

Managing tensions in assessment: moving beyond either–or thinking

Marjan J B Govaerts,¹  Cees P M van der Vleuten¹ & Eric S Holmboe²

CONTEXT In health professions education, assessment systems are bound to be rife with tensions as they must fulfil formative and summative assessment purposes, be efficient and effective, and meet the needs of learners and education institutes, as well as those of patients and health care organisations. The way we respond to these tensions determines the fate of assessment practices and reform. In this study, we argue that traditional ‘fix-the-problem’ approaches (i.e. either–or solutions) are generally inadequate and that we need alternative strategies to help us further understand, accept and actually engage with the multiple recurring tensions in assessment programmes.

METHODS Drawing from research in organisation science and health care, we outline how the Polarity Thinking™ model and its ‘both–and’ approach offer ways to systematically leverage assessment tensions as opportunities to drive improvement, rather than as intractable problems. In reviewing the

assessment literature, we highlight and discuss exemplars of specific assessment polarities and tensions in educational settings. Using key concepts and principles of the Polarity Thinking™ model, and two examples of common tensions in assessment design, we describe how the model can be applied in a stepwise approach to the management of key polarities in assessment.

DISCUSSION Assessment polarities and tensions are likely to surface with the continued rise of complexity and change in education and health care organisations. With increasing pressures of accountability in times of stretched resources, assessment tensions and dilemmas will become more pronounced. We propose to add to our repertoire of strategies for managing key dilemmas in education and assessment design through the adoption of the polarity framework. Its ‘both–and’ approach may advance our efforts to transform assessment systems to meet complex 21st century education, health and health care needs.

Medical Education 2019; 53: 64–75
doi: 10.1111/medu.13656



This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

¹Department of Educational Development and Research, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, the Netherlands

²Accreditation Council for Graduate Medical Education, Chicago, Illinois, USA

Correspondence: Marjan J B Govaerts, Department of Educational Development and Research, Faculty of Health, Medicine and Life Sciences, Maastricht University, PO Box 616, Maastricht 6200 MD, the Netherlands. Tel: 00 31 43 388 5746;

E-mail: marjan.govaerts@maastrichtuniversity.nl

 INTRODUCTION

Ultimately, the purpose of health professions education (HPE) is to benefit the quality of health systems by transforming learners into qualified professionals who not only have achieved standards of competence that are acceptable to the profession and the community, but are first and foremost committed to excellence, lifelong learning and the ongoing advancement of the field and high-quality care.^{1,2} In HPE, as in any other education system, assessment is seen as fundamental to achieving these education goals.³ In order to meet the demands of rapidly changing health care systems and the expectations of the public, approaches to education and assessment have evolved radically over the past few decades. Recent approaches to assessment reform aim to align assessment practices with models of outcome-based or competency-based education (CBE).^{4–6} Within the framework of CBE, assessment typically focuses on fostering the development of professional competence and ensuring robust decision making about learners' and physicians' fitness for practice. Competence-based assessment systems are typically complex as they rely on programmes of assessment that include multiple methods (standardised as well as non-standardised), contexts and assessors, and must be embedded in highly complex education as well as health care systems.^{6,7} Assessment systems are thus bound to be rife with tensions as they must fulfil formative and summative assessment purposes, be efficient and effective, and meet the needs of learners, education institutes, patients and health care organisations. Addressing these competing demands requires systems that are standardised as well as authentic, that allow for control as well as trust, and that foster cultures that enable and value learning as well as high-quality performance. As increasing pressures to reform HPE accentuate these multiple polarities inherent in modern assessment programmes, the resulting tensions seem to become even more salient in times of scarcity (increasing demands for high-quality performance while reducing costs) and plurality (e.g. multiple perspectives on the 'what' and 'how' in education reform), as illustrated by ongoing debates about what is right about how to educate and assess health professionals.^{8,9} Scholarly papers criticising or favouring concepts of CBE and workplace-based assessment, and discussions about the role of subjectivity and qualitative assessment approaches may very often reflect competing yet coexisting goals, tasks and roles, as well as individuals' emotions and cognitive frames shaped by

cultural and contextual factors.^{10–13} The way we respond to these tensions may very well be a fundamental determinant of the fate of assessment practices and reform.

A common response to tensions in organisational systems is the application of the so-called 'contingency approach' whereby researchers seek to capture a multifaceted reality with a perfect, internally consistent theory, and practitioners look upon tensions as 'problems that need [to] and can be fixed'.^{14,15} Through the contingency lens, challenging tensions in assessment systems become problems that will be solved and disappear if we can agree upon the single right answer. We then typically start searching for 'if-then' insights, striving to identify under which conditions either A or B needs to be emphasised or selected (e.g. standardisation or authenticity; quantitative or qualitative assessment approaches). This approach typically results in 'either-or' discussions, enabling rational decision making about assessment design and implementation. For example, using findings from assessment research as well as careful analysis of the US medical education context, Hanson and colleagues wrote a powerful plea for replacing numbers (grades) with words (narratives) to enable the achievement of education and assessment goals.¹⁶ Similarly, assessment experts argue that different information management strategies are to be used depending on assessment purposes and rationale,⁶ that we should refrain from grading if we use assessment for formative purposes,^{17–20} that we should focus on standardised assessments and criteria-based grading in the assessment of learning in order to ensure that our graduates are equivalent and fit for practice,^{21–25} and that we should favour qualitative over quantitative assessment approaches if we want to capture professional competence.²⁶ Proposed solutions then typically include measures to 'overcome' barriers to the successful implementation of the assessment ideal or to provide arguments to illustrate the fallibility of others' views and thinking.

It is increasingly recognised, however, that traditional 'fix-the-problem' approaches to coping with multiple tensions are generally inadequate as they fail to *sustainably* address the quantity and complexity of polarities in organisations. Likewise, the widespread ongoing and recurring debates in the HPE community suggest that many assessment tensions do not represent problems that can be resolved, but polarities that need to be carefully managed. This is the core premise of the Polarity Thinking™ (Polarity Partnerships, LLC, Auburn,

CA, USA) approach. This approach may offer promising and alternative ways of responding to tensions in complex organisational settings and help us further understand, accept and actually engage with the enduring tensions we face in education and assessment systems.

The purpose of this paper was to outline how Polarity Thinking™ may help assessment reform move forward by systematically leveraging assessment polarities as opportunities to drive improvement rather than as intractable problems. Drawing from literature in organisation science^{14,15,27,28} and health care reform,^{29,30} we will present a brief overview of the key concepts in Polarity Thinking™, followed by the identification and categorisation of common tensions in assessment. We will finally describe how the model can be used to map and manage key polarities in assessment.

