

reasoning. *N Engl J Med* 2006;**355**:2251–2.

27 Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance

in medicine and related domains. *Acad Med* 2004;**79** (Suppl):70–81.

28 Hatala RM, Brooks LR, Norman GR. Practice makes perfect: the

critical role of mixed practice in the acquisition of ECG interpretation skills. *Adv Health Sci Educ Theory Pract* 2003;**8**:17–26.

Assessment: do we need to broaden our methodological horizons?

Ayelet Kuper, Scott Reeves, Mathieu Albert & Brian David Hodges

Although medical education is a broad field of research and practice, it has come to be dominated by issues of assessment. Reasons for this emphasis range from the focus on accountability for educational outcomes¹ to the established relationship between assessment and student motivation.² Researchers in the domain, especially in North America, have largely focused on methodologies taken from psychometrics and have overlooked the broader social sciences literature devoted to the analysis of social behaviour and social interaction. In this commentary we provide a critique of the ubiquitous use of psychometric methodologies and perspectives and argue that the social sciences offer other rich methodological resources for the study of assessment.

Within medical education research, evaluation is almost always carried out using a set of appraisal tools that are collectively known as psychometrics. We talk about whether a test is valid (whether it measures the thing we

want to measure) and whether it is reliable (whether it measures it in a reproducible fashion). Psychometrics has been very successful in evaluating the assessment of many aspects of medical training. It has, for example, allowed the medical education community to systematically evaluate different measures of medical content knowledge,³ as well as to show that technical skills can be assessed in a reproducible, valid manner.⁴

The social sciences offer other rich methodological resources for the study of assessment

What is rarely made explicit, however, is what the use of psychometric analysis implies about that which is being assessed. More sophisticated psychometricians do stipulate that the latent traits they measure do not really 'exist in any physical or physiological sense'⁵ – that they're 'statistical constructs'.⁵ However, there is a longstanding implicit reification in the literature of the existence of these underlying internal traits that can be measured over time.^{6,7} In either case, it is clear that psychometric tools were initially developed by cognitive psychologists to be valid and useable for phenomena that could at least be conceptualised as stable traits within a single individ-

ual. They were designed for the assessment of personality traits such as intelligence, honesty and diligence. Despite issues of test–retest reliability and other methodological hurdles related to positive and negative changes in knowledge over time, they have since been extended for use in the assessment of knowledge and performance.

There is growing understanding that some aspects of medical education are better thought of as social constructs: instead of being considered as expressions of a single individual's abilities, they are conceived of as the products of interactions between two or more individuals or groups

With this caveat, psychometric tools have proven themselves to be very useful for assessing the aspects of medical training, such as content knowledge, that are more easily conceptualised as psychological constructs, as existing individually within each trainee. There is, however, a growing understanding that some aspects of medical education are better thought of as social constructs. That is, instead of being considered as expressions of a single individual's abilities, they are conceived of as the products of interactions between 2 or more individuals or groups.

The Wilson Centre for Research in Education, University of Toronto, Toronto, Canada

Correspondence: Ayelet Kuper, Sunnybrook Health Sciences Centre, 2075 Bayview Avenue, Room HG 08, Toronto, Ontario M4N 3M5, Canada. Tel: 00 1 416 480 4994; Fax: 00 1 416 480 4563; E-mail: ayelet94@post.harvard.edu

doi: 10.1111/j.1365-2923.2007.02945.x

From a social science (e.g. sociology, anthropology, social psychology) perspective, our ideas about how people should act in different situations are context-specific and culture-bound. They are constructed by the cultures and structures of the societies in which we live, as well as by the social groups to which we belong in those societies and by our ongoing interactions with others. As each person will have had a different set of experiences and positions within one or more societies, he or she will have a slightly or radically different set of perceptions and interpretations of appropriate behaviour in various contexts. Translated into medical settings, this means that our descriptions of competence in certain areas are socially constructed and may differ from those of our fellow doctors, our non-doctor colleagues, and our patients.

Our descriptions of competence in certain areas are socially constructed and may differ from those of our fellow doctors, our non-doctor colleagues, and our patients

This point of view changes the definitions of many abilities expected of trainees from being stable and internal to being socially constructed and historically transient and, as such, situational and interpersonal. Empathy, for example, can be described as the behaviour that causes another person, such as a patient, to perceive someone, such as their trainee doctor, as being empathic; that trainee doctor's empathy (or lack of) is constructed in the encounter rather than being an innate quality of that doctor as an individual. This construction comes from each person's perception of the other and of the situation, a

perception that, for its part, is grounded in each of their culture(s) and personal histories. These perceptions may differ radically not only between the trainee and the patient, but between either or both of them and an examiner who is observing the encounter.

