Workplace-Based Assessment and Assessment for Learning: Threats to Validity

MARJAN GOVAERTS, PhD

Over the past decades, competency-based education has become the norm in the medical education continuum. The rise of competency- or outcome-based models of education has been paralleled by an increasing interest in workplace-based assessment (WBA) to provide direct evidence of proficiencies of interest (ie, what the trainee will ultimately *do* in professional practice). WBA has become a cornerstone in the summative assessment *of* learning and professional competence. Competency-based education, however, also calls for a greater emphasis on formative assessment, or assessment *for* learning. Frequent feedback on performance should effectively steer and foster the acquisition of the necessary competencies. It therefore seems beyond dispute that assessment programs in competency-based medical education need to combine summative and formative assessment purposes.

In this issue of the *Journal of Graduate Medical Education*, the article by Donato et al,[1] "Validity and feasibility of the Minicard direct observation tool in 1 training program," illustrates how specific WBA tools may support both assessment functions. Findings from their study suggest that the Minicard direct observation tool facilitates identification of substandard performance and provides feedback to guide and stimulate residents' competence development. By obtaining evidence about the quality and quantity of feedback recorded on the Minicard, Donato and colleagues[1] aimed to develop a validity argument for the proposed use of the Minicard for formative assessment (ie, as a tool *for* learning).

Validity is essential in any assessment. In general, validity refers to the evidence that supports or refutes the interpretation and proposed use of assessment results. Results are more or less valid depending on what the intention was in using the assessment, at that particular point in time and for that particular population. Clearly, high-quality feedback is fundamental to formative assess-

ment. Feedback should not only include information about observed performance ("feed-back"), but also cues to directions for performance improvement in terms of performance goals and what needs to be done in order to achieve these goals (also known as feed-up and feed-forward, respectively).[2] However, if the main purpose of formative assessment is to stimulate further learning and use of feedback for performance improvement, one might argue that the key question to be addressed in the validity inquiry must be whether the assessment actually achieves these goals. Unfortunately, a wealth of research findings indicate that there is no simple answer to the questions of when, for whom, and for what feedback works.[3–5]

Research on WBA increasingly suggests that combining formative and summative assessment purposes in WBA programs is very difficult, some would say almost impossible to achieve. A recent study by Bok et al,[6] for instance, revealed that low-stakes formative assessments were increasingly being perceived as high stakes and summative, if these assessments were used not only to generate feedback but also as input into grading and pass-fail decisions.[6] In other words: Despite our best intentions, frequent formative assessments can very well be perceived as summative. This interferes with the acceptance and use of feedback. In fact, findings indicate that there are multiple threats to the validity of formative assessment, and some of the factors that compromise assessment *for* learning seem to be deeply engrained in medical training and WBA.

One of the factors to consider is our assumption that we can (and should) "measure" progress and learning outcomes, which leads to the focus on quantifiable assessment outcomes (eg, scores, grades). In WBA, assessment instruments typically require assessors to convert trainee performance into numerical scores, scoring levels, or grades on a performance rating scale. Space for narrative comments or written feedback, if present at all, tends to be limited. Obviously, grades and scores represent very poor feedback for learning (for instance, how does one interpret a "6" for communication?), but more importantly, research findings in (higher) education consistently indicate that a focus on grades may actually hinder learning and competence development. For example, grading student performance has been shown to diminish intrinsic motivation and

Marjan Govaerts, PhD, is Assistant Professor, Department of Educational Development and Research, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, the Netherlands.

Corresponding author: Marjan Govaerts, PhD, Maastricht University, Department of Educational Development and Research, PO Box 616, 6200 MD Maastricht, The Netherlands, +31 43 3885746, marjan.govaerts@maastrichtuniversity.nl

to reduce the quality of student learning. It may encourage students to focus on performance goals (''looking good'') rather than on learning, understanding, and mastery of the task. If grades are given, students will usually pay little attention to any supplemental feedback or comments, and they will not try to use comments for performance improvement.[7] It goes without saying that this finding is at odds with aspirations to develop our trainees to become self-regulated learners who are committed to excellence in patient care. To enhance the validity of WBA *for* learning, a shift from numbers to words, from rating scales to high-quality narrative evaluations, thus seems inevitable.

A second point refers to credibility as a key factor in the acceptance and use of feedback, and as a necessary condition for feedback to be perceived as meaningful.[8] Direct observation of trainee performance is a critical factor to enhance the credibility of feedback. Although the importance of direct observation in WBA is widely acknowledged, trainees are infrequently directly observed during clinical interactions with patients.[9] Some have argued that autonomy and independence, being core values in the current culture of medicine, may conflict with a learning culture that fundamentally values routine direct observation.[10] More importantly, however, the incorporation of direct observation into day-to-day working routines can be difficult because of competing demands, time pressure, and limited compensation for clinical teaching. For the same reasons, paperwork and recording of feedback may be delayed (sometimes for days to weeks after the assessment), resulting in feedback that is likely to be incomplete, inaccurate, and no longer meaningful.

