

Portfolios – attempting to measure the unmeasurable?

Portfolios have had a mixed press in medical education, with some finding them useful for training and education^{1,2} and others criticizing them as fads getting in the way of learning about doctoring by doctoring.³ Until now they have primarily been used for formative assessments, but they do hold attraction for summative assessment as they are based on work relevant experiences and track the learner over a protracted period of time.

This issue contains two contributions on portfolio learning in General Practice in the UK. Mathers *et al.*⁴ conclude that portfolios are an effective and efficient way of doctors pursuing continuing education. Completed portfolios were assessed for award of post-graduate education allowance (PGEA) by external assessment by a tutor. The external tutors judged the portfolios on whether there was enough evidence appropriate to the award of PGEA. It is always difficult to know what 'enough' means in this type of context, but as the portfolios were essentially being used as formative assessment tools, with the award of PGEA being an incentive to engage in learning rather than a summative judgement, it could be argued that assessment of content did not need to be precise.

Pitts *et al.*⁵ have taken the assessment of portfolios a step further by examining the inter-rater reliability of experienced general practice trainers in assessing portfolios of prospective trainers in a situation where summative judgements were being made on competency to become trainers. They used standardized assessment criteria and concluded that the degree of agreement between assessors did not reach a level where safe summative judgements could be made.

Both of these studies use fairly small sample sizes but raise the fundamental question of whether we should be attempting to summatively assess learning based on self-constructed material, and if so what are the best methods.

If we believe current thinking on the development of professionals,⁶ then portfolios are one way of encouraging reflection on clinical or educational practice as a stepping stone to professional mastery. They also hold promise in assessing medical teachers, particularly for

career development and progression, an area hitherto neglected in general. If, however, portfolios are going to be used widely as tools for such activities, can we assess their content by agreeing that participation and completion of a set of personal objectives is enough, or do we have to formally assess them in a standardized way? If portfolios are going to be used in final examinations, for accrediting trainers and other medical teachers or perhaps, as has been suggested,⁷ as one route to re-certification of doctors, then we have to devise robust methods of assessment. But do we have to seek assessments that meet present ideas of validity and reliability?

Current assessment methods are based on the views put forward by authors such as Cronbach⁸ and Cole & Moss.⁹ They argue that validity and reliability are dependent on measurements made in standardized situations, and that in such situations steps must be taken to reduce bias in these measurements. These thoughts have led to the growth of standardized measures, such as multiple choice examinations, which are designed to test specific parameters and rank doctors' performances with each other. But, assessment based on comparing doctors with each other and with issuing grades or marks does not fit with portfolios, which are non-standardized. Herein lies one of our biggest problems – how do we effectively assess material that is non-standardized? Is it sufficient to credit engagement in a process or must we strive to measure outcomes?

Pitts and his colleagues have illustrated just how difficult it is to assess portfolios if we rely on traditional methods of assessment. One of the reasons for this difficulty is that portfolios contain personalized material with few points of objectivity that allow comparisons to be made between doctors. Assessment is also labour intensive, requires careful reading and response to a learner's objectives and evidence of whether they have been met.

There is little research evidence available in the educational and medical educational literature on how to carry out large-scale portfolio assessments, and what has been done is largely within the American school system. One such example¹⁰ assessed students creative writing by developing a rubric on a scale of 1–6 encompassing the spectrum of 'exceptional achievement' to 'minimal evidence of achievement' with descriptors for each under the headings of: Focus/organization;

Correspondence: David Snadden, Tayside Centre for General Practice, University of Dundee, Kirsty Semple Way, Dundee DD2 4AD, UK

Development; and Mechanics. The descriptors were paragraphs defining what a student in each category would present as evidence within their portfolio. Assessors using this rubric to mark student work were shown, following repeated testing of their performances against each other, to provide results that met acceptable levels of validity and reliability. This is a complex and time consuming assessment to develop, however, which illustrates the difficulties in assessing portfolios and how new methods of assessment will need to be developed in the context of modern ideas on learning which value individual's experiences and their reflection on them. If we must assess portfolios formally then rubrics and descriptors are something worthy of exploration, but there is clearly a lot to be done in terms of exploring other methods, such as the qualitative approaches suggested by Pitts *et al.*⁵ as we strive to develop robust non-comparative methods of assessment.

In essence we really need to engage in a wider debate of how, and whether, we begin to assess effectively concepts such as professional mastery, clinical and educational performance and continuing professional development, and whether we have to look beyond our traditional horizons for effective solutions. At present our assessment methods stem from the reductionist philosophy that underpins our discipline, and we are, thus, trapped by our need to compare like with like. Until we can make a mental shift that allows us to include a more holistic approach to assessment, one which values the development of individuals over a period of time, we will continue to struggle to measure

the unmeasurable, and may end up measuring the irrelevant because it is easier.

David Snadden
Dundee

References

- 1 Challis M, Mathers NJ, Howe AC, Field NJ. Portfolio-based learning: continuing medical education for general practitioners – a mid-point evaluation. *Med Educ* 1997;**31**:22–6.
- 2 Snadden D, Thomas ML. Portfolio learning – does it work? *Med Educ* 1998;**32**:401–6.
- 3 Grant J. Training the trainers. *Lancet* 1998;**352**(Suppl. iv):19.
- 4 Mathers NJ, Challis MC, Howe AC, Field NJ. Portfolios in continuing medical education – effective and efficient? *Med Educ* 1999;**33**:521–530.
- 5 Pitts J, Coles C, Thomas P. Educational portfolios in the assessment of general practice trainers: reliability of assessors. *Med Educ* 1999;**33**:515–520.
- 6 Schön DA. *Educating the Reflective Practitioner*. San Francisco: Jossey-Bass; 1987.
- 7 Newble D, Jolly B, Wakeford R, eds. *The Certification and Recertification of Doctors*. Cambridge: Cambridge University Press; 1994.
- 8 Cronbach LJ, Wainer H, Braun HI, eds. *Test Validity: Five Perspectives on the Validity Argument*. Hilldale, NJ: Laurence Erlbaum Associates; 1988. p 3–17.
- 9 Cole N, Moss P, Linn R, eds. *Educational Measurement: Bias in Test Use*. 3rd edn. New York: McMillan; 1989. p 201–19.
- 10 Novak JR, Herman JL, Gearhart M. Establishing validity for performance based assessments: an illustration for collections of students writing. *J Educ Res* 1996;**89**:220–33.