Evaluating workplace-based assessment of interns in a Queensland hospital: does the current instrument fit the purpose?

TO THE EDITOR: An evaluation of 3390 prevocational progress assessment forms in New South Wales suggested that the assessment instrument appears unable to detect underperforming doctors, and may not aid their professional development.¹ Similarly, Queensland interns undertake five terms per year and, in each of these, they are assessed against 11 items in three domains: clinical competence, communication, and personal and professional skills. The purpose of the

assessment is to identify underperformers and subsequently to help improve their performance. However, there is limited literature on the psychometric properties of the instruments used. Our study aimed to evaluate the construct validity and reliability of the instrument as a measure of interns' performance.

We performed principal component analysis (PCA) of the assessment items from all 72 interns who completed the five assessment forms during their intern year of the University of Queensland (2005 to 2007), working in one teaching hospital. A reliability test was performed using generalisability theory, which measured the variation in scores due to differences in performance between interns. It is generally expressed as a G coefficient ranging from 0 to 1 (no reliability to perfect reliability).

Supervisors assess interns using the rating scale: "requires substantial assistance"; "requires further development"; "consistent with level of appointment"; "performance better than expected"; "performance exceptional"; or "not applicable or not observed". Supervisors' ratings were coded from 1 to 5 for the data analysis.

Our examination of the assessment forms found that the mean of the skill

“
the assessment
instrument
appears unable
to detect
underperforming
doctors
”

Zhang et al

items ranged from 3.69 to 4.16 (SD, 0.31–0.41). PCA identified only one factor from the 11 assessed items, and it explained 71% of variance in scores. The loadings of the coefficients ranged from 0.80 to 0.93 (Box) and the G coefficient was 0.16.

These data show that the instrument in this setting only measures a single global work performance factor, has low reliability, does not discriminate well between interns, and does not differentiate well between the different domains and assessment items, raising the question as to whether it is fit for purpose. The assessment program should be able to assess all these required attributes of junior doctors as outlined in the Australian Curriculum Framework for Junior Doctors in order to work safely in Australian hospitals and other health care settings.^{2,3} Other assessment methods involving direct observation, such as the mini clinical evaluation exercise⁴ and direct observation of procedural skills,⁵ could be integrated into the assessment to enhance its validity, but would require trained assessors and substantial time commitment for development and delivery.

Jianzhen Jenny Zhang Research Fellow

David Wilkinson Dean of Medicine and Head

Malcolm H Parker Associate Professor of Medical Ethics

Andrew Leggett Senior Lecturer

Jill Thistlethwaite Professor of Medical Education

School of Medicine, University of Queensland, Brisbane, QLD.
jenny.zhang@uq.edu.au

Acknowledgements: We thank staff from the Medical Education Unit at Princess Alexandra Hospital for their support and advice.

Competing interests: No relevant disclosures.

doi: 10.5694/mja11.11369

1 Bingham CM, Crampton R. A review of prevocational medical trainee assessment in New South Wales. *Med J Aust* 2011; 195: 410–412.

2 Confederation of Postgraduate Medical Education Councils. Australian curriculum framework for junior doctors. <http://curriculum.cpmec.org.au/> (accessed Aug 2010).

3 Hays R. Assessment in medical education: roles for clinical teachers. *Clin Teach* 2008; 5: 23–27.

4 Sidhu RS, Hatala R, Barron S, et al. Reliability and acceptance of the mini-clinical evaluation exercise as a performance assessment of practicing physicians. *Acad Med* 2009 (10 Suppl); 84: S113–S115.

5 Barnforth M, Agarwal M. Direct observation of procedural skills (DOPS). In: Bhugra D, Malik A, Brown N, editors. *Workplace-based assessments in psychiatry*. London: Royal College of Psychiatrists, 2007: 56.

Domains, assessed items, descriptive statistics and factor loading of principal component analysis of the assessment instrument used on 72 graduates

Domain	Skill items	Mean (range)	SD	Factor loading
Clinical skills	Knowledge base: demonstrates adequate knowledge of basic and clinical science	3.72 (3.00–4.40)	0.33	0.84
	Clinical skills: elicits and records accurate, complete history and clinical examination	3.87 (2.80–4.60)	0.37	0.90
	Clinical judgement/decision-making skills: applies knowledge base and clinical findings; organising, synthesising and acting on information	3.88 (2.80–4.80)	0.39	0.90
	Emergency skills: acts effectively and, when appropriate, acknowledges own limitations and seeks help	3.79 (3.15–4.40)	0.34	0.76
	Procedural skills: performs procedures competently	3.70 (3.20–4.40)	0.38	0.72
Communication skills	Patient and family: interacts effectively and sensitively with patients and families/caregivers	4.07 (3.40–4.80)	0.33	0.79
	Medical records/clinical documentation: provides clear, comprehensive and accurate records	3.92 (2.80–4.60)	0.41	0.88
Personal and professional skills	Professional responsibility: demonstrates appropriate attitudes and behaviour, including punctuality, reliability, honesty and self-care	4.15 (3.20–5.00)	0.34	0.85
	Teaching: demonstrates commitment to learning, reflective thinking, and teaching others	3.69 (3.20–4.40)	0.34	0.80
	Time management skills: organises and prioritises tasks to be undertaken	3.97 (3.20–4.60)	0.36	0.89
	Teamwork and colleagues: works and communicates effectively within a team	4.16 (3.20–4.80)	0.31	0.93