

Assessment, feedback and the alchemy of learning

Christopher J Watling¹  & Shiphra Ginsburg²

CONTEXT Models of sound assessment practices increasingly emphasise assessment's formative role. As a result, assessment must not only support sound judgements about learner competence, but also generate meaningful feedback to guide learning. Reconciling the tension between assessment's focus on judgement and decision making and feedback's focus on growth and development represents a critical challenge for researchers and educators.

METHODS We synthesise the literature related to this tension, framed around four trends in education research: (i) shifting perspectives on assessment; (ii) shifting perspectives on feedback; (iii) increasing attention on learners' perceptions of assessment and feedback, and (iv) increasing attention on the influence of culture on assessment and feedback. We describe factors that produce and sustain this tension.

RESULTS The lines between assessment and feedback frequently blur in medical

education. Models of programmatic assessment deliberately use the same data for both purposes: low-stakes individual data points are used formatively, but then are added together to support summative judgements. However, the translation of theory to practice is not straightforward. Efforts to embed meaningful feedback in programmes of learning face a multitude of threats. Learners may perceive assessment with formative intent as summative, restricting their engagement with it as feedback, and thus diminishing its learning value. A learning culture focused on assessment may limit learners' sense of safety to explore, to experiment, and sometimes to fail.

CONCLUSIONS Successfully blending assessment and feedback demands clarity of purpose, support for learners, and a system and organisational commitment to a culture of improvement rather than a culture of performance.

Medical Education 2019; 53: 76–85
doi: 10.1111/medu.13645



¹Department of Clinical Neurological Sciences, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada

²Department of Medicine, University of Toronto, Toronto, Ontario, Canada

Correspondence: Christopher J Watling, Office of Postgraduate Medical Education, Medical Sciences Building Room M103, Schulich School of Medicine and Dentistry, Western University, London, Ontario N6A 5C1, Canada. Tel: 00 1 519 661 2019; E-mail: Chris.watling@schulich.uwo.ca

INTRODUCTION

Expectations for assessment in medical education have never been higher. The pressure on assessment has long been immense as sound assessments of medical learners serve the public interest. However, increasingly, definitions of sound assessment practices highlight their potential to shape learning. Although assessment of learning (summative assessment) remains critical, assessment for learning (formative assessment) has assumed greater prominence. This dual purpose means that assessment information should be used both to adjudicate learning outcomes and to generate meaningful feedback to learners that they can use to shape their future performance. Feedback, in this equation, is the catalyst that transforms assessment into learning.^{1–3}

However, the emergence of learning from a cauldron of assessment and feedback can seem like alchemy. Conflating assessment and feedback can be problematic. Fundamentally, their purposes may seem at odds. Whereas feedback stresses development and learning, assessment stresses judgement and decision making. Mixing these distinct ingredients creates a risky recipe for learning. Managing that risk requires balancing the competing priorities and tensions related to assessment and feedback. The pressing need to reconcile these tensions is the product of a shifting research landscape that encompasses new ways of thinking about both assessment and feedback, increased scrutiny of the impact of our pedagogical approaches on learners, and heightened sensitivity to the imprint of culture on learning.

In this narrative review, we synthesise key literature related to assessment and feedback, with a particular focus on the points of convergence and divergence between these research conversations. The literature on assessment and feedback is vast and varied, and we have not attempted to be exhaustive in our coverage of that literature. Although our focus is the clinical learning environment in medicine, we draw on literature from across the medical education spectrum, and on selected work from outside the medical education domain.

THE SHIFTING ASSESSMENT CONVERSATION

The ‘assessment drives learning’ maxim is neither new nor specific to health professions education.

Since the 1950s, education researchers have cautioned teachers that the examinations they devise will influence the student learning that occurs.⁴ Although the negative influence of assessment on learning received much of the initial attention, the potential of assessment to have a positive influence on learner behaviour was soon recognised.⁴ Assessment, it seemed, could be used strategically.

In the late 1960s, a distinction was made between summative and formative assessment, giving specific language to the ‘for learning’ purpose of assessment.⁵ Refining the notion of formative assessment, Sadler outlined three necessary conditions for a learner to engage with formative assessment: (i) the learner needs an understanding of the standard of performance to which he or she aspires, (ii) the learner needs information about the gap between his or her performance and that standard, and (iii) the learner needs a range of strategies to remedy that gap.¹ Sadler stressed the central role of feedback in formative assessment, but also spoke of gradually freeing learners from dependence on the teacher by empowering them to grasp standards, make reflective comparisons with their own performance and take action.¹

Medical education’s embrace of formative assessment is a more recent phenomenon. Prior to the mid-1990s, medical education research was focused less on the potential learning value of assessments and more on their psychometric properties.⁶ Assessment’s role in predicting future professional performance was paramount. Assessment aimed to support decision making about learner selection, promotion and graduation to independent practice; given the high stakes of these goals, calls for research to enhance the psychometric robustness and predictive power of assessment strategies were made.⁷

