learner makes from acquiring knowledge to performing a task is illustrated in four stages. These are: knows, knows how, shows how, and does (Figure 1).

For example, if the aim is to examine a candidate’s factual recall (“‘knows’”), a multiple choice or extended matching item examination may be sufficient. If a candidate’s thought process is the target (“‘knows how’”), an essay format or oral examination may be useful by allowing a free and extended arena to formulate a response.

Figure 2

Figure 1. Millers Triangle

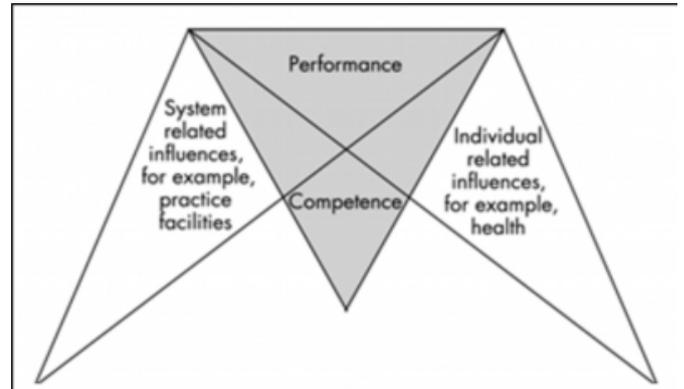


In clinical medicine it is important to distinguish between what a candidate knows and what they can do (“shows how”). Here, the clinical and practical assessment techniques are important. These techniques importance have lead to more objective approaches to clinical assessment over the past 30 years. The Objective Structured Clinical Examination (OSCE) and more recently the Objective Structured Assessment of Technical Skill (OSATS) and Patient Assessment and Management Examination (PAME) are well known examples of these.^{9, 16, 17}

Miller’s triangle assumes that competence will predict actual clinical performance (“does”). However, this may not be the case as many other factors can influence clinical performance. To address this, the “Cambridge Model” (Figure 2) expands Miller’s triangle to illustrate individual and system related influences that effect performance.¹⁸ This model distinguishes competence (what a candidate demonstrates during the examination) and performance (what the candidate demonstrates in real practice). With this model, the authors illustrate the need, when appropriate, to assess true clinical performance in addition to “in vitro” assessments of knowledge and practical skill.

Figure 3

Figure 2. The Cambridge Model for delineating performance and competence.



HOW WILL THE ASSESSMENT BE CATEGORISED AND ITS RESULTS JUDGED?

Another important consideration when developing an assessment method is how the assessment will be categorised and its results judged. In general, the two categories are formative and summative assessment, and judging can be either norm or criterion referenced.¹⁹

FORMATIVE ASSESSMENT

Formative assessment involves gathering findings from a variety of assessment sources. These findings are then used to chart a student’s progress through a particular course of learning. Importantly, formative assessment uses information to ‘feedback’ into the learning and teaching process which can be used for either student or program assessment.^{5, 20} In student assessment, a formative assessment is intended to give ongoing constructive feedback on a student’s strengths and weaknesses during a course. This feedback is student-centered and does not focus on the student’s ranking within a particular group. In program assessment, this assessment aims to improve the quality of the program. In neither student nor program assessment is formative assessment used to make pass or fail decisions.⁵

SUMMATIVE ASSESSMENT

In contrast to formative assessment, summative assessment is designed to accumulate information from all relevant sources and to determine whether course objectives have been adequately met. Summative recommendations usually occur at the end of a course and are used to make pass/fail/rank decisions. The intention here is to determine what has been learnt.

For more information on formative and summative

assessment please see Wanzel et al.⁵

NORM AND CRITERION REFERENCING

Norm and criterion referencing are two common methods of relating a student's raw performance against a standard so that comparisons or rankings can be drawn. Norm referencing is the more conventional method and is used to describe a candidate's performance in terms of their position in a group. Results are usually reported as a percentage of correct responses where the number of students to pass or be given a particular grade has been predetermined. Students often refer to this as the grade 'curve' for the class.

Conversely, criterion referencing has particular importance in professional education where there is more concern that the student attains a minimal level of competence rather than focusing on their ranking within a peer group. "Standards of performance are set using minimal levels of competence before the test is applied. Following the test there is no attempt to alter the percentage of pass or fail. The standards of performance are based upon uncompromising pre-set objectives."²¹

Both forms of judgment have merit in different circumstances. For example, norm referencing is useful when determining which candidates from a pool should be selected for a limited number of positions in a medical school. On the other hand, criterion referencing would be more appropriate in determining who graduates from medical school because a significant percentage or all of the candidates of that pool may be particularly skilled.

