

Workplace-based assessment and the art of performance

Einstein had it half right. He could have said:

'Not everything that counts can be measured. Not everything that can be measured counts ... but keep working at it!'

The challenge we have with assessing generalism, the most complex of medical specialties, is that although we recognise what a good GP is (it's someone we would trust to look after our family, isn't it?) we can't easily define what that means. However, if we want our assessments to be valid, we shouldn't be deterred by what Einstein said but should learn to count more appropriately. Or rather, we should learn to assess more appropriately because with complex abilities that are perceptible but not readily quantifiable, counting is not enough. Nowhere is this more evident than in workplace-based assessment (WPBA) in which qualitative and quantitative techniques coexist, construing the development and assessment of performance as both an art and an evolving science.

In this article, we will consider the place of WPBA in the nMRCGP (the GP licensing examination), the culture change in assessment and training that WPBA represents,¹ and the threats to WPBA's potential benefits.

WPBA may be new to GP assessment, but it represents an evolution of the structured trainers report previously used in GP training. WPBA is well tried and tested in non-medical vocational qualifications and is used by other medical Royal Colleges in their licensing examinations. The traditional examination elements of the nMRCGP are the machine-marked applied knowledge test (AKT) and the OSCE-style clinical skills assessment (CSA), which respectively test what a doctor thinks and how they deal with patient problems under simulated conditions. From these, we

might draw inferences about how a doctor could perform in real life. WPBA dispenses with the need to infer, by assessing how the doctor actually performs in the workplace, responding to a consensus of opinion in the literature that greater emphasis should be placed on performance-based assessment.^{2,3} Because WPBA tests what GPs are required to do in their working lives, it can potentially test the whole GP curriculum, unlike CSA and AKT, which are restricted by their examination formats.⁴ This does not mean that it must do so and time will tell which parts of the curriculum WPBA is best placed to assess.

So, if WPBA is so authentic and broad-ranging a test, why not dispense with CSA and AKT altogether? The reason is that WPBA does not yet have the track record to allow us to jettison traditional examinations.^{5,6} Additionally, there are arguments for high-stake assessments like the nMRCGP to include all levels of Miller's pyramid,⁷ which describes the various types of ability that could be assessed. These range from testing what doctors know, through what they say they would do in particular situations, to what they actually do in practice. The three components of the nMRCGP allow all levels of the pyramid of assessment to be represented.

The nMRCGP components therefore complement each other by illuminating the doctor's performance from different perspectives.⁸ WPBA is singular in testing real performance, but also brings other benefits to the nMRCGP mix. Let us consider three examples.

Firstly, WPBA greatly increases the range of clinical contexts in which doctors are tested.⁹ This is important because although many abilities in medicine are generic, many are context-specific (a sound knowledge-base in cardiology does not predict the same in gynaecology). Generalism is so broad that the range of the curriculum could not be

adequately tested by the other two components alone, although they serve an important function in standardising the exposure of candidates to important contexts. Secondly, WPBA allows patients and colleagues, who are the most valid raters of certain GP behaviours within, for example, communication, teamworking and developing relationships, to contribute to the assessment of performance. Lastly, there are certain behaviours that need to be witnessed over a period of time and therefore cannot be tested by 'biopsy' in AKT or CSA. The examples of such behaviours are not trivial but lie at the heart of our abilities as GPs and comprise clinical and non-clinical competencies. Clinical competencies include the ability to use time as a diagnostic tool, vary management options, and minimise risk as the patient's condition evolves. Non-clinical competencies include the ability to provide continuity of care, work cooperatively with the other members of the team, and preserve a sustainable work-life balance.

The opportunities that WPBA affords are only meaningful if we can trust the assessment process, and this brings us to a major challenge to WPBA. Can it be depended upon? Let us address some specific concerns.

IF WE ARE TO HAVE FAITH IN WPBA SURELY IT MUST HAVE A SIGNIFICANT FAILURE RATE?

To answer this, we need to understand the purpose of WPBA, which unlike CSA and AKT is not to provide an end of training snapshot of performance, but a continual assessment throughout the specialist training period. We therefore look for WPBA, particularly the staging reviews, to indicate whether the trainee is on course for becoming competent by the end of training or whether a problem exists that needs targeted assessment and possibly further training. In traditional

examinations, we would be aghast if examiners expected everyone to pass. Indeed, some would see a significant failure rate as being a mark of a satisfactory exam. However, if WPBA is working as intended, there should be no surprises at the end of training. Trainees who are not competent should be picked up much sooner for remedial help and those who are allowed to progress to the end of training should be more likely to be capable of passing, because they are building on a solid foundation. We can therefore see why it does not matter that the mandatory assessments in WPBA, such as case-based discussion are not used as discreet summative tests. They are chosen because they contribute powerfully to the emerging picture of performance that comes from the use of validated tools and, to a degree that is not always appreciated, from evidence that arises naturally in the course of working.

HOW CAN WE RELY UPON WPBA JUDGEMENTS IF THEY ARE SUBJECTIVE RATHER THAN OBJECTIVE?

