

# **Issues in the assessment of practice-based professional learning**

Mantz Yorke

A Report prepared for the  
Practice-based Professional Learning CETL  
at the Open University

November 2005

## Issues in the assessment of practice-based professional learning

### Contents

1	Issues for possible exploration	3
2	Work-based learning and curricula in higher education	6
3	The bifocal perspective	7
4	Purposes of assessment	12
5	Competence	14
6	Levels and criteria	18
7	Philosophical and methodological issues	21
8	Portfolios for summative assessment	27
9	Self-assessment	30
10	Assessing more than the individual	31
11	Some technical issues relating to assessment	31
12	Practitioners as assessors	37
13	The workplace as a learning opportunity	41
14	Concluding comment	43
	References	44

---

## 1 Issues for possible exploration

The literature review in succeeding sections of this Report suggests, in some places explicitly and in others more implicitly, a number of issues relating to practice-based professional learning [PBPL] that could be further explored. The list below, which stems from the review, offers some signals to the reader regarding points of possible interest: some issues refer directly to assessment, whereas in others there are implications for assessment. However, this list does not by any means exhaust the potential of the field for useful research and development work.

*Issues which are probably amenable to projects that could be completed within one year*

- A. What model(s) of assessment underlie the assessment of practice-based professional learning [PBPL]? Hager and Butler's (1996) ideal types could provide a starting-point for inquiry. A single approach may not provide an adequate understanding of the assessment process: if this is the case, how are different approaches combined in the assessment regime?
- B. How is the assessment of workplace achievement being undertaken in selected foundation degree programmes, the selection spanning subject areas in which there is and in which there is not a tradition of assessing in workplaces. What challenges do assessors perceive, as regards assessment practices, and how are these overcome?
- C. What proportion of workplace experiences accords with institutional expectations, and what does not? To what extent is the assessment regime operated by the institution able to cater for placement variability?
- D. How are portfolios used in PBPL, and in the assessment of student achievement? Where are there tensions, and where have these been resolved satisfactorily? Is there a bank of 'good practice' which can be compiled for general use?
- E. Is it possible to characterise placements according to their propensity to support students' growth as practitioners? The perceptions of students on PBPL placements could perhaps have value in the further development of placement experiences. There are more indirect implications for assessment, since judgements of 'student success' need to be made with reference to the level of demand made of students by the placement experience. A further use might be to encourage *students* to look actively at their placement, and to consider ways in which it could be developed into a better learning experience.

- 
- F. If the placement experience (*inter alia*) is regarded as a site for personal growth and the development of professional autonomy, to what extent does it foster the growth of metacognitive capability, including the capacity for self-assessment? What features of the placement experience encourage metacognition, and what discourage it?
- G. How, and under what circumstances, is credit awarded for work-based learning? The survey by Johnson (2004) indicates that there is growth in this aspect of higher education, but it provides no detail about what the award of credit entails.
- H. How does the assessment process deal with the marginal performer in PBPL? What considerations determine whether an assessee falls on one side or the other of the pass/fail boundary? In what circumstances is the pass/fail decision deferred, and what are the implications of any such deferral?
- I. There is little in the literature on students' experiences of being assessed in PBPL situations. Since what is written in assessment specifications (and understood by assessors) may be understood differently by students, 'assessment-in-use' as perceived by students may be a worthwhile topic for investigation.
- J. The literature suggests that the training of workplace assessors (and, to some extent, academics) has been patchy. As the use of PBPL increases, it seems important to gain a stronger appreciation of the extent of training in assessment for it, and of the strengths, weaknesses, opportunities and threats [SWOT] that apply.
- K. In relation to the above, particular attention might be given to the extent to which mentors in the workplace 'double up' the roles of formative and summative assessment, and the extent to which this is perceived by those involved a problem.
- L. The duality of role may be particularly marked in respect of the practitioner/teacher appointments that exist, largely (but not exclusively) in the health arena. An issue susceptible to research is how such appointees enact their role, with particular reference to their role as formative and/or summative assessors.
- M. How is the assessment of PBPL undertaken when the work undertaken is group-based? What are the strengths and weaknesses of the methods used, and what might be done to mitigate the latter?
- N. How are summative assessments of performances in the workplace, some of which may relate to evanescent actions, made available for moderation or confirmation?