REFRAMING PROBLEMS AS POLARITIES

The foundational premise of Polarity Thinking™ is the view that tensions must be accepted in order to achieve long-term success and sustained transformation because they are inherent to human behaviours in complex, dynamic and ambiguous systems.^{15,27–30} Key principles underpinning this perspective are summarised in Box 1.

Basically, polarities are two (or more) values or alternative views (called ‘poles’) that may appear as opposite or competitive but are interdependent and need each other to achieve a goal neither can reach alone.²⁸ Each of the poles brings positive outcomes to the overarching goals to be achieved, whereas emphasising one pole at the expense of the other will result in negative outcomes. In organisational systems, the coexistence of these different yet interrelated poles is commonly experienced as ‘tension’. Polarity Thinking™ entails a ‘both–and’ mindset rather than ‘if–then’ thinking or ‘either–or’ solutions in managing these tensions or dilemmas. Within assessment systems, assessment for learning (formative) and assessment of learning (summative) are typical examples of polarities, creating tensions between subsystems with different (sub)goals, functions and expectations. We need both, however, to achieve our ultimate goals of ensuring robust judgements and high-quality learning for high-quality patient care. In assessment, there can be no formative assessment without summative assessment: formative assessment is always about identifying the gap between actual performance and expected

Box 1 Adapted from Johnson,²⁸ and Wesorick and Shaha³⁰

Basic polarity principles

- 1 Polarities are interdependent pairs of different values or points of view (so-called ‘poles’)
- 2 The different values or poles need each other over time to reach the higher goal neither can achieve alone, even though there is tension between them
- 3 Both poles bring positive outcomes or an ‘upside’
- 4 Both poles have a potential ‘downside’
- 5 If one pole is neglected, there will be negative outcomes
- 6 Polarities are intrinsic to complex systems and organisations. Polarities are not problems to be solved; tensions between poles are unavoidable and must be leveraged

performance or required standards, defined as summative assessment criteria. Summative assessment is thus always embedded in formative assessment, either implicitly or explicitly.³¹ Likewise, people will argue that valid, fair assessment of learning is not possible without ongoing formative assessment.

Reframing problems as polarities (or paradoxes) is fairly well established in organisation science and research shows that leaders and organisations that manage organisational tensions using this perspective outperform those that do not.^{32–34} More recently, Polarity Thinking™ models have been successfully introduced to optimise health care reform, as traditional change efforts – aiming at fixing a polarity as a problem to be solved – did not result in the sustainable achievement of desired outcomes.^{30,35} Likewise, we propose the Polarity Thinking™ perspective may enable us to face and manage increasingly complex conditions for the assessment reform that is necessary to meet the rapidly changing demands of learners, organisations and the public.

POLARITIES AND TENSIONS IN ASSESSMENT: CATEGORISATION AND DEBATES

Based on our collective experience with assessment and drawing from the assessment literature within

and outside the domain of HPE, we identified and categorised common polarities and tensions in assessment systems (Fig. 1). It is not our intent to provide a systematic review of assessment research; rather, the purpose of our paper is to highlight exemplars of specific assessment polarities and tensions in order to demonstrate how Polarity Thinking™ can assist in addressing current tensions in assessment (in relation to CBE in particular).

We defined core activities and elements in assessment systems (i.e. assessment *for* learning; assessment *of* learning; assessment structure and organisation; cultural context) and identified tensions that arise within and between these elements. In fact, many tensions in assessment systems seem to reflect the complex interrelationships and interactions among these core elements, not infrequently driven by increasing pressures for accountability and cost-effectiveness.

For many decades, the significance of formative assessment as a powerful driver for learning has been acknowledged.^{17,36–38} With its emphasis on feedback, learning processes and student engagement in assessment, it aims to develop students and trainees into efficient, lifelong learners. High-quality formative assessments direct attention to goals to be achieved and how to achieve them, and thus act as incentives to align study and practice with the needs of individual learners. *Assessment for learning* tensions may then become salient if the emergence of individual learning trajectories involves flexible curricula and the reorganisation of work processes, impacting not only on the (scarce) financial, material and administrative resources of schools and health care institutions, but also on the quality of patient care.^{39–41} The need to create flexible and dynamic learning while maintaining stable working routines in the interwoven systems of education and health care and competing goals of efficiency and effectiveness may then create conflict ('strain') and ambiguity regarding assessment strategies.

Assessment of learning has for long been the almost exclusive focus of assessment content and approaches; issues of accountability, fairness and equivalence in the context of summative assessment have dominated assessment development and research in HPE until very recently. Tensions within *assessment of learning* may stem from the plurality of stakeholders' views on how to define what needs to be assessed in order to ensure high-quality care.

Scholars, as well as practitioners, may hold different views on what constitutes 'professional competence'^{42,43} and on how conceptualisations of competence are to be translated into assessment requirements.⁴⁴ The complexity of patient care may drive tensions between the individual and the collective because individual competence needs to be assured although teamwork and organisation performance largely contribute to patient outcomes.^{45,46} Similarly, efforts to reform assessment in CBE may drive tensions between cost-effectiveness (focus on 'fitness for practice' – being 'good enough') and excellence^{47,48} and/or between holistic and analytic approaches to competence. Although it is widely acknowledged, for example, that 'professional competence is more than a demonstration of isolated competencies' and that professional competence is to be defined as the inseparable and integrated use of knowledge, skills, norms, values, judgement and reasoning to serve the health care needs of the community,⁴⁹ the development of professional competence may require the description and evaluation of separate competencies or competency domains to ensure intended outcomes for high-quality care.^{5,6,50,51}

Tensions when designing (*structuring and organising*) assessment will surface if the achievement of the desired outcome requires different competing, yet coexisting processes and systems. Common tensions operate between standardisation and authenticity, or between (external) control (e.g. regulatory bodies) and trust in local assessment expertise as a critical resource in assessment decisions. Competing demands of internal and external stakeholders may raise tensions between equivalency (such as through the implementation of national licensing tests) and alignment with the medical school's or workplace curriculum. Serious dilemmas may then arise in contexts in which efforts to ensure quality of assessment for accountability through top-down, prescriptive (national) mandates or tests are interpreted to signify loss of programme autonomy and professional integrity, whereas non-compliance may result in the loss of programme accreditation.^{52,53}