Subjectivity and objectivity are part of a continuum, but let us consider the polarities. There is a commonly-held belief that objectivity is synonymous with reliability, but this is false. In fact, if sufficient subjective judgements are combined, the collated judgement about performance can be reliable.¹⁰ This has important implications. If we felt that objectivity was paramount, we might be tempted to do what Einstein warned about and produce lengthy checklists that run the risk of trivialising the assessment by not testing what is really important.¹¹

Many essential professional competencies, such as 'good teamworking' are multifactorial and inter-relational. They cannot be adequately assessed by checklists,¹¹ but require assessors to take a range of data into account sometimes over lengthy periods, ascribe appropriate weighting to the elements of the evidence, and then make global judgements. This complex process often appears subjective when it does not use an external rubric but relies upon internal deliberations. However, 'subjective' does not mean idiosyncratic.

For the judgements that arise from these deliberations to be acceptable they must be capable of being explained and of being shown to be rational (evidence-based) and in line with the gradings that might be given by competent adjudicators.

Each of these requirements must be met and each presents significant challenges. Trainers already have their internal frameworks and language of competence, but to explain, rationalise, and account for judgement to an external community, a common framework and language is needed. This is the purpose of the competency framework that underpins and informs WPBA. In this, a series of word pictures describe different stages in the development of competence within 12 domains that are rooted in the curriculum.¹² An example from the 'communication and consulting skills' domain is 'Works in partnership with the patient, negotiating a mutually acceptable plan that respects the patient's agenda and preference for involvement'.¹²

Explaining and justifying complex judgements will be easier initially in some areas of the competency framework than in others and this relates to whether a rubric and consensus of performance currently exists or not. For example, consulting skills are relatively widely understood through consultation models and performance criteria and present less of a challenge to justification than explaining and justifying a judgement made on the standard of holistic care. The latter is not impossible, but its rubric is at a much earlier stage of development and will need support.

WON'T WPBA DRIVE LEARNING, TO THE DETRIMENT OF TRAINING?

It has been said that 'Students don't do what you expect, they do what you inspect'. In WPBA, the 'inspection' is curriculum-wide, using the word pictures of the competency framework, a multitude of methods and a range of raters.¹³ It is therefore highly unlikely that candidates could prepare strategically for WPBA by becoming adept at any particular exam technique. Hitherto, learning and assessment in medicine

have been separate activities, often promoting superficial learning at the expense of deeper learning.¹⁴ In WPBA, the evidence that arises from assessment is used to provide formative feedback as well as to make summative judgements¹⁵ with the responsibility for the former lying with trainers and the latter with separate panels. Assessments have always had the potential to do both, but in WPBA the outcomes of assessment are purposefully harnessed to provide feedback throughout training in the specific and timely manner needed to appropriately tailor the trainee's educational journey. In effect, this is assessment for, as well as assessment of, learning.

This educational journey is a quest for meaning and trainees can develop their understanding at the following levels of the competency framework. Firstly, for each element of behaviour described in the framework, word pictures illustrate three levels of performance along a competency progression. The central level represents the licensing standard of 'competent' and the other two are below and above this level, thus allowing trainees to gauge where they stand in the continuum and encouraging developmental progression even for those deemed to be competent. Next, each competency progression is part of a set that between them describe that domain of competence, such as 'making diagnoses/making decisions'. Thirdly, each of the 12 domains of competence do not sit in isolation but interconnect with other domains, which between them describe the important abilities of the modern GP.

As described, WPBA clearly has great potential to reconnect education and assessment.¹⁶ However, there are threats to the achievement of this potential that we should be aware of. The competency framework is the key to the process and needs to be understood by trainees and educators.

WPBA is learner-led, with trainees self-assessing and presenting evidence that they believe will confirm competence. Trainees therefore need to understand the meaning of the word pictures within the various clinical contexts of the curriculum. They must also have sufficient powers of

reflection and insight for their self-assessments to be dependable. This is not a given and because of their importance, the ability to reflect on performance and to develop insight¹⁷ are vitally important for trainers and trainees to develop and monitor within WPBA.

The judgement of trainers is historically not in doubt. For years before WPBA, they have demonstrated their ability to recognise competence and, importantly, identify poorly performing registrars. The difference now is that trainers need to guide registrars with their evidence collection and justify their assessments of registrar performance. To do the latter, educators will need to become familiar with the language of the competency framework, so that it is clear how the evidence and the judgements based on this, relate to the word pictures. They will also need to develop insight into their assessment behaviour, in particular to recognise their personal biases, prejudices and over- or under- reactions to aspects of the trainee's persona and behaviour. This will need considerable 'training the trainer' investment, but by so doing, trainers are more likely to give appropriate weighting to what they perceive and thereby, increase the reliability of the assessment process.