*Some issues which are probably outside the scope of short-term projects*

- O. Institutions may have, in their record systems, data that would allow some estimation to be made of the extent to which there is a 'placement effect' on academic performance. The unavoidable weakness, however, is that the placement experience is confounded with student maturation, and teasing out the relative contributions of the placement experience and the extra period (typically a year) would be impossible. Whilst an analysis of existing data might help to focus on the value or otherwise of the placement, it may be possible to conduct a qualitative study of placement experiences and to link this to pre- and post-placement performance. The timescale of such a study, though, would seem to require in excess of one year.
- P. To what extent is the concept of 'level' meaningful in PBPL? Is it possible to produce descriptors that have a broad utility?
- Q. In what ways can learning in the workplace and in academic institutions be accorded a rough parity of value in award systems?
- R. What might a 'connective pedagogy' (Griffiths and Guile 2004) that engages with both academia and the workplace look like?
- S. How might summative assessments from different curriculum components (including PBPL) most usefully be brought together? ('Brought together' does not necessarily imply coalescence into a single index of performance.)

---

## 2 Work-based learning and curricula in higher education

Higher education exhibits a varied approach to work-based learning. In some subject areas (such as medicine, nursing, social work and teacher education), the development of professional practice is built into curricular structures, and students qualify as professionals provided that they demonstrate that they have developed an adequate level of professional capability in the relevant workplace. In others, such as business studies and engineering, the first degree programme is often of a 'sandwich' nature in which one or more periods of work placement are included prior to the final year of academic study. Here the position of work-based learning is less clear, since the academic credit that is earned can be disproportionately small in relation to the time spent on learning in a workplace environment. In a third group of subjects, experience in workplace and other situations outside higher education is treated as a separate matter from the academic programme, save where credit is awarded for, say, a module's worth of workplace experience unrelated to the academic discipline which has been distilled into a reflective account of what has been learned<sup>1</sup>.

The foundation degrees introduced in 2000 in England, Wales and Northern Ireland made it explicit that students would engage in work-based learning<sup>2</sup> and that this would be assessed as part of the qualification (HEFCE 2000). Nursing, social care and education have been particularly strong growth areas for foundation degrees, but these already have a tradition of assessing practice-based expertise. Other areas in which foundation degrees have been developed do not have such a tradition.

Since the then Manpower Services Commission introduced the Enterprise in Higher Education initiative in the late 1980s, there has been a growth of acceptance across higher education that students should develop the kinds of capabilities that employers claim to value (as expressed in, for example, Hawkins and Winter 1995; Harvey et al 1997; NCIHE 1997). Whilst the effects have become increasingly evident in first degree curricula and their associated pedagogic practices, formal valuation of student capabilities focusing on the wants of employers can often be found in separate qualifications, some run by individual institutions for their own students and others run by organisations for students from any institution. Institution-specific awards exist, for example, at York, Warwick, Essex, Glamorgan and Leicester Universities, whereas external award schemes have been devised by City and Guilds of London Institute and CRAC.

---

<sup>1</sup> See, for example, Marshall and Cooper (2001).

<sup>2</sup> In some instances, such as when small enterprises have been involved, practical difficulties have meant that this has had to be interpreted as 'work-related learning', with practice-relevant work being undertaken at the educational institution.

