Value in workplace-based assessment rater training: psychometrics or edumetrics?

J Eric Jelovsek

Workplace-based assessment (WBA) remains an important method of evaluating competencies in medical training programmes. It also continues to be the most acceptable way to reassure the public that medical trainees achieve a minimum level of competence in the delivery of patient care. Unfortunately, one of the most consistent findings in measurements of clinical competence is the tremendous variability that occurs in the rating of trainee performance across a set of tasks. This is not unique to the health professions as inter-rater reliability has also been found to be a dominant source of variability in other areas of science,[1] and in law[2] and military activities.[3] In this issue of *Medical Education*, Kogan *et al.* attribute the reliability issues found in many WBA tools to assessors who tend to 'value different aspects of performance', 'lack a clear standard for judging performance' or 'rely on a gut or gestalt feeling' and conclude that assessors do not 'correctly' apply assessment criteria.[4] They set out to explore how two rater training interventions might provide explanations for the challenges of WBA in the hope of improving reliability. There is an ironic twist to their results. Having allowed faculty staff to create important assessment elements and to participate in the consensus building process required to improve rater reliability, Kogan *et al.*[4] concluded that the biggest benefit may be the amount that faculty staff learned about the assessment process and their incorporation of this learning into useful tools.

Cleveland, Ohio, USA

Correspondence: J Eric Jelovsek, Obstetrics, Gynecology and Women's Health Institute, Cleveland Clinic, 9500 Euclid Avenue A81, Cleveland, Ohio 44195, USA.
Tel: 00 1 216 444 2488;
E-mail: jelovsj@ccf.org

doi: 10.1111/medu.12763

Inter-rater reliability is also a dominant source of variability in other areas of science, and in law and military activities

Learning through assessment, sometimes referred to as the edumetric properties of assessment, is not a new phenomenon. Researchers and users of assessment tools are often intent upon measuring or understanding the psychometric properties of tools. However, many authors frequently overlook the impact of the driving force of the assessment process itself on learning. For example, it is common for faculty members to report that medical trainees direct their learning towards whatever they are assessed on and pay less attention to things in which they are unlikely to be assessed. Similarly, faculty staff coincidentally report the important role of the assessment process in facilitating learning, particularly when the tools are integrated into the educational culture of a programme. Unfortunately, few studies formally report these edumetric properties when studying assessment tools. In a recent systematic review[5] of the psychometric and edumetric properties of WBA tools used to assess psychomotor skills on live patients, use of a tool was shown to: (i) positively affect individual learner performance by resulting in a change in knowledge, skills or attitudes in only five studies investigating four tools; (ii) positively affect or change a programme curriculum in only three studies, and (iii) provide specific actionable results regarded as useful by learners in two studies covering two tools.[5]

Many authors frequently overlook the impact of the driving force of the assessment process itself on learning

Edumetric properties of assessment in the context of faculty development are less characterised. Fortunately, Kogan *et al.*[4] give us descriptions of many of these properties, despite their primary focus on training to reduce rater variability. Examples of these properties can be categorised into four ways in which assessment drives learning.[6]

Firstly, assessment drives learning through its content. Faculty participants in Kogan *et al.*'s study[4] preferred being active participants in developing the assessment standards:

> What participants reported as valuable about PDT [performance dimension training] was not the end product of a checklist that could be used mindlessly as a reductionist tool. Rather, the value was in the process of creating the framework to understand the component parts of a skill in order to

> develop a holistic shared mental model…[4]

The content participants were charged with creating and coming to consensus on shaped their own learning. Faculty staff also reported improving their own clinical skills. Kogan *et al.* describe how the process of developing and reviewing frameworks enabled faculty members to 'review, refresh, gain confidence in and reinforce previously learned skills […] thereby improving and updating [their] own clinical scripts'.[4] From these comments, there appears to be tremendous value in having faculty derive and apply assessment standards if they are to understand and buy into those standards for future use in assessment.

