FOSTERING COMMUNITIES OF PRACTICE IN EXAMINING

A rationale for developing the use of new technologies
in support of examiners
by Andrew Watts

Paper to be presented at the conference

of the International Association for Educational Assessment,

in Singapore May 21st – 27th 2006

Contact details
Andrew Watts
Director, Cambridge Assessment Network,
1 Hills Road,
Cambridge,
CB1 2EU,
UK

Tel: +44 1223 552749
Email: watts.a@cambridgeassessment.org.uk
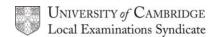Fax: +44 1223 552700
www.cambridgeassessment.org.uk

UNIVERSITY *of* CAMBRIDGE
Local Examinations Syndicate

Abstract

Examiners and assessors who work in teams to judge the quality of students' work in examinations, or of trainees' performance in assessments of competence, are frequently described as working in 'communities of practice'. Following Wenger (*Communities of Practice: Learning, Meaning and Identity*. 1998), this concept is used to describe the way examiners acquire their craft and maintain their competence in it. This paper discusses some of the literature about the place of communities of practice in examining, and seeks to clarify the rationale for them. It does this in the light of the significant changes taking place because of the introduction of new technologies to examining. The concept of communities of practice has often been put forward as a description of the strategies and procedures which lead to reliable marking. The use of e-technology could support such an aim. The paper argues that, at the same time, the necessity of fostering communities of practice to provide a context for valid assessments, can also be supported by new technologies.

Background: judgement and measurement

The choice of the words 'examiners' and 'examining' in the title of this paper deliberately places the discussion in the area of 'traditional' models of assessment, which most significantly require a high degree of judgement on the part of the examiners. The paper follows Spolsky's classification[1] in which he distinguishes a three-stage history of language assessment: 'the traditional or pre-scientific phase'; the 'modern' period, with greater application of psychometric principles; and a 'post-modern' period in which there is an attempt to unify psychometric principles with a greater recognition of 'functionally and socially contextualised testing of language use' (p.353-4).

Spolsky is cited, not because this paper will focus exclusively on language assessment, but because he highlights the place of the examiners' judgements in the traditional period, with 'the use of essays, open-ended examinations, or oral examining, with results determined intuitively by an authorised and authoritarian examiner' (p.353). This tradition goes back to the Chinese Imperial Civil Service examinations which had been in use for eight hundred years when Marco Polo visited China. One of the first mentions of these examinations in England was in Robert Burton's book 'The Anatomy of Melancholy' (1620). He had read about them in the writings of the Jesuit missionary-scholar, Matteo Ricci. Burton made examinations a part of his vision of an ideal society in which local magistrates and 'rectors of benefices [are] chosen out of the Universities, examined and approved, as the literati in China'. It was two hundred years however before examinations came to be similarly used in England, as they had also been in France and Germany, to select suitable people for the Civil Service. In England the use of examinations was extended when the universities, Oxford, Cambridge, Durham and London, began to run them for secondary school pupils. Of these bodies the University of Cambridge Local Examinations Syndicate (now called Cambridge Assessment), whose school examinations began in 1858, has since had a significant part to play in the assessment systems of many other countries.

In reviewing this history, Spolsky shows himself to be post-modern according to his own definition since he argues for assessments which use both counting and judgment. He asks whether assessing language ability is like measuring the 'the time of a race or the distance of a jump or the number of goals' or, in contrast, like judging

---

[1] Spolsky, B. (1995a) *Measured Words.* Oxford University Press, UK. p. 353 – 359.

'the performance of a diver, gymnast or skater'. In relation to students' use of language he says that, 'It is not too hard to realize that both sides of the argument are correct; that there are language abilities that are measurable and there are others that are only judgeable.' (p.353)

The significance of this for our present argument is that the examination systems set up in Europe in the nineteenth century were very much based on the judgements of examiners and there was an assumption that, if the examiners were of sufficient academic quality and reputation, their judgements would be trustworthy. It was not long, however, before doubts were being cast on the reliability of this kind of marking. In 1877 Henry Latham, a lecturer at Cambridge, wrote 'On the Action of Examinations Considered as a Means of Selection'. In this he raised the kinds of questions about the administration and impact of examinations that have since become familiar. One of these questions was the fact that different standards could be applied in their marking by different examiners. Later the economist, F.Y. Edgeworth, described in 'The Statistics of Examinations' (1888) and 'The Element of Chance in Competitive Examinations' (1890), both papers for the Royal Statistical Society, the inevitable error to be found in the judgemental marking of examinations. Edgeworth's description of such unavoidable error expressed the kind of thinking which led Galton, Cattell, Binet and Spearman to seek more objective measures of human ability, in the late C19th and early C20th.

