MEDICAL TEACHER
2019, VOL. 41, NO. 3, 243-248
https://doi.org/10.1080/0142159X.2018.1554898

The intersection of assessment, selection and professionalism in the service of patient care

Chris Roberts [a], Tim J. Wilkinson [b], John Norcini [c], Fiona Patterson [d], and Brian D. Hodges [e]

[a] Sydney Medical School, Faculty of Medicine and Health, University of Sydney, Sydney, Australia; [b] University of Otago, Christchurch, New Zealand; [c] FAIMER, Philadelphia, PA, USA; [d] Work Psychology Group, Derby UK; [e] University Health Network and University of Toronto, Toronto, Canada

CONTACT Chris Roberts christopher.roberts@sydney.edu.au Sydney Medical School, Faculty of Medicine and Health, University of Sydney, Sydney, Australia.

Introduction

The addition of a consensus statement on the assessment of professionalism (Hodges et al. 2018) in this edition of the journal complements the two statements published in an earlier issue: one on Assessment (Norcini et al. 2018) and the other on Selection and Recruitment (Patterson et al. 2018). The ways in which these three statements might intersect and bring new understandings about the relationships among professionalism, assessment, and selection were raised at the 2018 Ottawa-ICME Joint Conference on the Assessment of Competence in Medicine and the Healthcare Professions. In recognition of the lack of guiding literature, a compelling case emerged during the conference to further consider the ways in which the professionalism of health education students and graduates can be re-conceptualized within systems of selection and assessment. In this paper, we begin that conversation so that it might provide fruitful insights and stimulate future debate from a practice, theoretical, and research basis. Our guiding question was "What is the relationship between professionalism as a construct and systems of assessment and selection?" We begin by summarizing some of the key features recommended by the international working group on assessment in professionalism, and the ways in which their three levels of assessment resonate with the consensus statements in assessment and selection. We draw out what is common to all three statements, what is different, and the ways in which new understandings might lead to changes in selection and assessment systems, and further sharpen the ways in which the assessment of professionalism is considered.

The assessment of professionalism

Hodges et al. (2018) re-iterate three key discourses about professionalism that have implications for assessment at the level of the individual, the inter-personal, and the societal–institutional. This framing, developed after extensive consultation and informed by the extant literature, resonates with contemporary issues of selecting and assessing for professionalism. For example, the Physician Charter on Professionalism of the American Board of Internal Medicine (ABIM) Foundation and the European Federation of Internal Medicine, suggest physicians must uphold the best interests of patients whilst simultaneously assuring that health care resources are distributed justly (Weinberger 2011). This duality challenges educators to consider the professionalism of a healthcare professional or student at both the individual level and at a health systems level. In meeting this challenge, there has been much progress both in dispelling the ambiguity, confusion and controversy about what professionalism is, and in providing guiding principles as to what, how, and by whom professionalism might be assessed and critically at what level (Hodges et al. 2018).

Where professionalism is seen as an individual characteristic, then its assessment would focus on the measurement of certain traits, behaviors, or cognitive processes. Where professionalism is seen as a co-construction through inter-personal interaction, the focus is more on identifying unprofessional behavior that arises from particular kinds of social interactions. In both cases it cannot be assumed that judgements about these behaviors are generalizable across culture, context, and history. Importantly, behaving professionally in front of an assessor does not directly predict how individuals will be interacting with colleagues or when put into stressful healthcare contexts. Even if individuals do display behaviors that some might regard as unprofessional, we also cannot assume that these behaviors arise just from the individual. Assessors can, however, observe and provide judgement on individual student/trainee problem solving, and encourage reflection aimed at avoiding future lapses. However, both the assessor and the trainee/student must remain sensitive to the context in which judgements are made and narrative feedback should aim to articulate that.

When professionalism is viewed at the level of social structure, it can be understood in the context of the goals, aspirations and collective behaviors of healthcare and educational institutions and of the profession itself. Here, developing fit for purpose assessment requires extensive engagement through dialog and meaningful collective input from all stakeholders. Assessment would be based on the degree to which the profession (at the level of the class of students, a whole medical school, interdisciplinary faculty of health sciences, professional practice group, or even the profession as a whole) meets these expectations. This challenging goal resonates with the warning from Cronbach et al. (1997), over two decades ago, that "assessment programs intended to bring schools and students to high standards are adopting untraditional procedures and are proposing to support the kinds of decisions that have little precedent." Increasingly health care providers and patients expect their view to be included in setting those high standards for medical graduates' preparedness for practice at each transition in their training (Kostov et al. 2018).