This leads to an intriguing problem. In a domain where the touchstones have been inter-rater reliability (and a numerical understanding of validity that depends on such reliability), how do we account for the shifting, context-laden, socially constructed nature of trainee competencies, such as empathy and professionalism? Rather than trying to pin down the definitions of these abilities to a single artificial norm, how can we begin to capture their inherent variability and analyse it in a systematic, meaningful way?

There is tantalising scope for whole programmes of research in this area

There is a large body of literature in the social sciences devoted to the analysis of social behaviour and social interaction (such as the work of Pierre Bourdieu, Irving Goffman, Anselm Strauss, Howard Becker and Kurt Lewin, among others). Ethnography provides such a methodological approach.⁸ Ethnography has as its central focus the understanding of social processes, behaviours and perceptions that exist within groups and organisations. Participant observations and key informant interviews are commonly employed to explore and illuminate the social actions and interactions that occur within a specific context. Discourse analysis provides another possible methodological route. By focusing in depth on the verbal interactions that occur between individuals,

using techniques such as interviews, explanations of how individuals construct versions of the social world can be generated. These methodologies, among others, allow the generation of rich qualitative datasets, which can be used to create qualitative assessments.

For example, an analysis of a combination of semi-structured key informant interviews and focus groups, based in the ethnographic tradition, conducted on a medical ward with patients, nurses, students and other trainees, as well as with attending doctors, could generate rich, meaningful trainee assessments for certain aspects of clinical rotations. Such assessments would be particularly useful for non-medical expert doctor competencies, such as those of collaborator, communicator and professional.⁹ Unlike standard 360-degree assessments, the emphasis would be on capturing the range and perceptions of interpersonal behaviour taken in context. Although the results of such an assessment would not be psychometrically reproducible, the robustness of such a process could be evaluated in a rigorous way. Driesen *et al.* have previously described an evaluation method using strategies derived from qualitative research to show the credibility and dependability of a portfolio assessment process.¹⁰ Although their work was not grounded in a particular methodology, it should be possible to adapt the different criteria for judging the rigour of various methodological approaches in order to evaluate assessments carried out in those traditions. In this case, for example, one could look to the criteria used for evaluating ethnographic research in the social sciences. There is, indeed, tantalising scope for whole programmes of research in this area.

We need to think about the nature of the constructs and competencies we are trying to assess and choose our evaluation tools accordingly

Although we are not advocating an end to the use of psychometrics in medical education, we would like to propose that we go back to basics. We need to think about the nature of the constructs and competencies we are trying to assess and choose our evaluation tools accordingly. Rather than being tied to any single methodology, we should continue to focus our concerns on our ability, as a medical community, to know what it is that our trainees know and do, and to assess and evaluate it appropriately. That's all that matters – the rest is commentary.

REFERENCES

- 1 Chen FM, Bauchner H, Burstin H. A call for outcomes research in medical education. *Acad Med* 2004;**79** (10):955–60.
- 2 Swanson DB, Case SM. Assessment in basic science instruction: directions for practice and research. *Adv Health Sci Educ Theory Pract* 1997;**2**:71–84.
- 3 Wass V, van der Vleuten CPM, Shatzer J, Jones R. Assessment of clinical competence. *Lancet* 2001;**357** (9260):945–9.
- 4 Martin JA, Regehr G, Reznick R, Murnaghan J, Hutchison C, Brown M. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 1997;**84** (2):273–8.
- 5 Anastasi A, Urbina S. *Psychological Testing*, 7th edn. Upper Saddle River, NJ: Prentice-Hall 1997;73.
- 6 Brown FG. *Principles of Educational and Psychological Testing*, 3rd edn. New York: Holt, Rinehart & Winston 1983;3–16.
- 7 Murphy KR, Davidshofer CO. *Psychological Testing: Principles and Applications*, 5th edn. Upper Saddle River, NJ: Prentice-Hall 2001;39–46.
- 8 Hammersley M, Atkinson P. *Ethnography: Principles and Practice*, 2nd edn. London: Routledge 1995.
- 9 Frank JR, ed. *The CanMEDS 2005 Physician Competency Framework. Better Standards. Better Physicians. Better Care*. Ottawa: Royal College of Physicians and Surgeons of Canada 2005;15–28.
- 10 Driessen E, van der Vleuten C, Schuwirth L, van Tartwijk J, Vermunt J. The use of qualitative research criteria for portfolio assessment as an alternative to reliability evaluation: a case study. *Med Educ* 2005;**39** (2):214–20.