Not all the competency framework is applicable to secondary care. The content and rating scales may therefore need to be refined, so that hospital colleagues can comment upon those behaviours that are valid and feasible to rate in a secondary care setting. Increasingly, the assessors' judgements and justifications will need to be documented, the implication being that simply signing off the tools will not provide the sufficiency of justification that is required. In addition, educators will need to routinely share their judgements with their peers so that a communal understanding of what the standard of 'competent' means in contemporary practice can be created and maintained. The ability to use the assessment tools and the framework should not be assumed, but will need to be assured through adequate training of those in primary and secondary care who have a substantive role. Educators will rightly continue to approach training in a variety

of creative ways, using the competency framework to reference rather than prescribe their own particular road-map for guiding trainees through both curriculum and consultation (T Norfolk, personal communication, 18 May 2008). However, assessment practices will need to be standardised as part of the quality management of WPBA and this represents a major but necessary task.

The evidence-set upon which the judgements rest needs to be satisfactory in its scope and quality.¹⁸ There is a commonly held belief that the evidence forthcoming from the minimum mandatory application of the validated tools is in itself sufficient to demonstrate competence. This cannot be the case as the minimum application simply guarantees a degree of comparability and consistency between portfolios, not sufficiency of evidence. Additionally, evidence for many important facets of performance, such as the demonstration of reflection and continuing development, cannot be adequately gathered through the tools, but arises through the richness of naturally occurring evidence. The evidence-set appropriate for the demonstration of competence therefore needs to be broader and qualitatively much more developed than might be supposed. There is a danger that if we reduce the portfolio of evidence to the minimum requirements, the scope and quality of information required for making dependable judgements and adequately informing trainees' learning plans, will be impaired.

WPBA has been described here as being in the vanguard of a new approach to educational assessment.¹⁹ It will evolve to meet the needs of educational programmes and face challenges in gaining the respectability currently accorded to traditional examinations. Nevertheless, if those involved in the process understand the issues and continue to professionalise their approach, the art and science of performance assessment could become a fitting commentator on the art and science of generalism.

Amar Rughani

REFERENCES

1. Swanwick T, Chana N. Workplace assessment for licensing in general practice. *Br J Gen Pract* 2005; **55**: 461–467.
2. Wragg A, Wade W, Fuller G, *et al*. Assessing the performance of specialist registrars. *Clin Med* 2003; **3**: 131–134.
3. Feletti G (ed.). In-training assessment. In: Newble D, Jolly B, Wakeford R (eds). *The certification and re-certification of doctors. Issues in the assessment of clinical competence*. Cambridge: Cambridge University Press, 1994.
4. Schuwirth L. The use of clinical simulations in assessment. *Med Educ* 2003; **37** (suppl.1): 65–71.
5. Norcini JJ. Work-based assessment. *BMJ* 2003; **326**: 753–755.
6. Roberts C, Newble D, O'Rourke A. Portfolio-based assessments in medical education: are they valid and reliable for summative purposes? *Med Educ* 2002; **36**: 899–900.
7. Miller G. The assessment of clinical skills/competence/performance. *Acad Med* 1990; **65**(Suppl.): S63–67.
8. Southgate L, Grant J and working group. *Principles and standards for an assessment system for postgraduate medical training*. London: Postgraduate Medical Education and Training Board, 2004.
9. Rethans JJ, Norcini JJ, Baron-Maldonado M, *et al*. The relationship between competence and performance: implications for assessing practice performance. *Med Educ* 2002; **36**(10): 901–909.
10. Schuwirth LWT, Southgate L, Page GG, *et al*. When enough is enough: a conceptual basis for fair and defensible practice performance assessment. *Med Educ* 2002; **36**(10): 925–930.
11. Regher G, MacRae H, Reznick R, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Acad Med* 1998; **73** (9): 993–997.
12. Royal College of General Practitioners. nMRCGP WPBA competence framework 2007. Competency areas in detail. http://www.rcgp.org.uk/docs/nMRCGP_12%20Competency%20Areas%20in%20detail.doc (accessed 7 Jul 2008).
13. Jolly B, Grant J. *The good assessment guide: a practical guide to assessment*. Milton Keynes: Open University Centre for Education in Medicine, 1997.
14. Broadfoot, P. Assessment for lifelong learning: challenges and choices [editorial]. *Assessment in Education: Principles, Policy & Practice*; **8**(3): 5–7.
15. Wiliam D. An overview of the relationship between assessment and the curriculum. In: Scott D (ed.) *Curriculum and assessment*. Westport, CT: Ablex Publishing, 2001: 165–181.
16. Assessment Reform Group. *Assessment for learning: beyond the black box*. Cambridge: Cambridge University Press, 1999.
17. Hays RB, Jolly BC, Caldon LJM, *et al*. Is insight important? Measuring capacity to change performance. *Med Educ* 2002; **36**: 965.
18. Eraut, M. *Developing professional knowledge and competence*. London: Falmer Press, 1994.
19. Schuwirth LW, van der Vleuten CP. A plea for new psychometric models in educational assessment. *Med Educ* 2006; **40** (4): 296–300.

DOI: 10.3399/bjgp08X319783