Commonality and differences

The three consensus statements reflect the challenges for assessment, as professionalism has evolved from being viewed as a core competency at an individual level to a wider interpretation at the collective level. The assessment of individual professionalism (Hodges et al. 2018) resonates with the way professionalism is conceptualized in selection for personal characteristics (Patterson et al. 2018) and with programs of learning activities linked to professionalism (Norcini et al. 2018). In competency-based selection models, selection research has focused on individual competencies and traits (Patterson et al. 2018). Patterson et al. (2018), do not provide any particular definition of professionalism, rather professionalism is framed as a contextually specific construct, which is reflected in the differing empirical and theoretical literature underpinning the development of current selection methods. There is, however, an implicit assumption in the selection statement, that various personal characteristics and traits measured during selection and recruitment processes can collectively be understood to be representative of professionalism, an assumption which remains untested (Hodges et al. 2018). In selection, definitions of personal characteristics emphasize the importance of individual competencies in order to contrast the academic qualities of applicants for entry into health professional training programs, or clinical service roles. There is an assumption that assessments of personal characteristics would have a clear relationship with future outcomes. Patterson et al. (2018), report the weight of evidence for which assessments "add" predictive value for future professionalism. However, the science is unsettled as to which outcomes best capture the complex work of practicing health professionals, the so-called "criterion" problem.

In examining different selection methods, two assessment formats have the most compelling validity evidence for the purpose of determining individual characteristics: the Multiple Mini Interview (MMI) and the Situational Judgement Test (SJT) (Patterson et al. 2018). The MMI emerged from a distillation of faculty, community, and research informed notions of professionalism (Eva et al. 2004) within the format of

an Objective Structured Clinical Examination (OSCE) (Harden and Gleeson 1979). There is a body of evidence (Patterson et al. 2018) attesting to the MMI format retaining the robust measurement characteristics of the original OSCE format (Newble and Swanson 1988). By contrast, the SJT format has been used in organizational psychology for many years (Patterson et al. 2018). Their design is based on a stimulus format, which is contextualized to the selection. However, recent evidence suggests that personality traits are no longer regarded as stable deterministic predictors of behavior and are instead found to change across the life span and in response to environmental contexts (Patterson et al. 2018). There is limited evidence to robustly link the assessment of personality or "traits/characteristics" to subsequent professional behaviors (Hodges et al. 2018). We would argue that assessments of individual traits that are taken out of context are unlikely to add substantial predictive value either. MMIs and SJTs are best considered as parts of an assessment system where the design parameters for each method can vary significantly depending on the purpose of the recruiting institution (Patterson et al. 2018).

The concept of validity in assessment itself is evolving. The pursuit of universal evidence of construct validity (e.g. entry level professionalism) for large scale selection systems, have adhered to appropriate validity frameworks, for example Kane (Schuwirth and van der Vleuten 2012; Cook et al. 2015). Here, validity is a property of the proposed interpretations and uses of the assessment scores and not the tool itself. The Kane framework has been thought to be more appropriate for individual assessments than for assessment systems. More work in developing validity arguments in professionalism is required. In Kane's typology of inferences; scoring, generalizations, extrapolation, and decision rules, the types of inferences that can be drawn can be flexible. Were a case to be made, these could be inclusive of the catalytic effect (Norcini et al. 2018), political validity (Prideaux et al. 2011), cross cultural validity (Hodges et al. 2018), and validity as a social imperative (St-Onge et al. 2017). In reflecting on how judgements are being made about the professionalism of a student/trainee, what inferences are being made about the expression of particular indicators of professionalism in a particular context. What decisions are arising out of those judgements? How might patients view assessing of good or poor professional behavior versus the consequence of that behavior?