The content participants were charged with creating and coming to consensus on shaped their own learning

Secondly, assessment drives learning through its format. Kogan *et al.*'s participants described how they began to alter their own and the learner's format of assessment after the faculty training, and:

> …how PDT changed how they prepared themselves and their learners for direct observation. For them, it was helpful to have the frameworks available for use prior to direct observation, even when they were familiar with the framework skills. Reviewing frameworks before direct observation primed participants for their assessment role.[4]

One participant explained:

> …what I find most helpful is making sure I have all of my tools in front of me. […] I like having all the frameworks around me so that I can easily reference […] exactly what I'm looking for…[4]

Providers of faculty development may consider deliberately altering the format of their assessment procedures for the sole purpose of influencing the educational effect.

There is value in having faculty derive assessment standards if they are to understand and buy into those standards for future use in assessment

Thirdly, assessment drives learning through the information given: some faculty staff learned that sharing the assessment frameworks with their residents prepared the resident for observation during assessment. This gave information to the trainee on what ideal behaviour looks like. One participant said:

> The main thing I've done differently is to share a framework, what would make a terrific skill set, with the trainee before actually going in to do direct observation… Now, many times I will say: "Let's talk about the key elements for what we're about to do…"[4]

Faculty staff were given information on what competence is, which contributed to their own learning of how to define competence. One participant reported: 'I'm going to be able to have a more clear gold standard [on which] to compare my residents from now on. Having that definition of what competence is…'[4]

Assessors may want to consider using their assessment tools not to make decisions, but as instruments through which they can provide information and structure feedback to accomplish a learning exercise.

Providers of faculty development may consider altering the format of assessment procedures for the sole purpose of influencing the educational effect

Fourthly, assessments drive learning through their programming: van der Vleuten has described how the frequency, timing and regulations pertinent to trainee promotion are important elements in how assessment drives learning.[6] In the USA, formal Accreditation Council for Graduate Medical Education milestone reporting on trainee progress occurs every 6 months. Unfortunately, all resident assessments occur at this time and are in continuous competition with one another. These assessment requirements create bottlenecks in the programme during 'assessment time' by stretching the bandwidth of the faculty staff and resources required to perform accurate assessment. These comments were raised in Kogan *et al.*'s data as concerns for some faculty members that the elements being assessed were 'not realistic or feasible to achieve' in the time-limited clinical environment of a 20-minute office visit.[4] Many participants felt 'tension' in deviating from normative practices within their institution by considering residents to be learners and independent providers. In this context, many faculty members perceived assessment processes as hindering other important cultural or financial perspectives of medical training.

Assessors may want to consider using assessment tools not to make decisions, but as instruments through which they can provide information and structure feedback

Although psychometric properties such as reliability will always remain important aspects of any form of high-stakes assessment, Kogan *et al.*[4] demonstrate that medical educators should not forget that educational impact lies at the heart of assessment. Attempts to improve assessment through faculty development should also continue to consider assessment as part of the learning process in order to achieve the milestones set forth by the training programme. Faculty members appeared to observe these edumetric properties as having 'tremendous power'[4] to motivate residents to improve.

REFERENCES

1 Shavelson RJ, Baxter GP, Gao X. Sampling variability of performance assessments. *J Educ Meas* 1993;**30**:215–32.
2 Linn RL. Educational assessment: expanded expectations and challenges. *Educ Eval Policy Anal* 1993;**15** (1):1–16.
3 Shavelson RJ, Mayberry P, Li W, Webb NM. Generalisability of military performance measurements: marine corps rifleman. *Mil Psychol* 1990;**2**:129–44.
4 Kogan JR, Conforti LN, Bernabeo E, Iobst W, Holmboe E. How faculty members experience workplace-based assessment rater training: a qualitative study. *Med Educ* 2015;**49**:692–708.
5 Jelovsek JE, Kow N, Diwadkar GB. Tools for the direct observation and assessment of psychomotor skills in medical trainees: a systematic review. *Med Educ* 2013;**47**: 650–73.
6 van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ Theory Pract* 1996;**1** (1):41–67.