### 3 The bifocal perspective

Work-based learning covers a spectrum of activities, amongst which are sandwich courses, co-operative programmes<sup>3</sup>, job-shadowing, joint industry/university courses, new traineeships and apprenticeships, placements/practica, fieldwork, post-course internships, on-the-job training and work experience unrelated to academic study. PBPL involves some of these, but is focused specifically on work-based learning connected with the development of expertise as a professional, either as a part of first degree studies or in connection with continuing professional development. Some of this learning may be undertaken in formal educational programmes of varying kinds: however, where ongoing professional development is concerned, the larger proportion of learning may be undertaken in workplace settings and be less formalised.

PBPL (along with much other work-based learning) involves the fusion of two different perspectives on learning, the academic and the practice-based. Each has its own way of going about things and tends to view the other's as different, even alien. Billett (2004b) remarks that workplaces have their own practices and that these do not necessarily align with those of educational institutions, hence the latter sometimes see the former (inappropriately) as *ad hoc*. There is, with the exception of some well-established professions, a tension between education-led and employer-led views of curriculum.

An issue of importance which is not given a lot of explicit attention in the literature, is 'where the student is coming from' as regards the relationship of work-based learning to academic study. Students on formal academic programmes are likely to look at work-based learning from the standpoint of how what they have learned from their academic studies might (or might not) apply in work settings. In contrast, students engaging with higher education from a base as an employee are likely to look for ways in which (possibly multiple) academic disciplines can assist them in dealing with the challenges thrown up by the workplace, which will often be 'messy' and multidimensional<sup>4</sup>. This comes close to the contrast between what Gibbons et al (1994) characterised in terms of Mode 1 and Mode 2 approaches to knowledge generation<sup>5</sup>. As Lyons and Bement (2001, p.169) observe, in respect of employed students:

---

<sup>3</sup> Zegwaard et al (2003) note that co-operative placements are particularly useful for the appreciation of workplace cultures and for the development of the so-called 'soft skills'. They suggest that academics and workplaces emphasise different kinds of achievements; that the concern for rigorous evidence-based assessments distracts from other kinds of achievement that are harder to evidence; that the relatively trivial may be valued (because it can be evidenced) at the expense of the complex. They propose three models of assessment:

- a. a template of competences based on employers' preferences
- b. assessment based on a set of objectives agreed between the employer and the student (which, further, requires agreement on the meanings to be placed upon terms such as 'competent', and is accompanied by moderation on the part of the academic institution)
- c. a portfolio of experiences and achievements (which could be based upon Model [b], above).

<sup>4</sup> In such circumstances, Gwenlan's (1993) claim that educational institutions can only seek to assess competence for which they have provided learning opportunities seems misconceived.

<sup>5</sup> Lester (1995) contrasts 'technocratic' and 'competence-led' approaches to professional development, the former appearing to align with Mode 1, the latter with Mode 2. His terminology seems inappropriate.

... [work-based learning] is not about ‘experiencing the world of work’, or putting theory into practice, but about learning to develop career and professional standing and how to put practice in a theoretical context.

The corollary is that there can be a tension between academics and employers as to the merits of projects that span academic and employment interests. Lyons and Bement illustrate this in a vignette relating to the adaptation of a military satellite navigation system, in which a valued commercial development was not rated highly in academic terms (ibid, p.169).

The contrast between the two perspectives of the previous paragraph is schematised, albeit in probably too stark a manner, in Figure 1.