Given that all three papers (Hodges et al. 2018; Norcini et al. 2018; Patterson et al. 2018) have emphasized the importance of local context, there is an opportunity for all of the working groups to consider the theory and practice of longitudinal predictive validity studies which are often based on large datasets and typically operate at the individual applicant/student/trainee level. There is little current guidance available for designing assessments that might predict professional behaviors arising from particular kinds of social interactions. At the level of social structure, the selection working group (Patterson et al. 2018) have considered selection frameworks that can be understood in the context of the philosophy, goals, aspirations, and collective behaviors of healthcare and educational institutions, and of the professions themselves. Individual education and training institutions define their selection philosophy, and uniquely enact a selection policy (or policies) within the context of their national- or state-based regulations, their own history, mission and goals, and stakeholder organizations. They choose and implement a combination of specific methods from the range available to them. Making meaning where the purpose of selection has been revised to include the societal level, not only requires extensive engagement through dialog and meaningful input from all stakeholders, but also the development and enactment of policy at a tertiary educational, and health care systems level. We feel this is a task unlikely to be achieved by single-site research teams such that large scale inter-disciplinary research is required in future to address these issues more effectively

Systems of assessment

A significant update in the assessment statement is the development of a framework for systems of assessment (Norcini et al. 2018). A system of assessment is composed of multiple, coordinated individual

assessments and independent performances that are intended to be orderly, and aligned around the same purposes. This development is in response to the growing awareness of the importance of taking a systems approach to assessment. Such an approach incorporates broader notions of sampling, timing, and decision-making, the means of combining different kinds of information from different sources, and how progression decisions are made. Examples include licensure, programmatic assessment, and selection. Norcini et al. (2018), draw a careful distinction between the assessment *of* learning and the assessment *for* learning. It is noted that the weights given to the criteria, or elements of the framework, that apply to these two different purposes of assessment will also differ. For example, a criterion such as equivalence might be of greater importance for an assessment with a summative purpose rather than a formative purpose. The system of assessment for licensure ensures doctors have the knowledge and skills necessary to practice medicine safely and effectively. There is an emphasis at the system level on the accreditation of the medical school, and at the individual level through standardized tests of knowledge and skills, and supervised training and work-based assessment. Hodges et al. (2018), emphasize the importance of longitudinal assessment of professionalism and a greater focus on formative feedback, reinforcing the notion that professionalism is learnt not innate. This might be considered a feature of programmatic assessment, which Norcini et al. (2018) suggest serves both assessment and educational functions. Accordingly, the utility for each should be carefully considered in designing both assessments and their evaluation. Any measure of utility should be inclusive of the costs and the return on investment of the chosen methods of assessment. (Norcini et al. 2018)

A move towards a systems approach to selection (Patterson et al. 2018) has emerged recently, building on early work establishing the incremental validity of combinations of selection assessments over single formats. Researchers have called for developing standards for the evaluation of systems of selection in order to provide an evidence base to support further practice and policy development. There are significant differences in the ways data from individual assessments and independent performances are combined and co-ordinated, reflecting the flavor and philosophy of particular systems of assessment. For example, the selection consensus group draw the distinction between competency-based selection approaches (where assessments are designed and blue-printed using job analysis studies), and the North American approach of combining prior assessments that add predictive value. Yet, little research has addressed issues surrounding how best to weight and sequence the variety of assessment data collected at the point of selection to achieve the desired policy intentions. A system approach also raises the issue of whether research within advanced healthcare systems is generalizable to other international contexts, where fundamental research problems may be strikingly different. Some issues e.g. medical migration from countries with emerging healthcare markets to the UK, US, and Europe is a considerable issue for local medical workforce e.g. in Africa, and there is a dearth of research in this area (Patterson et al. 2018).

Assessing professionalism

The intersection between professionalism and systems of assessment inclusive of selection, graduation and licensure is a complex one. In order to provide some useful insights, we reflect on a vignette to illustrate the ways in which systems of assessment impact systems-level responses to degrees of learner difficulty and failure (See Box 1)