For more information on norm and criterion referencing please see Turnbull.²¹

WHO SHOULD DO THE ASSESSMENT?

The traditional form of assessment in medical education has been physician-assessing physicians. Under this umbrella falls peer assessment, self-assessment, and in the case of trainees, faculty staff assessment.⁵ Over time, non-physicians including other members of the healthcare team and members of the community (e.g., standardised patients) have been involved.

PHYSICIANS-ASSESSING PHYSICIANS

This method is the most widely utilised and accepted technique in evaluating medical and surgical trainees.⁵ The strength of this approach draws from an underlying belief that experts (i.e., physicians) are better able than non-experts to discriminate between the subtleties and intricacies of their field. This is supported by data which show that global

rating scales (which require the rater to understand of the content domain) are more accurate than checklists for assessing a technical skill when administered by expert.²²

SELF-ASSESSMENT

Self-assessment is theoretically appealing because it allows the learners to take ownership of their own educational process. In the medical community, continuing education is driven by the physician's assessment of their learning needs and is considered a professional requirement. Despite its appeal, however, results of self-assessment studies in higher education have been mixed. Several meta-analyses have concluded that students are poor to moderate judges of their own performance.^{23,24} Others, especially with respect to technical domains, show that self-assessment is quite reliable.²⁵

When reviewing the self-assessment literature, a number of factors must be considered. First, it is important to remember that the ability to assess and self-assess is not an innate skill. Although it would be convenient, it is unlikely that individuals are able to efficaciously use assessment methods without prior instruction in their use. In addition to this 'calibration,' the assessment method used must be rigorously validated. Only with this knowledge can the reader draw meaningful conclusions about the utility of self-assessment in higher education.

PEER-ASSESSMENT

Peer assessment has recently garnered attention as an assessment tool. While several exist, a useful definition of peer assessment is 'assessment of the work of others by people of equal status and power'.²⁶ While the potential development of learner maturity, critical skills and discernment are evident, there is some reluctance to initiate peer assessment as an assessment tool. Reasons for this reluctance include: learners' believing they do not have the necessary skills to assess others; tradition's dictating that the educator evaluates learners; interpersonal relationships' interfering with assessment; and a perceived unreliability of the results.²⁷ However, investigation from Belfast indicates that learners are able to make rational judgments about the work of their peers with correlation coefficients between peer and expert assessment being 0.89.²⁸ These results have been replicated in undergraduate as well as surgical and medical training. This indicates that peer assessment may have a significant role in assessment.²⁹⁻³²

NON-PHYSICIANS ASSESSING PHYSICIANS

Despite its intrinsic appeal, there are limitations to using content experts to examine. These include added expense, scheduling conflicts, and retention of experienced examiners due to changing interest and burnout. To address this, non-physicians have been increasingly utilised in assessment. As outlined by Wanzel, non-physicians have been utilised in both ward and OSCE settings in a cost effective and reliable manner after sufficient training.⁵ However, these settings rely predominantly on checklist ratings that may be inferior to global rating scales in the hands of experts.²²

WHEN SHOULD THE ASSESSMENT OCCUR?

Assessment can occur at any one, or all of three points during a course. It can occur at the beginning of a course (the pretest), during or throughout the course (continuous assessment), or at the end of the course (end of course assessment).⁴

The pretest is useful since it may indicate whether students have the necessary requisite knowledge to study the course, or it may indicate what those aspects of the course students have already grasped. Continuous assessment is useful because it allows a learners progress through a course to be monitored so that remediation may be taken if necessary. End of course evaluation affords measurement of the degree to which overall curriculum objectives have been met.

CONCLUSION

Assessment in medical education is a multi-faceted and dynamic process. While outwardly complex, sound appraisal is centered on basic principles that allow accurate, efficient and meaningful determinations of mastery. With this knowledge, physicians should feel confident in their ability to participate in trainee assessment, peer assessment, or self-assessment regardless of practice type, specialty, or intervals between assessing.