Obviously, well-described and well-researched tensions between assessment design and assessment *for* and *of* learning reflect competing demands of reliability, validity, feasibility and acceptance. These tensions surface in debates around qualitative versus quantitative assessment approaches,²⁶ the role of 'objectification' (objectivity) versus subjectivity^{54,55} and a psychometric versus 'edumetric' or

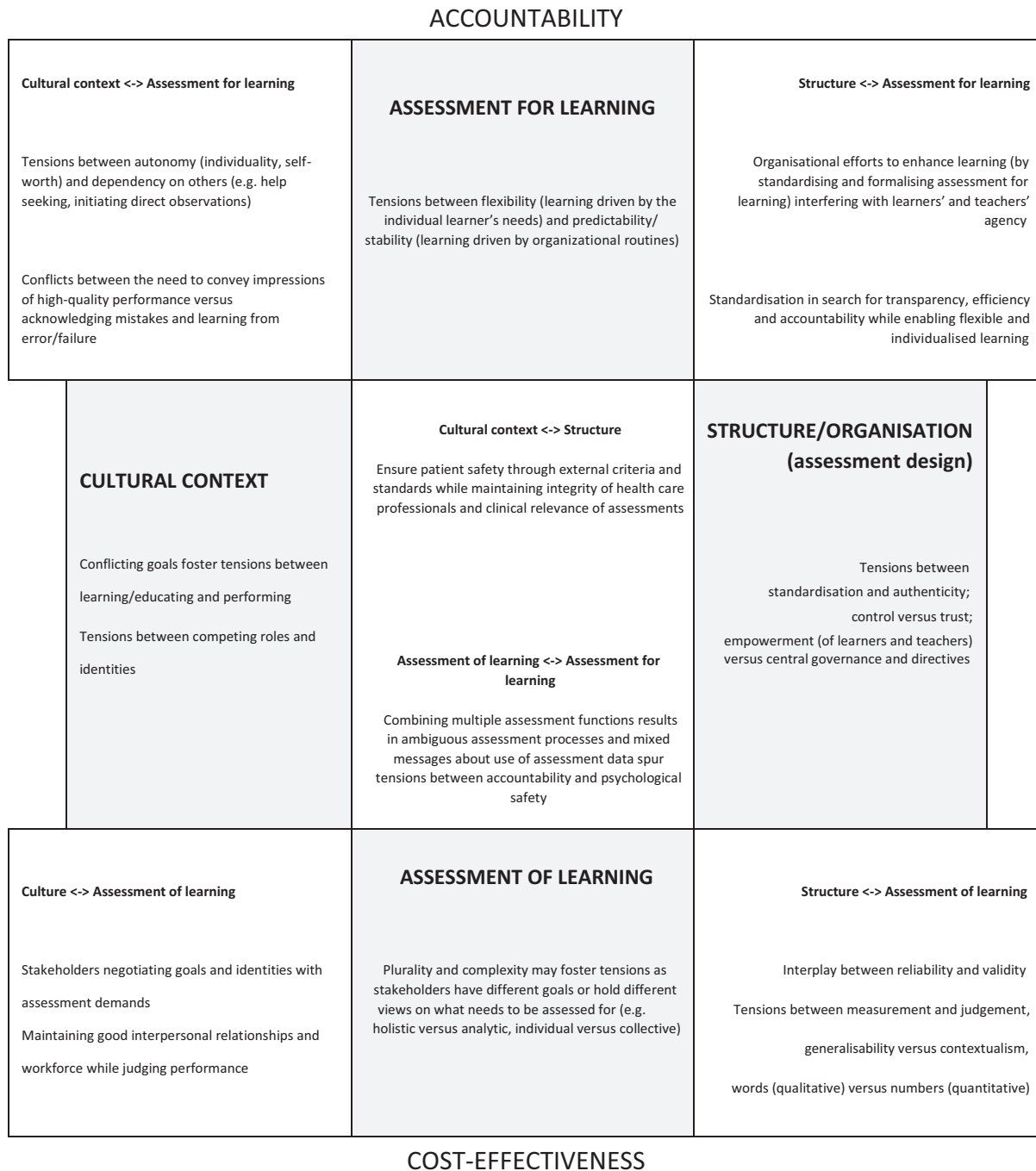


Figure 1 Categorisation of common assessment tensions (after Lewis and Smith¹⁵). Exemplars of key tensions that arise within and between core elements in assessment systems, driven by increasing pressures for accountability and cost-effectiveness. Shaded boxes present potential tensions within core elements, such as tensions within assessment for learning and assessment of learning, the structure and organisation of assessment (assessment design) and the cultural context in which assessment is embedded. The central and corner boxes present exemplars of tensions arising between these core elements

educational approach to assessment design.⁵⁶ At a more micro level, stakeholders' struggles with grades and narrative assessments, or the use of global rating scales and detailed checklists, illustrate efforts to navigate these tensions within specific

assessment contexts. The requirements of accountability and cost-effectiveness in particular may push towards objectified quantitative assessment outcomes, favouring grades and standardised assessments. As a consequence,

tensions may surface through differing, ambiguous and sometimes conflicting assessment strategies such as the ‘standardisation’ of inherently unstandardised assessments (e.g. standardised narratives for authentic assessments),^{57,58} ‘measurement’ of performance-in-context through the use of detailed rating scales and task-specific checklists,^{59,60} or the ‘objectification’ of competence and competence development through the use of entrustable professional activities and milestones.^{61,62}

In addition, education accountability may entail a ‘timekeeping function’ of assessment as assessment may (implicitly or explicitly) serve to increase the rate of learning: frequent assessments become ‘stopwatches’, controlling individual progress and the efficiency of learning processes.⁶³ Tensions between control and trust, or between standardisation and authenticity may then not only interfere with processes of teaching and learning in classroom settings or work environments,^{64–68} but also impact on stakeholders’ sense of autonomy, agency and engagement with assessment as a tool for learning.⁶⁹ Tensions between assessment design and assessment for learning may equally reflect conflicts between the need to enable individualised learning on the basis of meaningful feedback and the desire to structure and standardise in search of efficiency and accountability. Discussions about the use of numbers (grades) versus words (narratives) may be illustrative in this respect. In general, grades are considered to be the poorest form of feedback, whereas high-quality feedback is believed to be conveyed through the use of words. However, although the role of quality narratives in feedback for performance improvement is undisputed, research findings show that the standardisation of assessment and provision of grades that reflect progress towards competence may enhance learners’ sense of self-efficacy and competence, as well as perceptions of fairness and equivalence with respect to levels of competence upon graduation.^{20,24,70,71} Numbers can be very efficiently aggregated into individualised profiles of learners and learners’ progress, supporting self-assessment and inciting reflection and action for change.⁷² As Tekian et al. state: ‘Under appropriate circumstances, numbers can be more meaningful than a thousand pictures.’⁷² Various stakeholders may thus perceive the provision of numerical assessment data to be essential in maintaining a culture of excellence: numerical data enable the efficient and psychometrically sound ranking of learners, allowing learners to ‘objectively’

demonstrate excellence in a highly competitive environment and allowing stakeholders to defensibly select the ‘best-performing’ individuals for later admissions decisions. Decisions made when dealing with these tensions are thus inextricably linked with key characteristics of the cultural contexts in which assessment systems are embedded.