In 1996, van der Vleuten published a groundbreaking utility model for assessment, a model that remains influential.⁸ At the heart of this model were notions of compromise and trade-off. Reliability, validity, educational impact, acceptability and cost were seen as dynamic elements of utility; improvements in one sphere might come at the expense of another. van der Vleuten challenged designers of assessments to use their tools and approaches strategically to reinforce desirable learning behaviour. His model reflected a shifting view of assessment’s purpose; it acknowledged that assessment might be meant as either a judgement or a facilitator of learning, depending on the

circumstances, and that different elements of the equation might deserve more or less attention based on an assessment's intended purpose. He cautioned, however, that compromise on educational impact should be minimised.⁸

In the over 20 years that have followed, leading assessment scholars have heeded this caution, proposing models of assessment that foreground its learning impact. Norcini et al.'s 2011 criteria for good assessment, for example, included the related elements of educational effect (the extent to which an assessment motivates those who take it to prepare in a way that is educationally beneficial) and catalytic effect (the extent to which assessment offers feedback to enhance learning), solidifying expectations that assessment should both support sound decision making and facilitate learner progress.⁹ Baartman et al.'s 2006 framework for evaluation of assessment programmes included both 'meaningfulness' and 'educational consequences' among its proposed quality measures.¹⁰

Just how useful is assessment as a catalyst for learning? Considerable research on test-enhanced learning supports the notion that regular injections of test-based assessment into a curriculum can enhance both learning and retention.^{11,12} Within medical education, testing of learners has been shown to positively impact their diagnostic accuracy,¹³ their clinical reasoning¹⁴ and their clinical application of knowledge.¹⁵ Whereas testing appears to enhance learning even in the absence of feedback, the addition of feedback to testing enhances its benefits by correcting errors and confirming correct responses. For this reason, calls for more frequent testing in medical education typically warn of the need to link tests with feedback to optimise their benefit.¹⁶ In the clinical settings that are so central to medical learning, however, the educational impact of regular testing in the form of workplace-based assessment is less clearly established, although learners and teachers may perceive a useful effect.¹⁷

This shift in the assessment conversation toward foregrounding its formative potential has played out against the backdrop of the competency-based medical education (CBME) movement. Competency-based medical education orients itself toward outcomes rather than processes, and thus demands rigorous and robust assessment.^{18,19} However, competency assessment is a complex business: trustworthy and holistic assessment of the behaviour of medical professionals requires a menu

of assessment tools, strategically and coherently arranged. In response to this requirement, the assessment conversation has begun to shift further, away from individual assessment tools and their psychometric properties and toward assessment programmes that link multiple strategies to produce fulsome portraits of learner performance.^{20,21} van der Vleuten's 2012 model of programmatic assessment reflects both of these shifts in thinking about assessment, and aspires to simultaneously optimise assessment for learning and assessment for decision making.²⁰

Approaches to programmatic assessment de-emphasise infrequent high-stakes summative assessments in favour of very frequent low-stakes formative assessments.^{20,22} Taken individually, these formative assessment events are intended to serve a developmental purpose for the learner: used mainly for their learning and feedback value, they help learners to refine skills and build knowledge. Taken together, however, the formative becomes summative. Like pieces of a mosaic, each individual formative assessment contributes to a bigger picture; the collection of pieces serves as the basis for judgements about competence. Schuwirth et al. highlight the principle of proportionality in programmatic assessment: whereas low-stakes decisions might be based on single data points, high-stakes decisions demand multiple data points.²³ Rarely, however, are data points zero stakes, a feature that may be troublesome in the translation of theory to practice.

THE SHIFTING FEEDBACK CONVERSATION

As the assessment paradigm has shifted, so too has the conversation about feedback in medical education. Although feedback has long been championed as foundational to learning,^{24,25} researchers have cautioned against taking its value for granted.^{26–28} Two meta-analyses of feedback interventions showed that although feedback does improve performance, its overall effect is modest rather than dramatic; in fact, in a substantial minority of studies included in these analyses, feedback was actually detrimental to performance.^{29,30} Hattie's analysis of influences on student learning showed that feedback was among the most impactful of interventions, but even there the effect size was highly variable.³¹ Feedback clearly can be a powerful influence on student learning, but needs to be used with great care in order to fulfil its potential.

Enabling consequential feedback is thus recognised as a dicey proposition; as a result, feedback research has become increasingly self-conscious, exploring both the elements that make feedback meaningful and the hazards that make it fail. To be effective, feedback should be timely, specific, actionable and task-oriented rather than person-oriented.³²

Learners are more likely to use feedback they perceive to be credible and constructive, and may discard feedback they see as lacking in these qualities.²⁶

Credible feedback is typically informed by direct observation of learner performance and originates from a trusted source; constructive feedback contains an action plan that offers direction and guidance for the learner on how to improve. Uptake and use of feedback are challenging if it threatens learner self-esteem or self-image,³⁹ or if it triggers strong, negative emotions.^{33,34}