In making judgments about professional behaviors of trainees/students, Ellaway et al. (2018), suggest that these may be made within one of several subsystems that operate within the medical education institution, the University and/or the healthcare system. These subsystems include; (1) Selection and recruitment, (2) Success and completion, (3) Remediation, (4) Failure and Exclusion, and (5) Fitness to practice. In our scenario, judgements can be made to each of these levels about the professionalism of the protagonists, based on inferences about their expression of particular aspects of professionalism. The

resident made a mistake and should be prepared for feedback. At the individual level of assessment, the resident was incompetent in making the original error and had little insight into the impact of that mistake on patients and those who care for them and should continue self-monitoring her behavior. Having received feedback that is not constructive, the resident remains responsible to give feedback to colleagues, who may be under equal amounts of pressure from the system. At an interpersonal level, there are lapses of professional behavior by both the resident and the nurse arising from particular kinds of social interactions. The patient was unaware of a potential error. At a systems level, the resident and the nurse are over-worked and under-supported, and there is an imbalance of power due to professional hierarchy, gender, and culture, that make it difficult for them both to recognize or raise professionalism issues (Martimianakis et al. 2009). Expectations of performance at an individual level are conflicted with those at an institutional/systems level. Each level of assessment requires a different solution, and systems of selection and assessment focused on the individual level, need to consider both the interpersonal and the societal levels in order to address the reality of modern healthcare environments.

Improving the self and others and the system

In addressing the needs of selection systems to be context dependent, useful research at the individual level might look at representation of protected groups, likelihood of primary care (or research career), identification of students likely to graduate and successfully handle the challenges of the medical course, and to behave as professionals, providing humane, compassionate care. There is good work showing that a number of measures (MCAT-like admissions tests, the MMI, etc.) (Patterson et al. 2018) have broad applicability (validity generalization), with factors related to the mission of a school and the characteristics of the applicant pool resulting in the need to contextualize the admissions system.

How might thinking about professionalism from the perspective of good assessment (Norcini et al. 2018) encourage new research that takes account of context, the interpersonal and systems factor? One component of professionalism is the engagement of the student/trainee with improving both the self (on going learning) and others (teaching and support), and the system they work in (quality improvement and systems design) (Wilkinson et al. 2009). How well might an individual recognize situations where their professional behaviors might be put under pressure, how well might an individual recognize when their colleagues need extra help in order to maintain their professional behaviors and how well might an individual recognize when a systems-level solution might be needed?

To what extent is good assessment at the individual level integrated with the need to assess at the collective level? Medical education has tended to emphasize individual competence, which is a necessary but insufficient condition to impact the quality health care experienced by patients. Rather, patients experience good quality care by teams delivering collective competence, which may include the contribution of individuals at the level of team or hospital or ambulatory clinic setting (Lingard 2016). For example, there are increasing calls to consider the assessment of team collaboration in patient safety or chronic disease management (Jorm et al. 2016). At the level of the health system or society, including regulators, individual characteristics may be less important than the characteristics of the class as a whole. Kreiter et al., have asked the question whether current selection methods are promoting a homogeneity of the student body which is constraining diversity (Kreiter 2002; Kreiter et al. 2003) and in opposition to societal imperatives such as the widening access agenda (Patterson et al. 2018; Crampton et al. 2018). We propose looking beyond selection methods to a systems level of analysis and considering whether selection systems should aim to produce more diverse cohorts of students that reflect the nature of the communities they serve. Rather than making selection decisions applicant by applicant, it might make more sense to decide on the desired characteristics of the entering cohort as a whole. Then offering a selection system that could produce collective characteristics in the cohort, for example, cultural mix, numbers of doctors going into primary

care, or rural areas that reflect the communities that the medical school serves (Kreiter 2002; Kreiter et al. 2003). There are challenges in relating cohort level data from selection processes to meaningful health outcomes data in order to evaluate the impact of change (Patterson et al. 2018). If changes were made to selection systems in an effort to improve outcomes, it would be difficult to differentiate causal relations from associations because of the time elapsed between selection decisions and the availability of health outcomes.

So far, there has been little more than exploratory work (i.e. group selection methods) in selection and the notion of the interpersonal view of the assessment of professionalism. That is a consideration of the personal characteristics and institutional conditions that might promote or prevent lapses of professionalism. Two significant gaps in the literature have been in relation to the interpersonal view of the assessment of applicants/students/trainees; the perspective of patients on what professionalism is and, resolving the tension between individual versus institutional definitions of professionalism (Hodges et al. 2018).