Clearly, health care and health care systems represent complex contexts, in which different organisational identities, goals, norms and values may drive tensions between learning and performance. Obviously, the tensions of competing goals in the delivery of high-quality, efficient patient care, as well as high-quality, efficient teaching and learning, are likely to become more salient under conditions of limited resources and increasing pressures for accountability.⁷³ In health care settings, learning–performance tensions may surface if an individual is required to fulfil the multiple roles of learner as well as health care provider (trainees), or health care provider, coach and assessor (clinicians). In workplace settings, trainees learn through and for work, and assessment of their learning inevitably involves judgement of the quality of their work. Critical feedback can thus feel uncomfortable to learners.⁷⁴ Similarly, assessors’ use of linguistic strategies, such as hedging, in attempts to maintain smooth social interactions and working relationships with their trainees while providing constructive feedback may reflect supervisors’ competing goals and conflicts between internal (towards learners) and external (towards institutions and the public) accountability.^{75,76} At the level of the education institution, the competing goals and demands of different stakeholders may result in similar conflicts between the actual use of assessment results and the assessment purposes communicated to teachers and learners. For instance, although management may support the role of teachers as the profession’s gatekeepers, it may at the same time tell staff that there is concern about the retention of students and the related funding of the institution or training programme.⁷⁷ In addition, low or below average ratings are considered unacceptable in some organisations (including in HPE and health care settings) even if they are accurate, and implicit organisational norms and pressures for conformity may be a significant factor in inflating ratings in clinical settings.^{78,79}

Interactions between the cultural context and assessment may also spur conflicts between

autonomy and independence – core values in the current culture of medicine – and a learning culture that fundamentally values routine direct observation and feedback seeking.^{73,74,80} Tensions between performance (looking good) and learning (being willing to show weaknesses, to admit and learn from failure) may additionally result in learners grappling with the need to engage in trustful relationships, interdependence and collaboration to foster development and expertise, while being competitive and seeking to outperform others in the battle for placements in residency training programmes or top-tier hospitals and health care organisations.^{81,82}

Similar types of tension may surface when integrating assessment for and of learning in assessment programmes. In programmatic assessment approaches, for instance, enhancing developmental assessment functions through frequent low-stakes performance evaluation and feedback conflicts with the summative use of these assessment data to ensure robust decision making. Learners, as well as teachers, may then face tensions between psychological safety and accountability when experiencing mixed messages, conflicting goals and ambiguous processes. As a

result, learners and teachers may perceive low-stakes assessments as high stakes, which impacts negatively on learning and learning strategies.^{69,83–85} Similar tensions may occur at organisational levels: while the enhancement of efficient, self-regulated lifelong learning is a pressing priority in education reform, accountability pressures may result in education institutions focusing on assessment for the purposes of selection, discipline and control.

Although by no means complete, this overview identifies many conflicting yet interrelated elements across a range of assessment phenomena, highlighting the complexity of assessment systems. Using Polarity Thinking™ might help us to sustainably manage these tensions.

POLARITY THINKING™: APPLICATION TO ASSESSMENT

As indicated in the previous section, increasing pressures for accountability and cost-effectiveness drive tensions that arise from the complex interplay among competing or conflicting goals, beliefs and values, each of which represent equally

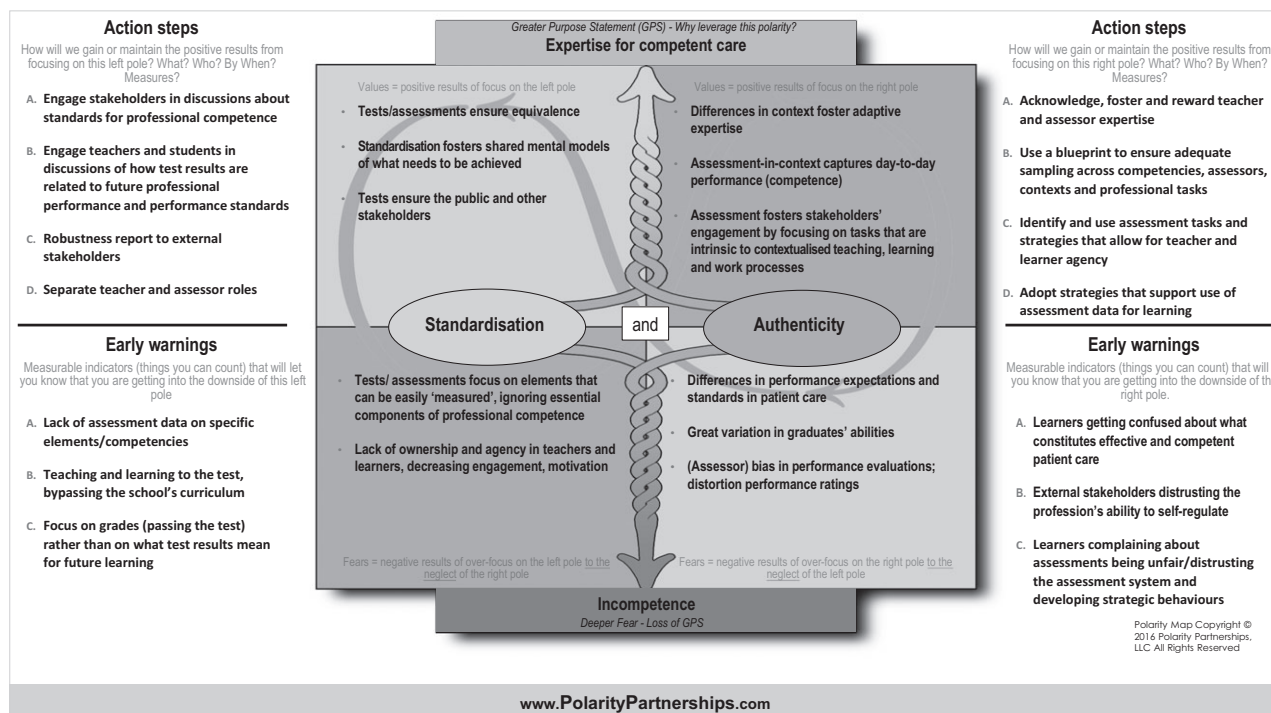


Figure 2 Example of a Polarity Map® for standardisation and authenticity. In this example, the poles are standardisation and authenticity. Several upsides (top cells) and downsides (bottom cells) are described and action steps suggested to maximise the upsides (benefits). Early warning signs are listed related to the downsides (potential harms)