Social, cultural and organisational influences can support or derail efforts to embed meaningful feedback in programmes of learning.³⁵ Learning environments that afford regular opportunities for teachers to observe learners are more likely to nurture good feedback; in clinical learning environments, however, the pattern is more often one of teachers and learners working in parallel, with limited moments of direct observation.³⁶

Education programmes that foster teacher–learner relationships also support quality feedback. Teacher–learner relationships based on trust create safety for learners to engage with feedback, which is especially important when that feedback is critical or corrective.^{37–39} Trust and safety do not simply materialise; they develop when there is clarity in the curriculum, for both the teacher and the learner, about the intent and goals of their interactions. Increasingly, the feedback literature advocates a coaching approach, which places the emphasis firmly on learner development.² Coaching models thrive on longitudinal, trusting relationships between teachers and learners, where both are clear about the developmental intent of the interaction.^{40,41}

However, clarity about the intent of teacher–learner interactions is often lacking in medical education. The lines between feedback and assessment blur: teachers are often asked to both assess and coach a learner at the same time and around the same task. The combining of these roles is, in part, a response to a cultural imperative: in medicine, teachers are accountable not only to learners, but also to the profession and to the public. This dual accountability makes our linking of the roles at least

understandable, even if it may not always be desirable. Yet, investing teachers with this difficult dual role evokes discomfort in both learners and the teachers themselves.^{42,43} Some have suggested that we should look for opportunities to divorce the roles of assessor and coach,^{36,44} and emerging evidence suggests that doing so may yield benefits in terms of perceived feedback quality.⁴⁰ However, models of programmatic assessment, if not implemented with care, risk blurring the lines between assessor and coach further, with their seeming embrace of the notion that the same information, and experiences, can serve both summative and formative purposes.

LEARNER PERCEPTIONS

Another key shift has been the increasing research focus on how learners experience, integrate and use both assessment and feedback. Researchers have explored, for example, how learners' goal orientation may influence their feedback-seeking behaviour,⁴⁵ and how learners' regulatory focus may influence their responses to feedback.⁴⁶ Learners interact with assessment based on their understanding of its purpose. Although educators may have a clear notion that some assessments are intended to be formative whereas others are intended to be summative, learners' perceptions do not always match educators' intentions. Medical learners often perceive assessment with formative intent as summative.^{42,47,48} That perception may then interfere with the intended educational effect; for example, feedback that feels like assessment may not be reflected upon in the same way, and may not drive learner development as hoped.⁴⁹ Bok et al.'s evaluation of a new competency-based curriculum in a veterinary medicine programme found that students had difficulty distinguishing their clinical supervisor's role of coach from that of assessor; the use of standard assessment forms as vehicles for feedback tended to emphasise the assessment role of the teacher, and students became more reluctant to engage with formative assessments as potentially useful learning events.⁴² Bok et al. cautioned that when all data points contribute to a final summative decision, learners tend to perceive all data points as primarily summative.⁴² That perception can be problematic; as Roberts warned, learners might seek out more lenient assessors or more straightforward cases in response to the knowledge that formative moments contribute to summative judgements.⁵⁰ Heeneman et al.'s exploration of student perceptions of programmatic assessment yielded

similar findings: formative assessment was perceived as summative, and elements of the programme that signalled summative assessment, such as pass/fail decisions, hindered students' acceptance of assessment as truly formative.⁵¹

Harrison et al. noted similar challenges with efforts to embed feedback into high-stakes assessments.⁴⁸ When feedback was made available to students after an OSCE, it was taken up the least by those who had barely passed the test, exactly the group with the most to gain. Furthermore, post-OSCE feedback was viewed primarily as relevant to future summative assessments, rather than as relevant to future practice. Harrison et al. blamed a dominant summative assessment paradigm for learners' general reluctance to engage with developmental feedback that arose from assessment events.⁴⁸

Recent research on direct observation further highlights the challenge of learner perceptions. Direct observation is fundamental to both trustworthy assessment and credible feedback. Learners, however, maintain some ambivalence about direct observation during their clinical education: they recognise that it has value, but also identify a certain discomfort with being observed that shapes how they respond.^{52–54} For learners, being observed typically feels like being assessed.⁵⁵ Because they feel as if they are being assessed, they may alter their routines by, for example, interacting differently with their patient or adopting a checklist-style approach to their history taking and physical examination as if they were performing in an examination setting rather than just doing their clinical work. These alterations to the routine may create a performance that feels inauthentic to the learner: a performance geared toward doing well on an assessment rather than serving as the basis for feedback or coaching. When the performance on which feedback is crafted feels inauthentic, the feedback may ring hollow.⁵⁵

For learners, engaging meaningfully with coaching involves elements of vulnerability and risk.⁵⁶ They must be willing to make themselves vulnerable, putting their fallibility, their weaknesses and their struggles on display in order to gain feedback that will enable them to tackle these challenges and refine their skills. They must be willing to risk trying new ways of doing things, even dismantling some of their habits in order to reconstitute sounder ways of working. They must feel safe to fail. Yet failing in an assessment setting feels distressing and dangerous. When an assessment system seems oriented toward

detecting and punishing incompetent individuals, Schuwirth and Ash wrote, learners will not feel supported to explore the boundaries of their knowledge,²² although this exploration is exactly the kind of learning we hope to encourage.