The necessity of communities of practice in a judgemental system
The term 'community of practice' when applied to examiners 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'[2]. His observation does not only apply to assessments of school attainment. Alison Wolf, writing about competence-based assessment describes how assessors 'operate in terms of an internalized, holistic set of concepts'[3]. With examples from a number of educational and vocational contexts she concludes '… how important and,

---

[2] Wiliam, D. (1996) *Standards in Examinations: a matter of trust?* The Curriculum Journal 7 (3) p. 293
[3] Wolf, A. (1995) *Competence-based Assessment.* Buckingham, Open University Press. p. 67

potentially, how effective assessor networks are. They are, in fact, the key element in ensuring consistency of judgement'(p.77).

Subjectivity and objectivity

Spolsky's requirement for a use of both counting and judgement in assessment, quoted above, interweaves the concepts of objectivity and subjectivity. Bourdieu, in his analyses of social practice, calls any division between these two concepts 'artificial' and particularly argues against the privileging of an 'objective' standpoint[4]. 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'.[5] 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, the place of assessment can be understood within a social theory of learning, such as Wenger's which recognises the place of components like 'community, identity, meaning and practice'[6]. Such a theory supports the view that a balancing of subjective and objective perspectives should be sought in making judgements, and that the community of practice provides an appropriate context for the assessment of complex tasks.

We shall now turn to look more closely at some of the features of traditional, judgemental examining which have enabled examiners' communities of practice to come into being and to maintain themselves.

---

[4] Bourdieu, P. (1990) *The Logic of Practice.* Stanford, CA, Stanford University Press. p.25
[5] Shay, S. (2005) *The Assessment of Complex Tasks: a double reading* Studies in Higher Education Vol. 30/6, pp 663-679
[6] Wenger, E. (1998) *Communities of Practice: Learning Meaning and Identity* (Cambridge, CUP) p. 5.

Subject knowledge

At the secondary and tertiary levels of examining, academic qualifications in the subject being examined are required of examiners or, if it is a vocational skill that is being assessed, the assessor must have practical knowledge and experience of the relevant competencies. In this way then examiners or assessors already have a professional identity as they belong to pre-existing communities of practice. Shay links this to Bourdieu's concept of 'habitus', which enables groups to have a common set of 'principles of vision and division' and a common classificatory framework. (Shay: 2005. p.667).

This then is one of the ways in which a community of practice in assessment can be brought together. We will discuss further ways below, but will also note that studies which have sought to establish the importance of these elements for the formation of communities of practice have not always confirmed their importance.

Teaching experience

Experience of teaching has also been seen to be important for assessment communities. This was investigated by Royal-Dawson in 2005[7]. Four groups of examiners were selected all of which shared the same academic background. The groups were: BA English graduates with no teaching experience; English graduates who had just completed their training to be English teachers; English teachers with three years of teaching experience, and English teachers who were also experienced markers of Key Stage 3 tests (for 14-year-olds). The focus of the study was the reliability of their marking and reliability was judged by how far each marker's marks on selected scripts agreed with the marks awarded by the Chief Marker to the same scripts. Overall, the study found 'no overwhelming differences between markers who had differing amounts of teaching experience' and it concluded that 'the training, standardisation and mark scheme are sufficiently rigorous to be effective in preparing non-teaching personnel' to mark at this level.[8] There was one area where teaching experience seemed to make a difference. This was in the marking of an essay on a Shakespeare play where those with teaching experience had higher reliability estimates. Perhaps this demonstrated a more realistic expectation of what 14 year-olds can produce in response to a Shakespeare play.

---

[7] Royal-Dawson, L. (2005) *Is Teaching Experience a Necessity for Markers of Key Stage 3 English?* Qualifications and Curriculum Authority, UK.
[8] Royal-Dawson, *ibid.* p.42.

Well-constructed mark schemes

In the study above a good deal of weight was put on the efficacy of the mark scheme to ensure the standardisation of the examiners' marks. This emphasis goes back to Hartog and Rhodes' critical studies in the 1930s, 'An Examination of Examinations' (1935) and 'The Marks of Examiners' (1936). These studies investigated the marking examination papers in History, Latin, French, Chemistry, English, Arithmetic and Mathematics. The following result from the Arithmetic paper gives the flavour of the findings. This paper had a Part A with '20 straightforward calculations', and a Part B which 'included problems'. Ten markers marked the same 150 scripts in the study. Out of the 150 candidates, 63 were given marks over 80 by one or more of the examiners. However, only 18 were given over 80 marks by all ten of the examiners. The section concludes, 'Supposing we regard 80 as a high mark intended to indicate scholarship level, we find complete agreement among the examiners in regard to only 18 candidates out of the 63 possible.'[9] As a result of the inconsistencies which the studies described, examining bodies began to pay more attention to the careful construction of mark schemes which clarified the criteria by which different levels of performance were to be judged.