valid alternative views of assessment reality. Tensions operate at and across different levels and stakeholders. For example, tensions between learning and performance – or related tensions of accountability and psychological safety – can exist at the level of the individual (learner, teacher) or the group (clinical workplace, clinical microsystem), as well as at the levels of programme management and the (education or health care) organisation. Attempts to manage key dilemmas through simple ‘either–or’ solutions or ‘if–then’ strategies are likely to have limited effectiveness (or might even fail), purely because simple ‘ifs’ may be hard to find in complex assessment systems. Rather, we would argue that the many chronic issues in assessment and assessment reform reflect multiple tensions between poles and polarities that are all necessary over time and should be managed to achieve sustainable and positive outcomes.

Use of the Polarity Thinking™ model to manage polarities involves three essential steps – seeing, mapping and tapping/leveraging – and requires the engagement of all key stakeholders in the process.^{28–30} The first step (*seeing*) is to know and accept that there are polarities, and to

identify them and understand how they work. Key polarities can be identified by reviewing the assessment literature (e.g. as presented in this paper) and by holding conversations with stakeholders about the challenges and dilemmas they face in day-to-day assessment practice. The next step (*mapping*) is to give a name to the different values or points of view (i.e. poles) underlying a dilemma and, through collaborative conversation, to identify the positive (upside) and negative (downside) outcomes associated with each of the poles. This will encourage stakeholders to explore the dilemma from multiple perspectives and as a whole. Step 3 (*tapping/leveraging*) involves engaging stakeholders in discussions about strategies or actions that need to be adopted to maximise the upsides of *both* poles while avoiding the downsides of each pole. At the same time, for each of the poles, early warning signs can be identified to indicate that one pole is being overly focused on to the neglect of the other. Polarity Thinking™ in assessment design or reform may thus facilitate buy-in from stakeholders through agreement upon strategies that must be adopted to achieve overarching goals and that accept and take multiple perspectives into account.

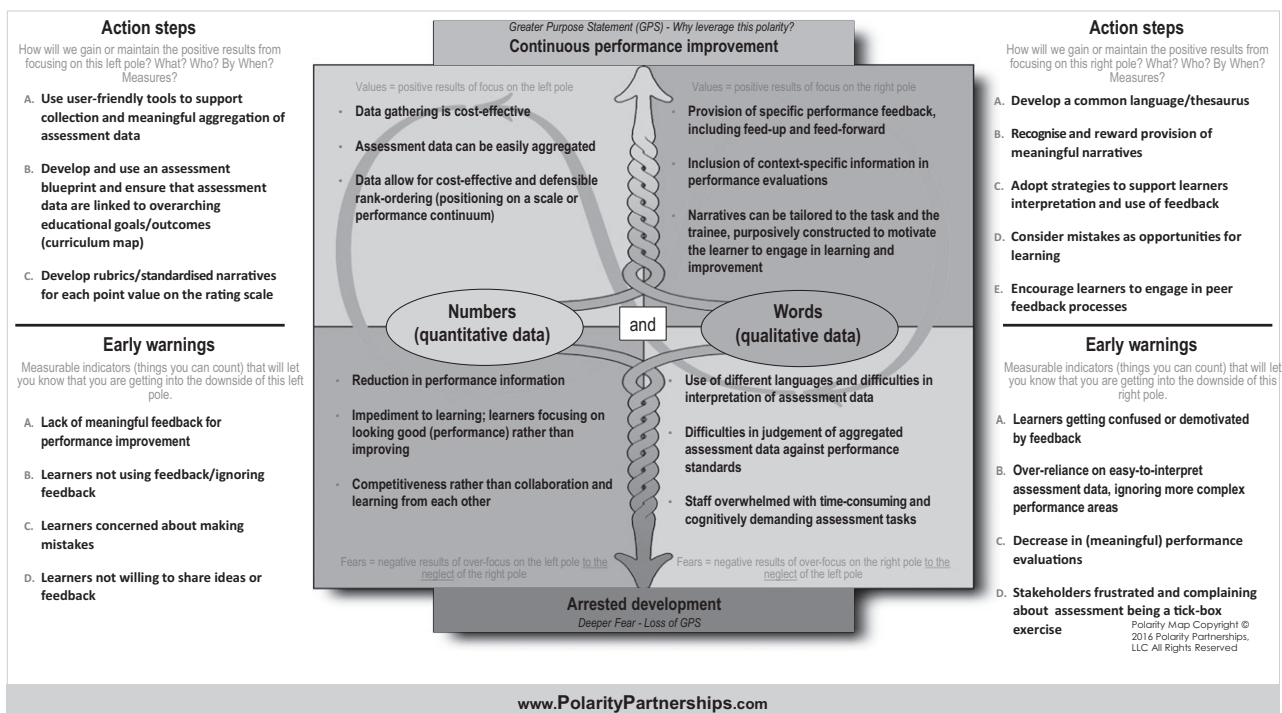


Figure 3 Example of a Polarity Map® for numbers (quantitative data) and words (qualitative data). For each pole, the upsides (top cells) and downsides (bottom cells) are described with suggested action steps to maximise the upsides (benefits) and early warning signs related to the downsides (potential harms)

Management of polarities can be supported by the use of a Polarity Map[®], which provides a structure for visualising and understanding polarities and how to address them. By definition, polarity maps are context-specific – to some extent at least – as the map needs to represent organisational reality. Figure 2 presents a (fictitious) example of a Polarity Map[®] for a common assessment polarity: the dilemma between standardisation and authenticity. Both these assessment approaches, although seemingly contradictory, are important in achieving the transformation of learners into high-quality health care professionals who are able to provide safe, high-quality patient care.