Programmatic assessment holds that this discomfort might be overcome by ensuring that the stakes for each individual assessment moment are extremely low. The stakes are raised only when these moments are summed together.²⁰ Because individual moments are not weighty for learners, the idea is that they will be able to engage with them in the formative spirit in which they were intended, comfortable with the vulnerability and risk that doing so entails. We caution, however, that stakes are in the eye of the beholder, and that low stakes may not feel so insignificant for learners, particularly if they know they need to accumulate a certain number of 'successful' low-stakes assessments in order to progress.

CULTURAL PERSPECTIVES

Research around assessment and feedback has not only shifted its perspective from teacher to learner, but also from individual learner to learning culture. Learners' individual experiences and perceptions are not readily untangled from the culture in which they learn; learners' notions of the stakes of an assessment, for example, are shaped as much by the culture in which they are learning as by their individual perceptions of the assessment experience. Sociocultural learning theories posit that learning is not an individual pursuit, at least not entirely. Learning is not solely the product of an individual's mind, but, rather, occurs as learners become part of a professional community, gradually adopting the practices, beliefs and values of that community.⁵⁷ Understanding how those practices and values shape the uptake, use and significance of assessment and feedback has begun to attract research attention.

Professions maintain distinct and sometimes singular approaches to educating their own. Shulman calls these approaches *signature pedagogies*; the routines that they embed tend to endure because they are deeply rooted in professional culture.⁵⁸ The education design decisions that comprise a signature pedagogy reflect the values of the profession. Shulman noted that signature pedagogies require periodic repair, particularly at moments of major change within professions.⁵⁹ Within medicine, the

rise of the patient safety movement, with its imperative to reduce errors, has provided a strong impetus to revisit our pedagogical routines, opening the door for new approaches such as CBME and programmatic assessment.^{60,61} These new approaches aim, in part, to support more consistent education outcomes, and to ensure that doctors entering unsupervised practice can really do all the fundamental things their jobs require.

Competency-based medical education and programmatic assessment demand attention to culture change, and to the system and organisational decisions that will catalyse that change. Eva et al. have called for an assessment system that recognises performance as continuous and dynamic, that guides learners while minimising threats to their self-concepts, and that embeds assessment in authentic work that reflects the practice domains into which learners are entering.³ Achieving this ideal is a tall order. Cultural challenges abound and are amplified, in fact, by our need to situate learning and assessment in the workplace.

Although the demonstrated value of test-enhanced learning adds weight to the ‘assessment drives learning’ ethos, the translation of this benefit to the clinical setting may not be straightforward. In most examples of effective test-enhanced learning, test performance is private and offers data that can shape learning by allowing a learner to calibrate his or her performance against a standard. Workplace-based assessments are different: performance is public and the stakes of underperformance for identity and professional confidence are higher. In the workplace, assessment can never be simply data; rather, workplace-based assessment is a complex social interaction.⁶²

One key element of this social interaction is the concept of *face*, which encompasses both the self-image an individual wishes to project (positive face) and his or her desire to be autonomous and unimpeded (negative face).⁶³ In social interactions people generally have a desire to enhance, or at least maintain, their face for themselves and others. We do this linguistically by using politeness strategies, such as by complimenting someone as a way to enhance their positive face, or apologising before asking a favour, to attend to their negative face.⁶³ In medical education, our written assessments of learners abound with politeness strategies, especially ‘hedging’, suggesting that we view assessment as a face-threatening act.⁶⁴ In a recent study of learners, a culture of ‘niceness’, although

appreciated on many levels, was also perceived as an impediment to honest feedback.⁶⁵ The strong desire to ‘save face’ serves a useful social purpose: to maintain harmony and to protect self-esteem. Thus, a tension is created between the necessity for politeness and the mandate to give and receive feedback. In this charged environment, we must ask whether our professional cultural practices are reinforcing the perception that feedback is high stakes, not low stakes, and that observation and coaching are actually assessment and judgement in disguise. Every observation, regardless of its intended stakes, may be an opportunity to lose face.

The widespread adoption of entrustment scales for observation-based assessment, including formative assessment, has been driven partially by the way the notion of trust resonates within medicine’s professional culture.⁶⁶ Supervisors may not always know exactly how to dissect a learner’s performance or determine how a learner stacks up against the expected standard for his or her stage of training, but they know whom they trust, or so goes the reasoning. Trust, however, may be ‘our most influential social judgement’.⁶⁷ Trust is meaningful in medicine because it relates to a willingness to share the heaviest responsibility in the profession: the responsibility to provide safe care that does no harm. Although the notion of entrustment is thus very attractive as an assessment outcome, we must ask whether it is equally suitable as a language for our formative assessment processes. When we use entrustment scales as vehicles for feedback, the messages contained within that feedback may inadvertently threaten self-esteem and identity, threats that are well known to diminish feedback’s effectiveness. In fact, whenever learners perform tasks infused with significance for their identity, assessment based on those tasks may feel summative, even if the intent is formative.³

RECONCILING TENSIONS

Exciting research in recent years has moved medical education closer to an enlightened perspective on assessment and feedback. Robust assessment of learner competence and coaching for learner development are increasingly recognised as necessary partners in effective clinical education. The challenge is how to align these strategies so that they complement one another, rather than pulling in opposite directions. Attention to the cultural and organisational supports required to reconcile these tensions is mandatory.