Training of examiners

The training of examiners is obviously of crucial importance and within this the reading of exemplar answers which illustrate the different levels of performance that the examiner is likely to meet. Wolf says that key requirements are '... exemplars and networks of assessors – plus a good deal of realism about what can be claimed and achieved'[10]. Discussion of scripts, with reasons given for the award of particular marks to specific answers, is thus seen as an important part of learning the standards of the group. Examiners work under the supervision of Team Leaders and Senior Examiners and samples of their marked scripts are checked to see that they are applying the agreed standards. Feedback on the examiners' own marking of particular scripts is also a crucial factor. These activities are thus ways in which new members are inducted into the community of practice and all are reminded for each marking session of its standards.

---

[9] Hartog, P.J. and Rhodes, E.C., (1936) *The Marks of Examiners.* Macmillan. London. p.77
[10] Wolf, A. (1995) *ibid* p.76

In a study reported on in 2004, Baird, Greatorex and Bell investigated the impact of work on exemplar scripts and of discussion of the mark scheme [11] during the training of examiners. Their first aim was to find if there was any effect on inter-marker reliability when groups of English markers for GCSE (the public examination for 16-year-olds in the UK) were either provided with or not provided with previously marked exemplar scripts. After marking the initial training scripts, each marker marked the same 150 scripts. All their marks were compared to the marks given to the same scripts by the Chief Marker.

The actual and absolute differences between each Assistant Examiner's marks and the Principal Examiner's marks were calculated and analysed. Actual differences are the difference between the Principal Examiner and the Assistant Examiner's marking. Absolute differences are the same as the actual differences but all the negatives are converted to positive values. 'Surprisingly, lack of exemplar scripts did not make marking less accurate, as there were no significant differences in accuracy of marking between groups when the absolute differences were analysed'[12] The above study thus showed that there was little difference in marking consistency between the groups.

In the second study, there were three experimental conditions and groups of GCSE History examiners were selected. One group marked after familiarising themselves with the mark schemes. The other two had the opportunity in addition to discuss some exemplar scripts and the mark schemes together at two different styles of co-ordination meeting. The study concluded that the mark scheme alone 'had a strong standardising effect' (p.344) and that there were no large differences between the reliability of the marks from the different groups. Discussion of the exemplar scripts and the mark scheme therefore appeared not to be an essential part of the markers' training.

However, the writers point out that the markers were already familiar with the format of similar examination papers to those used in the research and thus the communities of practice would already have existed because of their earlier experience. The conclusion of the paper therefore was that 'a well-developed

---

[11] Baird, J-A., Greatorex, J., Bell, J..F. (2004) *What makes marking reliable? Experiments with UK examinations.* Assessment in Education 11 (3)

[12] Baird, J-A., Greatorex, J., Bell, J..F. (2004) *What makes marking reliable? Experiments with UK examinations.* Assessment in Education 11 (3) p. 338-9.

community of practice may not need exemplar candidate work or discussion to bring about accurate marking'. The writers go on, 'What remains to be identified is the relative importance of experience, ownership, feedback, discussion and the other factors that have been discussed in the process of examiners coming to a common understanding of a mark scheme and high levels of inter-rater reliability...' (p.346)

Reliability of marking

The above research used GCSE papers (for 16-year olds) in History and English, whereas Royal Dawson's study above used English papers from national tests for 14-year-olds. For the marking of national tests in England, which began in their present form in 1995, there has been a greater concentration on the development of mark schemes and marker training materials than for other public examinations, not least because, with over 600,000 pupils taking each test, the expense of development is relatively low. In addition, the test development programme for the national tests allows three rounds of pre-testing before the tests go live. Even so, the claim is still made that the marking of these tests is unacceptably unreliable, though it has to be said that it is the marking of English which is often quoted to support this criticism, rather than the marking of Maths or Science. Wiliam claimed in 1995 that the inevitable error in marking national tests could lead to 'less than two-thirds of the students [being] awarded the 'correct level''.[13] He reiterated this judgement in a survey of the working of the national tests in England for the Association of Teachers and Lecturers in 2001, in which he said that 'It is likely that the proportion of students awarded a level higher or lower than they should be because of the unreliability of the test is at least 30% at key stage 2 (11 year-olds) and may be as high as 40% at key stage 3 (14 year-olds).[14] The possible conclusion then is that, even with thoroughly developed mark schemes, when it comes to judgemental marking, communities of practice find it difficult to maintain standards as reliably as is required. (It should be pointed out that the marked national tests are returned straight to schools by the markers, who have only a sample of their scripts checked. There is no scaling of individual's marks, as there is in the public examination system, to counteract the effects of markers' severities or leniencies.)