As the upper quadrants show, the poles of both standardisation and authenticity bring different, yet valuable, outcomes to the overarching purpose of ensuring health professionals are able to deliver high-quality care, whereby standardisation ensures equivalence, clarity of goals and performance standards, and authenticity fosters adaptive expertise and the capturing of contextualised, day-to-day performance. The figure also shows the potential limitations of each pole (lower quadrants) and the loss of positive outcomes when one pole is focused on and the other neglected (diagonal quadrants). The action steps describe what education organisations and stakeholders need to do to keep both poles strong, whereas early warning signs represent the ‘symptoms’ that may emerge if there is too much focus on either authentic or standardised assessment. The map in Fig. 3 is filled in to provide an example of how the polarity between quantitative (numbers) and qualitative (words) assessment approaches can be managed.

Polarity maps are not necessarily comprehensive, but, rather, focus on the key, most impactful elements of each pole, including both the upsides (i.e. benefits) and downsides (i.e. potential harms). Much like the failure mode and effects analysis used in quality improvement initiatives,^{86,87} polarity maps can be used proactively to anticipate the strengths and weaknesses of each pole when a training programme is planning to revise or implement an assessment, as well as to decide and agree upon strategies to maximise positive outcomes and monitor for early warning signs. By embracing both poles in assessment dilemmas, Polarity Thinking[™] may thus provide a pathway out of the ‘either–or’ tension landscape in HPE and facilitate the evolving of assessment programmes to meet complex 21st century health and health care needs.

CONCLUSIONS

Assessment polarities and tensions are likely to surface with the continued rise of complexity and change in education and health care organisations. With increasing pressures of accountability in times of stretched resources, assessment tensions and dilemmas will become more pronounced. Based on insights from organisation science and health care reform, we argue that we may need to add to our repertoire of strategies for managing key dilemmas in education and assessment design. The Polarity Thinking[™] model may serve as a useful perspective for examining and addressing tensions in complex assessment systems. The use of this model encourages individuals to accept and leverage multiple views, rather than to engage in futile discussions about who or what is right. We do not want to pretend that all problems in assessment are to be viewed through the lens of Polarity Thinking[™]. This will not replace all ‘if–then’ solutions, but it may help to address the recurrent and persistent dilemmas that so very often compromise assessment practices. Its ‘both–and’ approach may help researchers to clarify the tensions that exist, how they fuel debates and assessment reform, and how actors navigate these tensions to achieve desirable outcomes. In programmatic assessment, for example, research questions may address tensions arising from the use of low-stakes assessments for high-stakes purposes, and how these impact on learner and assessor behaviours. Similarly, increasing use of narratives in competence assessment may raise questions about when, why and how to use grades or performance scores. Likewise, the adoption of Polarity Thinking[™] and the reframing of problems in assessment from an ‘either–or’ to a ‘both–and’ perspective may help to advance our efforts to transform assessment systems to meet the needs of learners, education organisations and the public.

Contributors: MJBG, CPMvdV and ESH worked collaboratively to develop the primary content of this paper. MJBG wrote the initial draft of the manuscript. All authors contributed to revisions of the initial draft for intellectual content and clarity. All authors approved the final manuscript for publication.

Acknowledgements: the authors wish to thank Polarity Partnerships, LLC for granting permission for the use of the Polarity Map[®] (Polarity Thinking[™] and Polarity Map are registered trademarks of Polarity Partnerships, LLC. Copyright© 2016, all rights reserved). The authors would

also like to thank Reed Williams and Claire Touchie, and an anonymous reviewer for their valuable comments, which helped to improve the contents of the paper.

Funding: none.

Conflicts of interest: none.

Ethical approval: not applicable.

REFERENCES

- 1 Irby DM, Cooke M, O'Brien BC. Calls for reform of medical education by the Carnegie Foundation for the Advancement of Teaching: 1910 and 2010. *Acad Med* 2010;**85** (2):220–7.
- 2 Frenk J, Chen L, Bhutta ZA *et al.* Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. *Lancet* 2010;**376** (9756):1923–58.
- 3 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.
- 4 Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. *Med Teach* 2010;**32** (8):676–82.
- 5 Harris P, Bhanji F, Topps M, Ross S, Lieberman S, Frank JR, Snell L, Sherbino J; ICBME Collaborators. Evolving concepts of assessment in a competency-based world. *Med Teach* 2017;**39** (6):603–8.
- 6 Lockyer J, Carraccio C, Chan MK, 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.
- 7 Kogan JR, Conforti LN, Iobst WF, Holmboe ES. Reconceptualizing variable rater assessments as both an educational and clinical care problem. *Acad Med* 2014;**89** (5):721–7.
- 8 Boyd VA, Whitehead CR, Thille P, Ginsburg S, Brydges R, Kuper A. Competency-based medical education: the discourse of infallibility. *Med Educ* 2018;**52** (1):45–57.
- 9 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.
- 10 Krupat E. Critical thoughts about the core entrustable professional activities in undergraduate medical education. *Acad Med* 2018;**93** (3):371–6.
- 11 Fokkema JP, Scheele F, Westerman M, van Exel J, Scherpbier AJ, van der Vleuten CP, Dörr PJ, Teunissen PW. Perceived effects of innovations in postgraduate medical education: a Q study focusing on workplace-based assessment. *Acad Med* 2014;**89** (9):1259–66.
- 12 Ramani S, Post SE, 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.
- 13 Ginsburg S, McIlroy J, Oulanova O, Eva K, Regehr G. Toward authentic clinical evaluation: pitfalls in the pursuit of competency. *Acad Med* 2010;**85** (5):780–6.
- 14 Poole MS, Van de Ven AH. Using paradox to build management and organization theories. *Acad Manage Rev* 1989;**14** (4):562–78.
- 15 Lewis MW, Smith WK. Paradox as a metatheoretical perspective: sharpening the focus and widening the scope. *J Appl Behav Sci* 2014;**50** (2):127–49.
- 16 Hanson JL, Rosenberg AA, Lane JL. Narrative descriptions should replace grades and numerical ratings for clinical performance in medical education in the United States. *Front Psychol* 2013;**4**:668.
- 17 Black P, Wiliam D. Assessment and classroom learning. *Assess Educ* 1998;**5** (1):7–74.
- 18 Kohn A. The case against grades. *Educ Leadersh* 2011;**69** (3):28–33.
- 19 Matthews R, Noyes A. To grade or not to grade: balancing formative and summative assessment in post-16 teacher trainee observations. *J Furth High Educ* 2016;**40** (2):247–61.
- 20 O'Connor A, Cantillon P, McGarr O, McCurtin A. Navigating the system: physiotherapy student perceptions of performance-based assessment. *Med Teach* 2017;https://doi.org/10.1080/0142159x.2017.1416071. [Epub ahead of print.]
- 21 Wass V. Ensuring medical students are 'fit for purpose': it is time for the UK to consider a national licensing process. *BMJ* 2005;**331** (7520):791–2.
- 22 Swanson DB, Roberts TE. Trends in national licensing examinations in medicine. *Med Educ* 2016;**50** (1):101–14.
- 23 Bajammal S, Zaini R, Abuznadah W *et al.* The need for national medical licensing examination in Saudi Arabia. *BMC Med Educ* 2008;**8**:53.
- 24 Koczwara B, Tattersall M, Barton M *et al.* Achieving equal standards in medical student education: is a national exit examination the answer? *MJA* 2005;**182** (5):228–30.
- 25 Sadler DR. Interpretations of criteria-based assessment and grading in higher education. *Assess Eval High Educ* 2005;**30** (2):175–94.
- 26 Govaerts M, van der Vleuten CP. Validity in work-based assessment: expanding our horizons. *Med Educ* 2013;**47** (12):1164–74.
- 27 Smith WK, Lewis MW. Toward a theory of paradox: a dynamic equilibrium model of organizing. *Acad Manage Rev* 2011;**36** (2):381–403.
- 28 Johnson B. Reflections: a perspective on paradox and its application to modern management. *J Appl Behav Sci* 2014;**50** (2):206–12.
- 29 Wesorick BL. Polarity thinking: an essential skill for those leading interprofessional integration. *J Interprof Healthcare* 2014;**1** (1):12.
- 30 Wesorick B, Shaha S. Guiding health care transformation: a next-generation, diagnostic