Dannefer's description of one medical school's embrace of a new assessment culture highlights the organisational commitment required for this shift in thinking. Emphasising assessment for learning, and aiming for a shift in responsibility for learning from teacher to student, their examination- and grades-free approach was founded on four fundamental building blocks: (i) clear expectations in the form of well-articulated competency standards; (ii) frequent formative assessment; (iii) strong learner support in the form of a faculty advisor for each student, and (iv) learner responsibility, in the form of a learning portfolio that the learner maintains.⁵⁶ Students reported engaging with a reflective, developmental approach to learning, albeit after some initial culture shock.⁶⁸ On external measures of performance, such as the US Medical Licensing Examination (USMLE), students have performed as well or better than students in more traditional curricula.⁶⁹ As this example shows, culture change requires much more than adopting new assessment forms or enacting guidelines for programmatic assessment; it requires thoughtful consideration of the organisational supports necessary to drive and sustain change.

Learner support appears to be non-negotiable if we are to succeed in blending the decision-making and the developmental purposes of assessment within a single programme. However, what kind of support investments will yield the highest returns? Many authors have championed the value of coaching, academic advising or mentoring in bolstering the educational impact of assessment.^{2,3,19,70} Coaches or mentors guide learners in interpreting, integrating and acting upon assessment information, potentially accelerating learning. Faculty members require support as well: we must train coaches and mentors, rather than assuming that any well-intentioned teacher will suffice in these key roles. Training matters to outcomes. Sargeant et al., for example, have shown that training coaches in a model of conducting a reflective feedback conversation can improve acceptance and uptake of the feedback offered.⁷¹

Committing time and resources to learner support in the form of coaching and mentoring also supports the establishment of trusting, longitudinal relationships between teachers and learners. Within such relationships, more consistently useful feedback is exchanged,^{38,72} and learners may be less inclined to view feedback offered with formative intent as a summative threat. Supported by trusting relationships, learners may feel freer to fail. They may more readily adopt a desirable learning goal

orientation, rather than falling prey to a performance goal orientation, and their professional development may be stronger for it.⁴⁵

Learners' perceptions that they need to perform, in fact, represent a critical hindrance to meaningful learning. Programmatic assessment built on the notion that no assessment moments are zero stakes should be applauded for its explicitness and clarity but may inadvertently contribute to a culture of performance over a culture of improvement. If we are not careful, we may even risk requiring learners to 'perform' that they are good receivers of formative feedback in order to meet the requirements of their assessment programme.

We should ask if there is room in our curricula for zero-stakes moments of learning, discovery and experimentation. Examples exist already. Simulation in postgraduate training may be distinctly set up as not being about assessment, but rather as offering an opportunity for practice, with the benefit of post-event debriefing to support more effective learning.^{73,74} Coaching programmes in which faculty members pair with learners to observe them work, in order to offer reflections on what they have seen and to make recommendations for future development, can be successfully implemented in ways that allow learners to let their guard down and engage with the process of continuous improvement.⁴⁰

Perhaps the very term 'formative assessment' is problematic.⁵¹ For learners, assessment implies judgement and consequences. Words matter, and the impact of the word 'assessment' may offset the meaning of the 'formative' qualifier. Perhaps we are best to reserve 'assessment' for activities associated with judgement and decision making, and to use different and less threatening terminology to refer to activities intended to promote learner development. *Coaching*, for example, may capture formative assessment's collection and use of data about performance to drive teaching and learning while limiting learners' perceptions that they need to perform to succeed. The value of adopting coaching terminology may, in fact, be as much rhetorical as conceptual: the very word 'coaching' conjures images of activities that are driven by developmental goals.

CONCLUSIONS

Assessment may indeed drive learning, but must it be the only driver? We think not. Properly

harnessed, assessment strategies enhance learning, but a focus on assessment alone ignores the equally critical need for coaching strategies that foreground learning and development. Sound programmes of assessment reasonably aspire to provide both, but it is in the execution of this double-barrelled plan that tensions become evident. By grappling with these tensions, we can move productively toward new assessment approaches that encourage learner growth and development. The words of van der Vleuten et al. are helpful here: ‘... all actors in programmatic assessment must know what they are doing, why they are doing it, and why they are doing it this way.’²⁰ We must be explicit about the purpose of a teacher–learner interaction, the intent of any assessment or teaching tool we use, and the consequences for the learner in relation to how he or she performs. We must embrace and routinely reinforce an improvement model of learning and of working, so that performing confidently is replaced by striving for improvement as a guiding professional value.