---

[13] Wiliam, D. (1995) *It'll all end in tiers.* British Journal of Curriculum and Assessment 5 (3) 21-4
[14] Wiliam, D. (2001) *Level best? Levels of Attainment in National Assessment.* Association of Teachers and Lecturers. UK. p.16

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. Systems are currently in use which involve the scanning of scripts, the separation of different kinds of questions and their distribution to different markers.[15] This enables questions to be marked in different ways. The distinction is made between questions which can be marked automatically, such as multiple choice questions, and those that require professional judgement. Between these two categories there can be classes of question which could be marked by trained clerical markers who do not have any subject expertise, or by 'subject markers', who are graduates in the subject but who have no teaching or examining experience. The fact that the 'expert markers' can then concentrate on the answers which demand more professional judgement would suggest that the job can thus be done more productively.

One of the main arguments advanced for this kind of use of technology is that the training and supervision of markers can be more intensively done. There is the potential for any number of training scripts to be delivered on-line to examiners, without the time and expense of them travelling to training meetings. Standardising scripts, already marked by the Senior Marker, can be introduced into the on-line marking allocations to check that the examiners have not drifted from the common standard. Team Leaders can read the work of the examiners and intervene if necessary at any point in the process. All these procedures are seen as ways of improving the reliability of the marking. They could, however, also be seen as undermining the argument for maintaining the community of practice. If examiners can be trained on-line, using downloaded training scripts, why should they come together at all?

On-line communities of practice

It is important in the light of the innovations described above to be clear about what a community of practice of examiners is and how it contributes to the marking process. Firstly, however, we must distinguish between those markers required to make judgements ('expert' or 'subject' markers) and those performing a merely clerical role. Provided such classes of questions can be distinguished, it will be those markers making judgements that will still need to part of an examiners' community of practice.

---

[15] See the claims made, for example, by DRS Data and Research Services plc, for their 'e-Marker®' product. UK. 2005

What we have said above about communities of marking practice will then still apply to the groups of expert markers. Thus it will be important to note what pre-exists for this particular community. They will be members of an academic tradition and they may well have teaching experience. For the future they will be encouraged to gain experience in marking and not just do the job for a year or two. Their training will include the careful study of exemplar, marked scripts and the marking of further sample scripts followed by discussion with Team Leaders, or maybe their peers, though now the discussion may be on-line.

Acceptability of new marking practices

This leads to the question of how far such communities of practice depend on face-to-face meetings. It could certainly be that training could be more conveniently delivered on-line, and more thoroughly too. Meetings, if they are needed at all, might then be run for other purposes, like professional development for those teaching a particular subject. Broader issues will have to be considered, though, like the effect on the motivation of examiners from the possible loss of the collegial relationships of marking teams and of meeting with others who teach their particular subject. Baird et al. noted in their study that 'questionnaire responses revealed that examiners did not like marking without the opportunity the coordination meeting provides to maintain a community of practice through discussion with colleagues.' [16]

Another broader issue is the way the public will perceive such changes. In England, press reaction to the idea that examination scripts have been marked by 'non-qualified markers' (i.e. they were not teachers of the subject being examined) has been negative. As we noted above, Wiliam has stated that those involved in standard setting must be trusted by the users of assessment results. In that article he links the valuing of the test results by their users to the idea that standard setting is 'neither objective nor subjective, but rather intersubjective' and he suggests that tests have to demonstrate their 'provenance' to their users [17]. We might apply these comments to those marking the scripts, as well as to those responsible for setting the standards, and suggest that the public has a sense of the kind of community of practice they will trust to examine. Thus, assessment agencies need to gain the support of the users for new approaches to examining, and this will include explaining the part that examiners' judgements will play in a more technology-supported system.

---

[16] Baird et al. (2004) *ibid.* p.345.
[17] Wiliam, D. (1996) *ibid*. Pages 293 and 304.