Another less well explored area, is the relationship between selection and the curriculum (Patterson et al. 2018) and by inference assessment both *of* learning and *for* learning. Current literature (Van Mook et al. 2009) provides in depth understandings of what learning and teaching principles might impact professional behavior in the curricula of medical schools and residency programs. One way to address the link between selection and curriculum might be to select out on those attributes that are important, but which can't be easily taught or learnt during the program/curriculum. Developing assessments for those attributes that cannot be gamed by applicants would be a daunting task, even if there were agreement on what the attributes were. It may be better to assume most things are "learnable" and invest curricular resources in promoting their attainment. That said, there is little or no research available to directly guide schools/residency programs in this respect and it is likely that there is a spectrum of attributes that are harder or easier to teach and learn.

Implications for future research

All three 2018 Ottawa Consensus statements are predicated on a rich understanding of the strengths and weakness, and possibilities of testing individuals using single formats. The call across all three papers for a widening of the framings currently used in assessment to take account of social/contextual perspective, may well be a paradigm shift for medical education generally.

The three statements take varied approaches to considering the philosophical underpinnings of their perspectives. Norcini et al. (2018), focuses on the pragmatic aspects of assessment, leaving the philosophical debate, around issues of the theory of knowledge (epistemology) and of reality (ontology) to local contexts for educators to provide a better fit of assessment to learning. Hodges et al. (2018), create an epistemological matrix that has begun to be more widely used in health professions education to map out epistemological assumptions where there is a diversity of approaches to assessment. In selection research, guidance is provided on differing approaches to evaluative frameworks that take account of different epistemological underpinnings, for example including assessment properties and social accountability (Patterson et al. 2018). Additionally, the selection statement (Patterson et al. 2018) proposes the lens of critical realism, which holds that the theory of knowledge, or epistemology, is different from a theory of being, or ontology. It seems reasonable to think that in systems of assessment where purposes of supporting and judging learning are being considered, the epistemic and ontological can be reconciled in both the design and the evaluation stages.

In both the selection and professionalism statements, there are calls for both theory development and innovation in assessment methodologies. Patterson et al. (2018), suggest that the development of interdisciplinary theoretical frameworks might facilitate the future development of policy and practice using appropriate methodologies which are sensitive to local contextual priorities. Others have gone further and

suggested that we are entering an era where human judgment of assessment should prevail over the mechanistic. Interdisciplinary conversations need to be continued to ensure rigorous systems of assessment that acknowledge competence is contextual, constructed, evolving and, at least in part, subjective and collective (Hodges 2013; Sebok-Syer et al. 2018).

There are measurement challenges to establish sufficient evidence of validity for individual judgements in complex settings, for example, where work-based assessments of professionalism are nested in raters, within disciplines, and within hospitals. Assessments of professionalism and their interpretation are similar to the issues in work-based assessments such as the Mini-Cex (Norcini and Burch 2007). Often the impact of the institution is left out of the generalizability estimates of the contribution of noise obscuring estimates of the competence of the student/trainee. Research that has traditionally been focused on individual characteristics might refocus on the impact of system-based variables such as hospital, or training program, country, or culture to draw out insights into the impact of the interpersonal or institutional levels (Hodges et al. 2018). More broadly, the use of mixed research methods that combine regression analyses, e.g. Generalizability or Rasch models with qualitative approaches, may serve to better understand the manner in which people represent and make determinations about others (Gingerich et al. 2011).

An imperative in reframing the relationship between selection, professionalism and assessment, from both a system and an individual perspective, is considering patient perspectives in the assessment of professionalism – that is health outcomes and what is important to patients. Next is the derivation of the attributes of the health system and from that emerges desired health practitioner attributes. Once these attributes are settled, it becomes easier to design curricula and training programs, but it also becomes easier to know which attributes of a practitioner cannot be easily delivered by a program/curriculum – and that therefore informs selection. Table 1 lists some recommendations to evolve the relationship between professionalism as a construct and systems of assessment and selection.