Contributors: CJW conceptualised the paper, reviewed and analysed the literature relevant to the work, and wrote all drafts. SG participated in conceptualising the paper, added meaningfully to the analysis and presentation of the literature reviewed, and critically revised all drafts of the paper. Both authors gave final approval to the submitted paper. Both authors agree to be accountable for all aspects of the work and for ensuring its integrity.

Acknowledgements: none.

Funding: none.

Conflicts of interest: none.

Ethical approval: not required.

REFERENCES

- Sadler DR. Formative assessment and the design of instructional systems. *Instr Sci* 1989;**18** (2):119–44.
- Konopasek L, Norcini J, Krupat E. Focusing on the formative: building an assessment system aimed on student growth and development. *Acad Med* 2016;**91** (11):1492–7.
- Eva KW, Bordage G, Campbell C, Galbraith R, Ginsburg S, Holmboe E, Regehr G. Towards a program of assessment for health professionals: from training into practice. *Adv Health Sci Educ Theory Pract* 2016;**21** (4):897–913.
- Alderson JC, Wall D. Does washback exist? *Appl Linguist* 1993;**14** (2):115–29.
- Scriven M. The methodology of evaluation. In: Tyler R, Gagne R, Scriven M, eds. *Perspectives on Curriculum Evaluation*. AERA Monograph Series – Curriculum Evaluation. Chicago, IL: Rand McNally 1967;39–83.
- Hodges BD. Assessment in the post-psychometric era: learning to love the subjective and collective. *Med Teach* 2013;**35** (7):564–8.
- McGuire C. Perspectives in assessment. *Acad Med* 1993;**68** (2 Suppl):S3–8.
- 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.
- Norcini J, Anderson B, Bollela V et al. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. *Med Teach* 2011;**33** (3):206–14.
- Baartman LKJ, Bastiaens TJ, Kirschner PA, van der Vleuten CPM. The wheel of competency assessment: presenting quality criteria for competency assessment programs. *Stud Educ Eval* 2006;**32** (2):153–70.
- Roediger HL III, Karpicke JD. Test-enhanced learning: taking memory tests improves long-term retention. *Psychol Sci* 2006;**17** (3):249–55.
- Butler AC. Repeated testing produces superior transfer of learning relative to repeated studying. *J Exp Psychol Learn Mem Cogn* 2010;**36** (5):1118–33.
- Baghdady M, Carnahan H, Lam EWN, Woods NN. Test-enhanced learning and its effect on comprehension and diagnostic accuracy. *Med Educ* 2014;**48** (2):181–8.
- Raupach T, Andresen JC, Meyer K, Strobel L, Koziol M, Jung W, Brown J, Anders S. Test-enhanced learning of clinical reasoning: a crossover randomized trial. *Med Educ* 2016;**50** (7):711–20.
- Larsen DP, Butler AC, Lawson AL, Roediger HL III. The importance of seeing the patient; test-enhanced learning with standardized patients and written tests improves clinical application of knowledge. *Adv Health Sci Educ Theory Pract* 2013;**18** (3):409–25.
- Larsen DP, Butler AC, Roediger HL III. Test-enhanced learning in medical education. *Med Educ* 2008;**42** (1):959–66.
- Miller A, Archer J. Impact of workplace-based assessment on doctors’ education and performance: a systematic review. *BMJ* 2010;**341**:c5064.
- Lockyer J, Carraccio C, Chan M-K, Hart D, Smee S, Touchie C, Holmboe ES, Frank JR; ICBME Collaborators. Core principles of assessment in competency-based medical education. *Med Teach* 2017;**39** (6):609–16.
- Holmboe E. Work-based assessment and co-production in postgraduate medical training. *GMS J Med Educ* 2017;**34** (5):Doc58.
- van der Vleuten CPM, Schuwirth LWT, Driessen EW, Dijkstra J, Tigelaar D, Baartman LKJ, van Tartwijk J. A model for programmatic assessment fit for purpose. *Med Teach* 2012;**34** (3):205–14.
- Dijkstra J, van der Vleuten CP, Schuwirth LW. A new framework for designing programmes of assessment. *Adv Health Sci Educ Theory Pract* 2010;**15** (3):379–93.