Validity as a prime concern

Shay describes assessment as a 'socially situated interpretive act'[18]. 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 a conference in Cambridge 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.[19] The question of what our assessments mean can thus be overlooked. In an article about judging learners' work on screen, Johnson and Greatorex note that much of the literature appears to conclude that marking scores are comparable whether the marking is done on paper or on screen. But they also point out that what matters more is the validity of the judgements made by assessors rather than the reliability of their marking. They conclude, 'Our review suggests that judgements made on screen and conventionally on paper are qualitatively different. If it is accepted that validity is more important than reliability then it is a priority to investigate how assessors make assessment judgements and what they judge, to identify what is actually measured and how technological developments might facilitate this.'[20]

Validity and communities of practice

In considering the place of communities of practice in establishing the validity of assessments, let us return to Spolsky's writing about the assessment of language competence. In the introduction to a comparison of English tests designed according to examination and to psychometric test models, he suggests that all such assessments must take into consideration 'the close relations of tests to teaching, the

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

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

[20] Johnson, M. and Greatorex, J. (2005) *Judging learners' work on screen. How valid and fair are assessment judgements?* Paper presented to the British Educational Research Association annual conference, University of Glasgow. p.10.

emphasis on content validity, the recognition of the need for differences in standards of judgement and taste.'[21] In his history of objective language testing[22], Spolsky concludes by moving from his three historical periods to a description of three competing sets of forces. They are:

1. feasibility: the institutional context, requiring compromise to satisfy practical concerns;
2. usability: the need for test results to be straightforward to understand and 'feel fair' to the users;
3. reliability: the recognition of Edgeworth's principle of 'unavoidable uncertainty'.

He says that all three of the above forces constrain the possibility of developing a valid test, and all three are important. Spolsky also draws attention to a post-modern 'sincere, ethically-driven consideration of the potentially deleterious effects of testing on the test taker, on the instructional process, and on other facets of the social context in which we test'. He concludes that the tension between these forces is 'worth maintaining' because a complex human behaviour, like language use, requires recognition of the 'contextualisation, dynamism, reciprocity' which are parts of knowing. All of the above factors will 'mitigate against measurement', and it will be the task of examiners to judge how to balance them within a single assessment.

Wenger's 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

---

[21] Spolsky, B. (1995b) in Studies in Language Testing 1: *An investigation into the comparability of two tests of English as a Foreign Language*, by Bachman, Davidson, Ryan and Choi. Series ed. Milanovic, M. Cambridge University Press. UK. p.13
[22] Spolsky, B. (1995a) *Measured Words.* Oxford University Press. UK. p. 354 - 356

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.' [23]

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 will not be overlooked.

New technology to support valid assessment

We have already noted that technology could play a part in facilitating communities of practice, with the added value that problems of time and space will not seriously hamper them. In a paper entitled 'Creating a virtual community of assessment practice: towards 'on-line' examiner reliability', Shaw[24] describes the kind of on-line communication tools, such as video conferencing, chat rooms, web-based bulletin boards and e-mail discussion lists, which could be used with on-line assessment communities. These certainly could be used to create and foster examiners' communities of practice, and Shaw describes plans that are being put into place to effect that. It is noteworthy that he proposes this for the sake of greater reliability in the assessments. The argument of this paper is that such communities will, in addition, be a significant support for a socially-based process of examining, and therefore for the validity of the assessments.

Conclusion

It is evident that in assessment systems in which judgements are required, the role of communities of practice for examiners will continue to be important, alongside the technological changes that are taking place with the intention of improving the reliability of marking. There are potential advantages in clarifying the kinds of marker needed for different kinds of questions. Such distinctions might well highlight the importance of 'expert' judgements and reaffirm that they cannot be made satisfactorily from within an under-professionalised, 'cottage industry'[25]. Examiners who will be able to respond to the demands set out by the writers whom we have

---

[23] Wenger, E. (1998) *Communities of Practice: Learning, Meaning and Identity*. Cambridge, UK. Cambridge University Press. p.265

[24] Shaw, S. (2004) *Research Notes*, Issue 15, February 2004, p. 18-20 University of Cambridge ESOL Examinations.

[25] Spolsky, B. (1995a) *ibid.* quoting Alan Maley. p.341.

discussed above, will need to have an understanding of the place of their work in its wider context and knowledge of the underlying principles of assessment. They will also need to be committed to involvement in a community of markers and be willing to analyse and improve their own performance, to develop their knowledge of assessment and to play a part in passing on their knowledge to others. It is a model of a community of practice at work with which teams of 'traditional' examiners will be familiar.