References

1. Epstein RM: Assessment in medical education. *N Engl J Med*; 2007; 356(4): 387-96.
2. Brown: Trends in Assessment. In: Harden R Hart IR MH, ed. *Approaches to the Assessment of Clinical Competence Part 1*. Dundee: Center for Medical Education; 1992.
3. Broadfoot: Editorial. *Assessment in Education*; 1994; 1: 3-10.
4. Harden R: How to Assess Students: An Overview. *Medical Teacher*; 1979; 1(2): 65-70.
5. Wanzel KR, Ward M, Reznick R: Teaching the surgical craft: From selection to certification. *Current Problems in Surgery*; 2002; 39(6): 573-659.
6. Turnbull J, Gray J, MacFadyen J: Improving in-training evaluation programs. *Journal General Internal Medicine*;

- 1998; 13(5): 317-23.
7. Brown S, Race, P and Smith, B: *500 Tips on Assessment*. London: Kogan Page; 1996.
8. *Assessing Learning in Australian Universities*. Centre for the Study of Higher Education, The University of Melbourne, 2002. (Accessed March 9, 2010)
9. Harden R: How to... Assess Clinical Competence- An Overview. *Medical Teacher*; 1979; 1(6): 289-96.
10. Gallagher AG, Ritter EM, Satava RM: Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. *Surgical Endoscopy*; 2003; 17(10): 1525-9.
11. Downing SM: Validity: on meaningful interpretation of assessment data. *Med Educ*; 2003; 37(9): 830-7.
12. Phillips S: Legal issues in performance assessment. *West's Education Law Quarterly*; 1993; 2(2): 329-58.
13. Frank JE: The CanMEDS 2005 physician competency framework. Better standards.
14. *Better care*. Ottawa: The Royal College of Physicians and Surgeons of Canada; 2005.
15. Hunt D: Functional and dysfunctional characteristics of the prevailing model of clinical evaluation systems in North American medical schools. *Academic Medicine*; 1992; 67(4): 254-9.
16. Miller: The assessment of clinical skills/competence/performance. *Academic Medicine*; 1990; 65 (Supplement): S63-S7.
17. MacRae H, Regher G, et. al.: A new assessment tool for senior surgical residents: the patient assessment and management examination. *Surgery*; 1997; 122: 335-43.
18. Martin R, Reznick, et al.: An objective structured assessment of technical skills (OSATS) for surgical residents. *The British Journal of Surgery*; 1997; 84: 273-8.
19. Rethans JJ, Norcini JJ, Baron-Maldonado M, et al.: The relationship between competence and performance: implications for assessing practice performance. *Medical Education*; 2002; 36(10): 901-9.
20. *Assessing Learning in Australian Universities*. Centre for the Study of Higher Education, The University of Melbourne, 2002. (Accessed March 9, 2010)
21. Gipps. *Quality in Teacher Assessment*. In: Harlen, ed. *Enhancing Quality in Assessment*: British Educational Research Association; 1994.
22. Turnbull J: What is... Normative Versus Criterion-Referenced Assessment. *Medical Teacher*; 1989; 11(2): 145-50.
23. Regehr G MH, Reznick RK, Szalay D: Comparing the Psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Academic Medicine*; 1998; 73: 993-7.
24. Falchikov B: Student self-assessment in higher education: a meta-analysis *Review of Educational Research*. 1989; 70: 395-430.
25. Gordon: A review of the validity and accuracy of self-assessments in health professions training. *Academic Medicine*; 1991; 66: 762-9.
26. Mandel LS, Goff BA, Lentz GM: Self-assessment of resident surgical skills: is it feasible? *American Journal of Obstetrics and Gynecology*; 2005; 193(5): 1817-22.
27. Brown, Bull, Pendlebury: *Assessing student learning in higher education*. Routledge, London; 1997.
28. Habeshaw G, Habeshaw: *53 Interesting Ways to Assess Your Students* (3rd edition). Melksham, U.K.: Cromwell Press; 1993.
29. Stefani: Peer, Self and Tutor Assessment: relative reliabilities. *Studies in Higher Education*; 1994; 19(1): 69-75.
30. Morton JB, Macbeth WA: Correlations between staff,

peer and self assessments of fourth-year students in surgery. *Medical Education*; 1977; 11(3): 167-70.

31. Linn BS, Arostegui M, Zeppa R: Peer and self assessment in undergraduate surgery. *The Journal of Surgical Research*; 1976; 21(6): 453-6.

32. Martin, Hodges, McNaughton: Using videotaped benchmarks to improve the self-assessment ability of family practice residents. *Academic Medicine*; 1998; 73: 1201-6.

33. Risucci DA, Tortolani AJ, Ward RJ: Ratings of surgical residents by self, supervisors and peers. *Surgery, gynecology & obstetrics*; 1989; 169(6): 519-26.

Author Information

Ashley Vergis, MD, MMed, FRCSC

Section of General Surgery, University of Manitoba

Krista Hardy, MD, MSc, FRCSC

Section of General Surgery, University of Manitoba