- remediation tool for leveraging polarities. *Nurs Outlook* 2015;**63** (6):691–702.
- 31 Taras M. Summative assessment: the missing link for formative assessment. *J Furth High Educ* 2009;**33** (1):57–69.
- 32 Dodd D, Favaro K. *The Three Tensions: Winning the Struggle to Perform Without Compromise*. New York, NY: Wiley & Sons 2007.
- 33 McGrath RG. How the growth outliers do it. *Harv Bus Rev* 2012;**90** (1/2):110–6.
- 34 Heracleous L, Wirtz J. Singapore Airlines: achieving sustainable advantage through mastering paradox. *J Appl Behav Sci* 2014;**50** (2):150–70.
- 35 Wesorick B, Doebbeling B. Lessons from the field: the essential elements for point-of-care transformation. *Med Care* 2011;**49** (12 Suppl 1):49–58.
- 36 Crooks TJ. The impact of classroom evaluation practices on students. *Rev Educ Res* 1988;**58** (4):438–81.
- 37 Shute VJ. Focus on formative feedback. *Rev Educ Res* 2008;**78** (1):153–89.
- 38 Hattie J, Timperley H. The power of feedback. *Rev Educ Res* 2007;**77** (1):81–112.
- 39 Nousiainen MT, Caverzagie KJ, Ferguson PC, Frank JR; ICBME Collaborators. Implementing competency-based medical education: what changes in curricular structure and processes are needed? *Med Teach* 2017;**39** (6):594–8.
- 40 van Rossum TR, Scheele F, Sluiter HE, Bosman PJ, Rijksen L, Heyligers IC. Flexible competency based medical education: more time efficient, higher costs. *Med Teach* 2018;**40** (3):315–7.
- 41 van Rossum TR, Scheele F, Sluiter HE, Paternotte E, Heyligers IC. Effects of implementing time-variable postgraduate training programmes on the organisation of teaching hospital departments. *Med Teach* 2018; <https://doi.org/10.1080/0142159x.2017.1418850>. [Epub ahead of print.]
- 42 Hodges B. Medical education and the maintenance of incompetence. *Med Teach* 2006;**28** (8):690–6.
- 43 Morcke AM, Dornan T, Eika B. Outcome (competency) based education: an exploration of its origins, theoretical basis, and empirical evidence. *Adv Health Sci Educ Theory Pract* 2013;**18** (4):851–63.
- 44 Tekian A, Hodges BD, Roberts TE, Schuwirth L, Norcini J. Assessing competencies using milestones along the way. *Med Teach* 2015;**37** (4):399–402.
- 45 Lingard L. Rethinking competence in the context of teamwork. In: Hodges BD, Lingard L, eds. *The Question of Competence: Reconsidering Medical Education in the Twenty-First Century*. Ithaca, NY: Cornell University Press 2012;42–70.
- 46 Hodges B. Assessment in the post-psychometric era: learning to love the subjective and collective. *Med Teach* 2013;**35** (7):564–8.
- 47 Leung WC, Diwakar V. Learning in practice. Competency-based medical training: the baby is thrown out with the bathwater. *BMJ* 2002;**325** (7366):693–6.
- 48 Talbot M. Monkey see, monkey do: a critique of the competency model in graduate medical education. *Med Educ* 2004;**38** (6):587–92.
- 49 Epstein RM. Assessment in medical education. *N Engl J Med* 2007;**356** (4):387–96.
- 50 van den Eertwegh V, van Dalen J, van Dulmen S, van der Vleuten C, Scherpbier A. Residents' perceived barriers to communication skills learning: comparing two medical working contexts in postgraduate training. *Patient Educ Couns* 2014;**95** (1):91–7.
- 51 Egener BE, Mason DJ, McDonald WJ, Okun S, Gaines ME, Fleming DA, Rosof BM, Gullen D, Andresen ML. The charter on professionalism for health care organizations. *Acad Med* 2017;**92** (8):1091–9.
- 52 Peck CA, Gallucci C, Sloan T. Negotiating implementation of high-stakes performance assessment policies in teacher education: from compliance to inquiry. *J Teach Educ* 2010;**61** (5):451–63.
- 53 Price T, Lynn N, Coombes L, Roberts M, Gale T, Regan de Bere S, Archer J. The international landscape of medical licensing examinations: a typology derived from a systematic review. *Int J Health Policy Manag* 2018;**7** (9):782–90.
- 54 Norman GR, van der Vleuten CPM, Graaff E. Pitfalls in the pursuit of objectivity: issues of validity, efficiency and acceptability. *Med Educ* 1991;**25** (2):119–26.
- 55 Eva KW, Hodges BD. Scylla or Charybdis? Can we navigate between objectification and judgement in assessment? *Med Educ* 2012;**46** (9):914–9.
- 56 Schuwirth LW, van der Vleuten CPM. A plea for new psychometric models in educational assessment. *Med Educ* 2006;**40** (4):296–300.
- 57 Regehr G, Ginsburg S, Herold J, Hatala R, Eva K, Oulanova O. Using 'standardized narratives' to explore new ways to represent faculty opinions of resident performance. *Acad Med* 2012;**87** (4):419–27.
- 58 Rawlings A, Knox AD, Park YS, Reddy S, Williams SR, Issa N, Jameel A, Tekian A. Development and evaluation of standardized narrative cases depicting the general surgery professionalism milestones. *Acad Med* 2015;**90** (8):1109–15.
- 59 Cunnington JPW, Neville AJ, Norman GR. The risks of thoroughness: reliability and validity of global ratings and checklists in an OSCE. *Adv Health Sci Educ Theory Pract* 1996;**1** (3):227–33.
- 60 Igen JS, Ma IW, Hatala R, Cook DA. A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. *Med Educ* 2015;**49** (2):161–73.
- 61 Carraccio C, Englander R, Gilhooly J, Mink R, Hofkosh D, Barone MA, Holmboe ES. Building a framework of entrustable professional activities, supported by competencies and milestones, to bridge the educational continuum. *Acad Med* 2017;**92** (3):324–30.

