References

Bramley, T. (2007). Quantifying marker agreement; terminology, statistics and issues. *Research Matters: A Cambridge Assessment Publication*, **4**, 22–28.

Massey, A.J. & Raikes, N. (2006). *Item-Level Examiner Agreement*. Paper presented at the 2006 Annual Conference of the British Educational Research Association, 6–9 September 2006, University of Warwick, UK.

Newton, P.E. (1996). The Reliability of Marking of General Certificate of Secondary Education Scripts: Mathematics and English. *British Educational Research Journal*, **22**, 4, 405–420.

Fostering communities of practice in examining

Andrew Watts Cambridge Assessment Network

This is a shortened version of a paper given at the International Association for Educational Assessment Conference in May 2006.

The necessity of communities of practice in a judgemental system

The term 'community of practice', when applied to examining in a traditional system, is usually used to denote the system of induction, cooperative working, supervision and development of examiners that aims to overcome the error to which their judgements are prone. Dylan Wiliam wrote in 1996 that 'maintenance of standards requires that those responsible for setting standards are full participants in a community of practice, and are trusted by the users of assessment results'. His observation does not only apply to assessments of school attainment. Alison Wolf (1995), writing about competence-based assessment, describes how assessors 'operate in terms of an internalised, holistic set of concepts'. With examples from a number of educational and vocational contexts she concludes '… how important and, potentially, how effective assessor networks are. They are, in fact, the key element in ensuring consistency of judgement' (p.77).

Subjectivity and objectivity

It has been common to characterise the judgements made in assessment as 'subjective' in contrast to more automated assessments which are 'objective'. Pierre Bourdieu (1990) however, in his analyses of social practice, calls any division between these two concepts 'artificial' and particularly argues against the privileging of an 'objective' standpoint. Sueellen Shay (2005) applies Bourdieu's analysis to the case of a university Engineering Department's assessment of undergraduates' final year theses, which she describes as 'complex tasks'. She describes such assessments within the logic of social practice and asserts that 'all judgement is both objectively and subjectively constituted'. She writes that this kind of professional judgement requires 'a double reading … an iterative movement'. From an objective perspective, assessors can 'observe, measure and map reality independent of the representations of those who live in it'. Subjectively, on the other hand, assessment is 'an embodiment of the assessor'; it is 'relational', 'situational', 'pragmatic' and 'sensitive to the consequences of [the] assessment'. Such 'double readings' enable the judges to assess a 'socially constituted, practical mastery' (p.675).

Shay's concept of a socially based 'double reading' presents us with a *requirement* for assessment to take place within a community of practice. Thus, assessment is understood within a social theory of learning, such as Wenger's (1998), which recognises the place of components like 'community, identity, meaning and practice' (p.5). This supports the view that a balancing of subjective and objective perspectives should be sought in making assessments, and that the community of practice provides an appropriate context for the assessment of complex tasks.

Reliability and the use of new technologies

Concern for greater reliability has motivated the search for more automated ways of managing and marking examination scripts. Paper scripts can be scanned and the images transmitted via a secure Internet connection to markers working on a computer at home. There is then the potential for all examiners to mark the same training scripts online, and for a Team Leader to call up instantly any script that an examiner wishes to discuss with them. Team Leaders may more closely monitor and support examiners during marking, since all marked scripts, together with the marks and annotations, are instantly available. Standardising scripts, with marks already agreed by senior examiners, can be introduced 'blind' into online marking allocations to check that examiners have not drifted from the common standard, and statistical methods for flagging potential aberrant marking may be employed. All these procedures may improve the reliability of marking, but they might also undermine the argument for maintaining a community of practice amongst all examiners. If the bulk of examiners can be trained and effectively monitored online, do they need to come together at all?

Validity as a prime concern

Shay (2004) describes assessment as a 'socially situated interpretive act'. She argues that validation of the assessment is what matters crucially and that the on-going process of evaluating the soundness of our interpretations is a community process. She quotes Bernstein, stating that validation requires 'the existence of a community of enquirers who are able, willing and committed to engage in the argumentation'. She argues that the 'typical technologies of our assessment and moderation systems … privilege reliability' and we fail to use these technologies as 'opportunities for dialogue about what we really value as assessors, individually and as communities of practice' (p.676).

In a paper delivered to the first Cambridge Assessment Conference in October 2005, Alison Wolf noted that 'very often we discuss assessment as an essentially technical affair'. We pursue reliability and lose sight of broader issues like the limitations of what we are testing and the effect of our assessments on those being assessed.

Validity and communities of practice

Wenger's (1998) description of the concept of communities of practice is a dissertation on human learning. Its most challenging thoughts concerning assessment do not refer to the way examiners should learn their trade but to the conditions in which true learning might take place. He says that school curricula, in order to make the process of learning orderly and manageable, often 'reify' the process and thus decrease the possibility that learning which is committed and involved might take place. This can then result in only a limited kind of learning being assessed. Wenger concludes:

> [such learning] can be misleading in that evaluation processes reflecting the structure of a reified curriculum are circular. Students with a literal relation to a subject matter can reproduce reified

knowledge without attempting to gain ownership of its meaning. An evaluation process will become more informative regarding learning that has actually taken place to the extent that its structure does not parallel that of instruction too closely, but instead conforms to the structure of engagement in actual practice and the forms of competence inherent in it. (p. 265)

Whether the performance of a candidate in an assessment 'conforms to the structure of engagement in actual practice' in a domain of knowledge will be something, as we noted in Shay's comments above, that only members of a community of practice will be able to judge. It will therefore be essential that, in the coming changes to assessment practice, the importance of fostering these groups is not overlooked.

References

Bourdieu, P. (1990). *The Logic of Practice*. Stanford, CA: Stanford University Press.

Shay, S. (2004). The assessment of complex performance: a socially-situated interpretive act. *Harvard Educational Review*, **74**, 3, 307–329.

Shay, S. (2005). The Assessment of Complex Tasks: a double reading. *Studies in Higher Education*, **30**, 6, 663–679.

Wenger, E. (1998). *Communities of Practice: Learning, Meaning and Identity*. Cambridge, UK: Cambridge University Press.

Wiliam, D. (1996). Standards in Examinations: a matter of trust? *The Curriculum Journal*, **7**, 3, 293.

Wolf, A. (1995). *Competence-based Assessment*. Buckingham: Open University Press.

Wolf, A. (2005). *What can we measure? And what should we be measuring?* Paper delivered to the Cambridge Assessment Conference. October 17 2005.