Credibility of feedback is similarly dependent on the relationship between supervisor (feedback provider) and trainee (feedback receiver). Acceptance and use of feedback is enhanced in settings in which trainees and supervisors have been able to develop trusting professional relationships, in which trainees feel safe and confident that they and their supervisors are working together to achieve shared goals, in health care as well as in learning and competence development.[8,11] Fragmented learning, however, seems to be typical of many medical training programs. Trainees often rotate to different sites and learning settings and to different supervisors within short periods of time. The validity of WBA for formative purposes therefore calls attention to the way we organize supervision and learning in medical training. We not only need to create time for direct observation and feedback in crowded training schedules, we also may need to reorganize medical training to foster the development of extended trusting supervisor-trainee relationships in communities of learning and professional practice. In doing so, high levels of account-ability would be combined with high levels of psychological safety.

A third and related point is the need to engage our trainees in the assessment and feedback process. Without engagement, feedback is likely to be perceived as meaningless. While it is common knowledge that to learn, students must do more than just listen or simply read the information provided to them, we often assume that providing learners with information about what went well and what went poorly will automatically result in behavior change. Understanding, accepting, and using feedback, however, requires learners to review and reflect, comment on, and discuss feedback. Learners need to actively engage in setting their own learning goals, discussing performance requirements, and seeking feedback on task performance. Effective (valid) formative WBA therefore needs to conceptualize feedback as a *dialogue*, a 2-way process, rather than as 1-way transmission of information.[7,12]

So what does this mean for the validity of formative WBA? First of all, we need to realize that feedback processes are complex, and that the acceptance and use of feedback by learners is influenced by many complex and interrelated factors in seemingly unpredictable ways: More feedback does not necessarily imply more learning. The conclusion can be drawn, however, that effective formative WBA requires a change in culture. It will not work to structure and formalize feedback processes through the use of feedback forms or by mandating that learners collect and document feedback in their portfolios. As shown in the study by Bok and colleagues,[6] and similar studies in the United Kingdom,[13] this may actually result in general disappointment or cynicism about the utility of WBA *for* learning. Rather, we should invest in a feedback culture in which our trainees' learning and competence development drives the assessment, and in which both trainees and supervisors are committed to high-quality feedback as the basis for high-quality patient care.

References

1 Donato AA, Park YS, George DL, Schwartz A, Yudkowsky R. Validity and feasibility of the Minicard direct observation tool in 1 training program. *J Grad Med Educ*. 2015;7(2):225–229.

2 Ramaprasad A. On the definition of feedback. *Behav Sci*. 1983;28(1):4–13.

3 Kluger AN, DeNisi A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychol Bull*. 1996;119(2):254–284.

4 Hattie J, Timperley H. The power of feedback. *Rev Educ Res*. 2007;77(1):81–112.

5 Shute VJ. Focus on formative feedback. *Rev Educ Res*. 2008;78(1):153–189.

6 Bok H, Teunissen P, Favier R, Rietbroek N, Theyse L, Brommer H, et al. Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Med Educ*. 2013;13(1):123.

7 Nicol DJ, Macfarlane-Dick D. Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Stud High Educ*. 2006;31(2):199–218.

8 Watling C, Driessen E, van der Vleuten C, Lingard L. Learning from clinical work: the roles of learning cues and credibility judgements. *Med Educ*. 2012;46(2):192–200.

9 Hauer KE, Holmboe ES, Kogan JR. Twelve tips for implementing tools for direct observation of medical trainees' clinical skills during patient encounters. *Med Teach*. 2011;33(1):27–33.

10 Watling C, Driessen E, van der Vleuten CP, Vanstone M, Lingard L. Music lessons: revealing medicine's learning culture through a comparison with that of music. *Med Educ*. 2013;47(8):842–850.

11 Govaerts MJ, van der Vleuten CP, Schuwirth LW, Muijtjens AM. The use of observational diaries in in-training evaluation: student perceptions. *Adv Health Sci Educ Theory Pract*. 2005;10(3):171–188.

12 Watling CJ, Kenyon CF, Zibrowski EM, Schulz V, Goldszmidt MA, Singh I, et al. Rules of engagement: residents' perceptions of the in-training evaluation process. *Acad Med*. 2008;83(suppl 10):97–100.

13 Bindal T, Wall D, Goodyear HM. Trainee doctors' views on workplace-based assessments: are they just a tick box exercise? *Med Teach*. 2011;33(11):919–927.