Assessment of Specialty Competence

Cees van der Vleuten, PhD
University of Maastricht, The Netherlands

Introduction

The assessment of clinical competence has a rich history of research and development. A traditional view on clinical competence that has long driven these developments is the notion that competence consists of a set of relative distinct characteristics or component and that each of these components can be best measured by different methods. Three components are discussed in this paper.

The *assessment of problem solving* has the longest history of development. It started with the use of written patient simulations in the sixties in which examinees were required to work through a patient problem. Their choices and responses throughout the patient problem were scored and taken as an indication of problem-solving ability. There were a number of complications with this type of assessment, but the most dominant one was that problem-solving performance was highly dependent on the problem content. Problem solving ability appeared to be less generic and independent as was assumed initially, but varied substantially from problem to problem. This has been termed the content specificity problem of clinical competence and has later been found in any form of assessment of clinical competence. A second problem was that these problem-solving measures correlated strongly with scores on regular multiple-choice tests. Again it appeared that problem solving was not an independent construct. In light of these research outcomes the development of problem solving proceeded with using more efficient testing formats. When performance varies across problem contexts it is necessary to assess many different contexts in order to achieve a reproducible performance estimate. More efficient testing formats were proposed that allowed the assessment of a larger sample of patient problems. Instead of working through one integral patient problem, these formats used short patient scenarios prompting examinees to the most pertinent decisions of each of the patient problems, allowing more time to assess more problems. This efficiency approach has been called the *key feature approach* to problem solving assessment.

Clinical skills assessment developed strongly with the introduction of the Objective Clinical Examination (OSCE) in the seventies. In the OSCE examinees are brought in a variety of simulated but highly authentic clinical situations. Each of these situations is called a station and examinees are required to rotate through each of these stations. Real or simulated patients (SPs) may be used to portray clinical problems, or, depending on the clinical situation, dummies, clinical models, manikins, etc. can be used. Examinee performance is directly observed and either instantaneously scored by a faculty examiner or by the SP following the patient contact.

OSCE assessment has developed into a true "OSCE-ology". The technology of running OSCEs, scoring performance, and setting standards, etc. has well progressed and is available

from a vast amount of literature. The OSCE is plagued by the content specificity problem and therefore quite a few stations are required for obtaining reproducible scores. An increasing tendency is noted to assess clinical skills more authentically in an integrated fashion in longer lasting stations, rather than assessing isolated skills in shorter stations. This is particularly true for a postgraduate training context.

The *assessment of direct performance in practice* is still in its infancy. Traditional supervisor ratings lack reproducibility while indirect measures (chart audits, outcome measures, etc.) have unclear validity. A start is made with a few new assessment methods. In video assessment a video system records patient encounters in actual practice and a sample of these encounters are scored for assessing the quality of performance. In incognito standardized patient-based assessment real or simulated patients visit the practice of physicians. The physicians are unaware of the fact that they are visited by non-real patients (although they have given prior consent). The patients play their role in a standardized way and they are trained to score the physician's performance after the encounter. In patient ratings real patients are asked to complete a questionnaire following the clinical encounter. Since untrained lay people are involved the assessment typically focuses on communication aspects and issues of professionalism of the physician. Finally, clinical work samples are ratings of focused clinical activities that are directly observed in actual clinical practice and are completed by any health professional involved (clinician, nurse, peer, etc.). The ratings are simple and easy to complete. By collecting a sample of these ratings the multiplicity of occasions and raters will provide a reproducible outcome. The assessment technology to assess performance in practice is still an emerging one and more research and development is clearly needed in the near future.

The research that was associated with each of these components of assessment yielded a number of consistencies:

Regardless of the assessment format, performance badly generalizes from task to task (or item, station, case, observation, etc). To be able to infer reproducible conclusions from the assessment one therefore needs a large sample of tasks to be included in the assessment. This usually requires substantial testing time per examinee.

Once reproducible scores are obtained with a particular assessment method, another rather consistent finding was that scores across different methods correlate strongly (such as the finding between MCQs and problem solving simulations), particularly when they are intended to measure similar competencies. A distinction can be made between the stimulus format (the task presented to the examinee) and the response format (the way the response to the task is captured; e.g., MCQ, free response). The stimulus format is more important for what is actually measured by an assessment method than the response format. Professional authenticity of the stimulus format is the hallmark in any assessment method.

A final consistent finding was that every assessment action is followed by an educational reaction: assessment drives learning. Whatever is being assessed will determine what and how examinees will prepare and what they will learn. When a mismatch occurs between the objectives of the educational program and the assessment program, the assessment program will inevitably "win". Congruence between both types of objectives is therefore a mandatory requirement.

In conclusion, the traditional notion of competence as distinct, generic and stable set of components is refuted by empirical evidence. Competence is less generic, highly context dependent and more interrelated than we originally assumed.

Based on these research and development outcomes a number of practical suggestions can be made:

- Do not rely on short tests
- Sample broadly (across content, in time, examiners, patients)
- Consider efficiency in:
  - selection of test format
  - construction test items
- Don't be married to a single method (the method itself is less important than the content)
- Select the tasks which resemble clinical practice most (worry less about the competency being measured)
- Avoid complexity in simulations and scoring
- Verify the educational consequences of the assessment (evaluate the evaluation)
- Use assessment strategically to achieve desirable learning effects
- Consider assessment as an educational design problem

Further Reading

Assessment in general

Kane MT. The assessment of professional competence. Evaluation and the Health Professions 1992; 15: 163-82.

Hager P, Gonczi A, Athanasou J. General issues about assessment of competence. Assessment & Evaluation in Higher Education 1994; 19: 3-16.

Newble D, Dawson B, Dauphinee D et al. Guidelines for assessing clinical competence. Teaching and Learning in Medicine 1994: 6: 213-20.

Newble D, Jolly B, Wakeford R, eds. The certification and recertification of doctors: issues in the assessment of clinical competence. Cambridge: Cambridge University Press, 1994.

Miller GE. The assessment of clinical skills/competence/performance. Academic Medicine (Supplement) 1990; 65: S63-S7.

Vleuten, C.P.M. van der (1996) The assessment of professional competence: theoretical developments, research and practical implications. *Advances in Health Sciences Education, 1*, 41-67.

Problem solving assessment

Swanson D, Norcini J, Grosso L Assessment of clinical competence: Written and computer-based simulations. *Assessment and Evaluation in Higher Education 1987; 12: 220-46.*

OSCE

Vleuten, C.P.M. van der & Swanson, D.B. (1990) Assessment of clinical skills with standardized patients: State of the art. Teaching and Learning in Medicine, 2, 58-76.

Vu, N.V. & Barrows, H.S. Use of standardized patients in clinical assessments: recent developments and measurement findings. Educational Researcher, 23, 23-30.

Assessment in Practice

Rethans, J.J., Sturmans, F., Drop, M.J. & Vleuten, C.P.M. van der (1991) Assessment of performance in actual practice of general practitioners by use of standardised patients. British Journal of General Practice, 41, 97-99.

Ram, P., Grol, R., Rethans, J. J., Schouten, B., van der Vleuten, C. P. M., & Kester, A. (1999). Assessment of general practitioners by video observation of communicative and medical performance in daily practice: issues of validity, reliability and feasibility. Medical Education, 33(6), 447-454.