- 22 Schuwirth L, Ash J. Assessing tomorrow's learners: in competency-based education only a radically different holistic method of assessment will work. Six things we could not forget. *Med Teach* 2013;**35** (7):555–9.
- 23 Schuwirth L, Valentine N, Dilena P. An application of programmatic assessment for learning (PAL) system for general practice training. *GMS J Med Educ* 2017;**34** (5):Doc56.
- 24 Ende J. Feedback in clinical medical education. *JAMA* 1983;**250** (6):777–81.
- 25 Cantillon P, Sargeant J. Giving feedback in clinical settings. *BMJ* 2008;**337**:a1961.
- 26 Watling CJ, Driessen E, van der Vleuten C, Lingard L. Learning from clinical work: the role of learning cues and credibility judgement. *Med Educ* 2012;**46** (2): 192–200.
- 27 Bing-You RG, Paterson J. Feedback falling on deaf ears: residents' receptivity to feedback tempered by sender credibility. *Med Teach* 1997;**19** (1):40–4.
- 28 Sargeant J, Mann K, Ferrier S. Exploring family physicians' reactions to multisource feedback: perceptions of credibility and usefulness. *Med Educ* 2005;**39** (5):497–504.
- 29 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):254–84.
- 30 Bangert-Drowns RL, Kulik C-LC, Kulik JA, Morgan MT. The instructional effect of feedback in test-like events. *Rev Educ Res* 1991;**61** (2):213–38.
- 31 Hattie JA. Influences on student learning. (Inaugural professorial address, University of Auckland, New Zealand, 1999). <http://www.education.auckland.ac.nz/webdav/site/education/shared/hattie/docs/influences-on-student-learning.pdf>. [Accessed 7 February 2018.]
- 32 Lefroy J, Watling C, Teunissen PW, Brand P. Guidelines on feedback for clinical education: the dos, don'ts, and don't knows of feedback for clinical education. *Perspect Med Educ* 2015;**4** (6):284–99.
- 33 Eva KW, Armson H, Holmboe E, Lockyer J, Loney E, Mann K, Sargeant J. Factors influencing responsiveness to feedback: on the interplay between fear, confidence, and reasoning processes. *Adv Health Sci Educ Theory Pract* 2012;**17** (1):15–26.
- 34 Sargeant J, Mann K, Sinclair D, van der Vleuten C, Metsemakers J. Understanding the influence of emotions and reflection upon multi-source feedback acceptance and use. *Adv Health Sci Educ Theory Pract* 2008;**13** (3):275–88.
- 35 Bowen L, Marshall M, Murdoch-Eaton D. Medical student perceptions of feedback and feedback behaviors within the context of the 'educational alliance'. *Acad Med* 2017;**92** (9):1303–12.
- 36 Watling CJ. Unfulfilled promise, untapped potential: feedback at the crossroads. *Med Teach* 2014;**36** (8):692–7.
- 37 Telio S, Ajjawi R, Regehr G. The 'educational alliance' as a framework for reconceptualizing feedback in medical education. *Acad Med* 2015; **90** (5):609–14.
- 38 Bates J, Konkin J, Suddards C, Dobson S, Pratt D. Student perceptions of assessment and feedback in longitudinal integrated clerkships. *Med Educ* 2013;**47** (4):362–74.
- 39 Watling C, Driessen E, van der Vleuten C, Vanstone M, Lingard L. Beyond individualism: professional culture and its influence on feedback. *Med Educ* 2013;**47** (6):585–94.
- 40 Voyer S, Cuncic C, Butler DL, MacNeil K, Watling C, Hatala R. Investigating conditions for meaningful feedback in the context of an evidence-based feedback programme. *Med Educ* 2016;**50** (9):943–54.
- 41 Watling C, Driessen E, van der Vleuten CPM, Lingard L. Learning culture and feedback: an international study of medical athletes and musicians. *Med Educ* 2014;**48** (7):713–23.
- 42 Bok HGJ, Teunissen PW, Favier RP, Rietbroek NJ, Theyse LFH, Brommer H, Haarhuis JCM, van Beukelen P, van der Vleuten CPM, Jaarsma DADC. Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Med Educ* 2013;**13**:123.
- 43 Blake JM, Norman GR, Smith EK. Report card from McMaster: student evaluation at a problem-based school. *Lancet* 1995;**345** (8954):899–902.
- 44 Cavalcanti RB, Detsky AS. The education and training of future physicians: why coaches can't be judges. *JAMA* 2011;**306** (9):933–4.
- 45 Teunissen PW, Stapel DA, van der Vleuten C, Scherpbier A, Boor K, Scheele F. Who wants feedback? An investigation of the variables influencing residents' feedback-seeking behavior in relation to night shifts. *Acad Med* 2009;**84** (7):910–7.
- 46 Watling CJ, Driessen E, van der Vleuten C, Vanstone M, Lingard L. Understanding feedback responses: the potential and limitations of regulatory focus theory. *Med Educ* 2012;**46** (6):593–602.
- 47 Harrison C, Wass V. The challenge of changing to an assessment for learning culture. *Med Educ* 2016;**50** (7):704–6.
- 48 Harrison CJ, Könings KD, Schuwirth LWT, Wass V, van der Vleuten CPM. Changing the culture of assessment: the dominance of the summative assessment paradigm. *BMC Med Educ* 2017;**17**:73.
- 49 Driessen E, Overeem K, van Tartwijk J. Learning from practice: mentoring, feedback, and portfolios. In: Dornan T, Mann K, Scherpbier A, Spencer J, eds. *Medical Education: Theory and Practice*. Edinburgh: Churchill Livingstone-Elsevier 2010;211–28.
- 50 Roberts T. Assessment est mort, vive assessment. *Med Teach* 2013;**35** (7):535–6.
- 51 Heeneman S, Oudkerk Pool A, Schuwirth LWT, van der Vleuten CPM, Driessen EW. The impact of programmatic assessment on student learning: theory versus practice. *Med Educ* 2015;**49** (5):487–98.
- 52 Watling C, LaDonna K, Lingard L, Voyer S, Hatala R. 'Sometimes the work just needs to be done':