Table 1. Summary of recommendations to evolve the Ottawa Consensus statements on selection and good assessment of health professional students in the context of professionalism. (Table view)

1	The importance of systems-level thinking emerges from all three consensus statements providing a significant challenge for medical educators to address.
2	To evolve assessment systems within medical schools to assure the quality of evidence that graduates are able to adjust to complex situations, work in teams, and meet a wide range of societal needs.
3	The mapping out of the different philosophical (including epistemological and ontological) assumptions underlying assessment systems for licensure and selection to create a common language with researchers looking at the assessment of professionalism.
4	Critical work is needed, to theorize about and provide empirical data, on the linkage between health outcomes and personal and collective attributes in the context of healthcare systems.
5	Develop consensus in what constitutes validity evidence for assessments of professionalism where students/trainees are nested within raters, within disciplines, and within institutions.
6	To establish principles of good assessment which regardless of specific systems; licensure, or selection, favor assessment *for* learning at multiple time points on the continuing pathway to becoming a healthcare professional.
7	To consider the impact of measures of collective competence on the health care system and compare with the impact of measures of individual competence.
8	Selection and professionalism researchers should revisit how they design, implement and assess desirable knowledge, behaviors and attitudes that indicate success as a healthcare practitioner, and which are contextualized and relevant across and between individuals and at the health systems level.
9	To explore the opportunities and challenges for expanding the repertoire of psychometric practice when it comes to assessing and selecting for professionalism both in program applicants and student/trainees, including the use of mixed methods in evaluations.
10	To develop ways of assessing professionalism at the interpersonal level which explores the notion of professional lapses and reflection aimed at avoiding future lapses.

Conclusions

In synthesizing the recent three Ottawa Consensus statements, we have seen much that deeply resonates among them and have provided some key issues that could usefully guide future work. These include widening the lens through which we view assessment to include large societal considerations, expanding the collective concept of validity amongst medical educators, and addressing concern with the outcomes of medical schools. An interdisciplinary approach can build on disciplinary perspectives to offer creative solutions for those who grapple with both selecting and ensuring the competence and performance of medical students. It is important this includes the needs of the diverse societies, which medical schools serve. The importance of systems-level thinking emerging from all three papers provides a significant challenge for medical educators to address. We have noted critical differences where agreement is elusive and suggested these as a good basis for further debate and scholarly input.

Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article

Notes on contributors

Chris Roberts, MBChB, MMedSci, FRACGP, PhD, The University of Sydney, Faculty of Medicine and Health, Sydney, New South Wales, Australia.

Tim Wilkinson, MB, ChB, M Clin Ed, PhD, FRACP, Professor of Medical Education, Deputy Dean and Director of the MBBCh program, Otago University, NZ, and consultant physician in geriatric medicine.

John Norcini, PhD, FAIMER, Philadelphia, Pennsylvania, USA.

Fiona Patterson, BSc, MSc, PhD, CPsychol, FRCGP (Hon), Work Psychology Group, Derby, United Kingdom of Great Britain and Northern Ireland.

Brian D. Hodges, MD PhD FRCPC, Executive Vice President and Chief Medical Officer, University Health Network and University of Toronto, Canada.

ORCID

Chris Roberts https://orcid.org/0000-0001-8613-682X

Tim J. Wilkinson https://orcid.org/0000-0002-4080-4164

John Norcini https://orcid.org/0000-0002-8464-4115

Fiona Patterson https://orcid.org/0000-0002-1031-130X

Brian D. Hodges https://orcid.org/0000-0003-4168-2725

References

Cook DA, Brydges R, Ginsburg S, Hatala R. 2015. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Med Educ*. 49:560–575. Crossref. PubMed.

Crampton P, Weaver N, Howard A. 2018. Holding a mirror to society? Progression towards achieving better sociodemographic representation among the University of Otago's health professional students. *N Z Med J (Online)*. 131:59–69. PubMed.

Cronbach LJ, Linn RL, Brennan RL, Haertel EH. 1997. Generalizability analysis for performance assessments of student achievement or school effectiveness. *Educ Psychol Meas*. 57:373–399. Crossref.

Ellaway RH, Chou CL, Kalet AL. 2018. Situating remediation: accommodating success and failure in medical education systems. *Acad Med*. 93:391–398. Crossref. PubMed.

Eva KW, Rosenfeld J, Reiter HI, Norman GR. 2004. An admissions OSCE: the multiple mini-interview [Research Support, Non-U.S. Gov't]. *Med Educ*. 38:314–326. Crossref. PubMed.

Gingerich A, Regehr G, Eva KW. 2011. Rater-based assessments as social judgments: rethinking the etiology of rater errors. *Acad Med*. 86:S1–S7. Crossref. PubMed.