- 62 Norman G, Norcini J, Bordage G. Competency-based education: milestones or millstones?. *J Grad Med Educ* 2014;**6** (1):1–6.
- 63 Kvale S. Contradictions of assessment for learning in institutions of higher learning. In: Boud D, Falchikov N, eds. *Rethinking Assessment in Higher Education: Learning for the Longer Term*. New York, NY: Routledge 2007;57–71.
- 64 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.
- 65 Massie J, Ali JM. Workplace-based assessment: a review of user perceptions and strategies to address the identified shortcomings. *Adv Health Sci Educ Theory Pract* 2016;**21** (2):455–73.
- 66 Ali JM. Getting lost in translation? Workplace-based assessments in surgical training. *Surgeon* 2013;**11** (5):286–9.
- 67 Harrison CJ, Könings KD, Schuwirth L, Wass V, van der Vleuten C. Barriers to the uptake and use of feedback in the context of summative assessment. *Adv Health Sci Educ Theory Pract* 2015;**20** (1):229–45.
- 68 Harden RM. Five myths and the case against a European or national licensing examination. *Med Teach* 2009;**31** (3):217–20.
- 69 Schut S, Driessen E, van Tartwijk J, van der Vleuten C, Heeneman S. Stakes in the eye of the beholder: an international study of learners' perceptions within programmatic assessment. *Med Educ* 2018;**52** (6):654–63.
- 70 Crocker J, Karpinski A, Quinn DM, Chase SK. When grades determine self-worth: consequences of contingent self-worth for male and female engineering and psychology majors. *J Pers Soc Psychol* 2003;**85** (3):507.
- 71 Nowacki AS. Making the grade in a portfolio-based system: student performance and the student perspective. *Front Psychol* 2013;**4**:155.
- 72 Tekian A, Watling CJ, Roberts TE, Steinert Y, Norcini J. Qualitative and quantitative feedback in the context of competency-based education. *Med Teach* 2017;**39** (12):1245–9.
- 73 Watling C, LaDonna KA, Lingard L, Voyer S, Hatala R. 'Sometimes the work just needs to be done': socio-cultural influences on direct observation in medical training. *Med Educ* 2016;**50** (10):1054–64.
- 74 Watling C, Driessen E, van der Vleuten CP, Lingard L. Learning culture and feedback: an international study of medical athletes and musicians. *Med Educ* 2014;**48** (7):713–23.
- 75 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.
- 76 Sebok-Syer SS, Klinger DA, Sherbino J, Chan TM. Mixed messages or miscommunication? Investigating the relationship between assessors' workplace-based assessment scores and written comments. *Acad Med* 2017;**92** (12):1774–9.
- 77 Hawe E. It's pretty difficult to fail: the reluctance of lecturers to award a failing grade. *Assess Eval High Educ* 2003;**28** (4):371–82.
- 78 Dudek NL, Marks MB, Regehr G. Failure to fail: the perspectives of clinical supervisors. *Acad Med* 2005;**80** (10):84–7.
- 79 Larocque S, Luhanga FL. Exploring the issue of failure to fail in a nursing program. *Int J Nurs Educ Scholarsh* 2013;**10** (1):115–22.
- 80 Watling C, Driessen E, van der Vleuten CP, Vanstone M, Lingard L. Beyond individualism: professional culture and its influence on feedback. *Med Educ* 2013;**47** (6):585–94.
- 81 Lempp H, Seale C. The hidden curriculum in undergraduate medical education: qualitative study of medical students' perceptions of teaching. *BMJ* 2004;**329** (7469):770–3.
- 82 Harrison CJ, Könings KD, Dannefer EF, Schuwirth LW, Wass V, van der Vleuten CPM. Factors influencing students' receptivity to formative feedback emerging from different assessment cultures. *Perspect Med Educ* 2016;**5** (5):276–84.
- 83 Bok HG, Teunissen PW, Favier RP, Rietbroek NJ, Theyse LF, Brommer H, Haarhuis JC, van Beukelen P, van der Vleuten CP, Jaarsma DA. Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Med Educ* 2013;**13** (1):123.
- 84 Heeneman S, Oudkerk Pool A, Schuwirth LW, van der Vleuten CPM, Driessen EW. The impact of programmatic assessment on student learning: theory versus practice. *Med Educ* 2015;**49** (5):487–98.
- 85 Harrison CJ, Könings KD, Schuwirth LW, Wass V, van der Vleuten CPM. Changing the culture of assessment: the dominance of the summative assessment paradigm. *BMC Med Educ* 2017;**17** (1):73.
- 86 Spath PL. Using failure mode and effects analysis to improve patient safety. *AORN J* 2003;**78** (1):16–37.
- 87 DeRosier J, Stalhandske E, Bagian JP, Nudell T. Using health care failure mode and effect analysis™: the VA National Centre for Patient Safety's prospective risk analysis system. *Jt Comm J Qual Improv* 2002;**28** (5):248–67.

Received 13 February 2018; editorial comments to author 16 April 2018, accepted for publication 8 June 2018