- sociocultural influences on direct observation in medical training. *Med Educ* 2016;**50** (10):1054–64.
- 53 Ross V, Mauksch L, Huntington J, Beard JM. Interdisciplinary direct observation: impact on precepting, residents, and faculty. *Fam Med* 2012;**44** (5):318–24.
- 54 Madan R, Conn D, Dubo E, Voore P, Wiesenfeld L. The enablers and barriers to the use of direct observation of trainee clinical skills by supervising faculty in a psychiatry residency program. *Can J Psychiatry* 2012;**57** (4):269–72.
- 55 LaDonna K, Hatala R, Lingard L, Voyer S, Watling C. Staging a performance: learners' perceptions about direct observation during residency. *Med Educ* 2017;**51** (5):498–510.
- 56 Dannefer EF. Beyond assessment of learning toward assessment for learning: educating tomorrow's physicians. *Med Teach* 2013;**35** (7):560–3.
- 57 Lave J, Wenger E. *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press 1991.
- 58 Shulman LS. Signature pedagogies in the professions. *Daedalus* 2005;**134** (3):52–9.
- 59 Shulman LS. Pedagogies of uncertainty. *Lib Educ* 2005;**91** (2):18–25.
- 60 Holmboe ES, Sherbino J, Englander R, Snell L, Frank JR; ICBME Collaborators. A call to action: the controversy of and rationale for competency-based medical education. *Med Teach* 2017;**39** (6):574–81.
- 61 Makary MA, Daniel M. Medical error – the third leading cause of death in the US. *BMJ* 2016;**353**:i2139.
- 62 Govaerts M, van der Vleuten CP. Validity in work-based assessment: expanding our horizons. *Med Educ* 2013;**47** (12):1164–74.
- 63 Brown P, Levinson SC. *Politeness: Some Universals in Language Usage*. New York, NY: Cambridge University Press 1987.
- 64 Ginsburg S, van der Vleuten C, Eva KW, Lingard L. Hedging to save face: a linguistic analysis of written comments on in-training evaluation reports. *Adv Health Sci Educ Theory Pract* 2016;**21** (1):175–88.
- 65 Ramani S, Post S, Könings K, Mann K, Katz JT, van der Vleuten C. 'It's just not the culture': a qualitative study exploring residents' perceptions of the impact of institutional culture on feedback. *Teach Learn Med* 2017;**29** (2):153–61.
- 66 Rekman J, Gofton W, Dudek N, Gofton T, Hamstra SJ. Entrustability scales: outlining their usefulness for competency-based clinical assessment. *Acad Med* 2016;**91** (2):186–90.
- 67 Gingerich A. What if the 'trust' in *entrustable* were a social judgement? *Med Educ* 2015;**49** (8):750–2.
- 68 Altahawi F, Sisk B, Poloskey S, Hicks C, Dannefer EF. Student perspectives on assessment: experience in a competency-based portfolio system. *Med Teach* 2012;**34** (3):221–5.
- 69 Bierer SB, Dannefer EF, Taylor C, Hall P, Hull AL. Methods to assess students' acquisition, application, and integration of basic science knowledge in an innovative competency-based curriculum. *Med Teach* 2008;**30** (7):e171–7.
- 70 Driessen E, Scheele F. What is wrong with assessment in postgraduate training? Lessons from clinical practice and educational research. *Med Teach* 2013;**35** (7):569–74.
- 71 Sargeant J, Lockyer J, Mann K et al. Facilitated reflective performance feedback: developing an evidence- and theory-based model that builds relationship, explores reactions and content, and coaches for performance change (R2C2). *Acad Med* 2015;**90** (12):1698–706.
- 72 Hauer KE, Oza SK, Kogan JR, Stankiewicz CA, Stenfors-Hayes T, ten Cate O, Batt J, O'Sullivan PS. How clinical supervisors develop trust in their trainees: a qualitative study. *Med Educ* 2015;**49** (8):783–95.
- 73 Eppich W, Cheng A. Promoting excellence and reflective learning in simulation (PEARLS): development and rationale for a blended approach to health care simulation debriefing. *Simul Healthc* 2015;**10** (2):106–15.
- 74 Rudolph JW, Simon R, Dufresne RL, Raemer DB. There's no such thing as 'nonjudgmental' debriefing: a theory and method for debriefing with good judgment. *Simul Healthc* 2006;**1** (1):49–55.

Received 10 February 2018; editorial comments to authors 4 April 2018; accepted for publication 24 May 2018