Harden RM, Gleeson F. 1979. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ*. 13:39–54. Crossref.

Hodges B. 2013. Assessment in the post-psychometric era: learning to love the subjective and collective. *Med Teach*. 35:564–568. Crossref. PubMed.

Hodges B, Paul R, Ginsberg S. 2018. Assessment of professionalism: from where have we come – to where are we going? *Med Teach*.

Jorm C, Nisbet G, Roberts C, Gordon C, Gentilcore S, Chen TF. 2016. Using complexity theory to develop a student-directed interprofessional learning activity for 1220 healthcare students. *BMC Med Educ*. 16:199. Crossref. PubMed.

Kostov CE, Rees CE, Gormley GJ, Monrouxe LV. 2018. 'I did try and point out about his dignity': a qualitative narrative study of patients and carers' experiences and expectations of junior doctors. *BMJ Open*. 8:e017738. Crossref. PubMed.

Kreiter CD. 2002. The use of constrained optimization to facilitate admission decisions. *Acad Med*. 77:148–151. Crossref. PubMed.

Kreiter CD, Stansfield B, James PA, Solow C. 2003. A model for diversity in admissions: a review of issues and methods and an experimental approach. *Teach Learn Med*. 15:116–122. Crossref. PubMed.

Lingard L. 2016. Paradoxical truths and persistent myths: reframing the team competence conversation. *J Contin Educ Health Prof*. 36:S19–S21. Crossref. PubMed.

Martimianakis MA, Maniate JM, Hodges BD. 2009. Sociological interpretations of professionalism. *Med Educ*. 43:829–837. Crossref. PubMed.

Newble DI, Swanson DB. 1988. Psychometric characteristics of the objective structured clinical examination. *Med Educ*. 22:325–334. Crossref. PubMed.

Norcini J, Anderson MB, Bollela V, Burch V, Costa MJ, Duvivier R, Hays R, Palacios Mackay MF, Roberts T, Swanson D. 2018. 2018 Consensus framework for good assessment. *Med Teach*. 1–8. DOI: 10.1080/0142159X.2018.1500016 PubMed.

Norcini J, Burch V. 2007. Workplace-based assessment as an educational tool: AMEE guide no. 31. *Med Teach*. 29:855–871. Crossref. PubMed.

Patterson F, Roberts C, Hanson M, Hampe W, Eva K, Ponnamperuma G, Magzoub M, Tekian A, Cleland J. 2018. 2018 Ottawa consensus statement: selection and recruitment to the healthcare professions. *Med Teach*. 1–11. DOI: 10.1080/0142159X.2018.1498589 PubMed.

Prideaux D, Roberts C, Eva K, Centeno A, Mccrorie P, Mcmanus C, Patterson F, Powis D, Tekian A, Wilkinson D. 2011. Assessment for selection for the health care professions and specialty training: Consensus statement and recommendations from the Ottawa 2010 Conference. *Med Teach*. 33:215–223. Crossref. PubMed.

Schuwirth LW, van der Vleuten CP. 2012. Programmatic assessment and Kane's validity perspective. *Med Educ*. 46:38–48. Crossref. PubMed.

Sebok-Syer SS, Chahine S, Watling CJ, Goldszmidt M, Cristancho S, Lingard L. 2018. Considering the interdependence of clinical performance: implications for assessment and entrustment. *Med Educ*. 52:970–980. Crossref.

St-Onge C, Young M, Eva KW, Hodges B. 2017. Validity: one word with a plurality of meanings [journal article]. *Adv in Health Sci Educ*. 22:853–867. Crossref. PubMed.

Van Mook WN, de Grave WS, van Luijk SJ, O'Sullivan H, Wass V, Schuwirth LW, van der Vleuten CP. 2009. Training and learning professionalism in the medical school curriculum: current considerations. *Eur J Intern Med*. 20:e96–e100. Crossref. PubMed.

Weinberger SE. 2011. Providing high-value, cost-conscious care: a critical seventh general competency for physicians. *Ann Intern Med*. 155:386–388. Crossref. PubMed.

Wilkinson TJ, Wade WB, Knock LD. 2009. A blueprint to assess professionalism: results of a systematic review. *Acad Med*. 84:551–558. Crossref. PubMed.