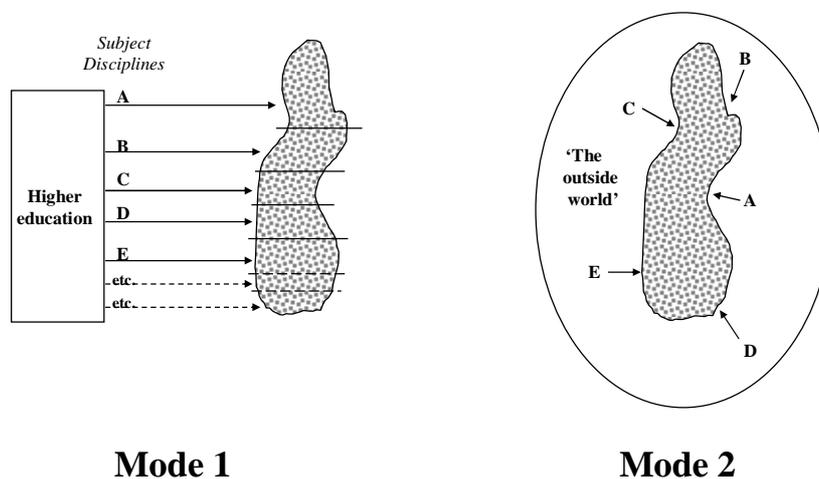


Figure 1 A schematisation of the difference between Modes 1 and 2 of knowledge generation and how the approaches to a problem might differ according to whether it is approached from an academic or an employer perspective. Whereas individual disciplines might attack slices of the problem, solving problems in everyday life may require a multidisciplinary attack.

Eraut (2004b) draws attention to the view held by some that learning relates to formal education and training, and that working and learning are activities that never overlap, but points out that he and colleagues have found the opposite – i.e. that most workplace learning occurs on-the-job rather than off-the-job. A lot of informal learning goes on in formal environments, whether they be educational or workplace environments (Billett 2004b). Hager (1998) makes the distinction between workplace learning and on-the-job training. He places the latter alongside formal education because of, *inter alia*, the

control that the teachers exercise over the prescription of what is to be learned and the learning itself. Workplace learning is much more contextualised, and is often collaborative. Formal and on-the-job learning is about theory and knowledge and their application, whereas workplace learning relates to the gaining of seamless 'know-how' which is closer to the Aristotelian concept of practical wisdom, or 'phronesis'. Hence he finds it unsurprising that formal learning and training are valued more highly than workplace learning.

Boud and Symes (2000, p.23) say

One of the more important ... challenges is that of acceptance, of establishing the credentials of work-based learning and gaining its legitimacy in the university setting. Unless a work-based award is in some sense equivalent to other qualifications and accepted as such, the reason for having the university involved lessens...

However, some smaller organisations may be less enthusiastic about formal credentials than their larger counterparts. Billett et al (2003, p.152) noted that 'the need for credentials, around which so much of vocational education and training provisions are based, may be of limited importance to small business operators.'<sup>6</sup> Eraut et al (2000) noted that the learning environments of small organisations were sufficiently different from the larger ones they were investigating that a separate study would be needed for them.

The focus of employers' interest, as expressed in different ways, often centres upon a range of personal qualities and attributes: the view that comes across could be summed up in words such as 'give me an intelligent and able person, and the organisation will provide the development to enable them to fulfil the required role. What is usually unclear is the extent to which being a graduate in Subject X is taken for granted.'<sup>7</sup>

Gerrish et al (1997) discussed the possibility of awarding a degree on the basis of an even division between academic and practical work, but indicated their concern that this would devalue the standing of the degree, which is perhaps why Canham (2002) was able to state that, in the nursing field, very few institutions ascribed an equivalence of value when assessing theory and practice. There may be a connection with empirical observations that gradings for coursework tend to be higher than those awarded for examinations (e.g. Elton 1998; Bridges et al 2002).

In its *Post-qualifying framework for social work education and training*, the General Social Care Council makes very clear its intention that assessment should incorporate a

---

<sup>6</sup> Billett et al are writing of the Australian context, where a small business is characterised by fewer than 20 employees, in contrast to the UK where the criterion is fewer than 100 employees.

<sup>7</sup> See, for example, the survey of employers by Harvey et al (1997), and surveys of advertisements for graduate jobs conducted by Purcell and Pitcher (1996) and Bennett (2002). Further, in a presentation in 2004 to the Enhancing Student Employability Co-ordination Team [ESECT], Norman Mackel of the Federation of Small Businesses provided survey data that aligned with the comment of Billett et al (2003) about credentials and earlier findings by Yorke (1999) from a survey of small and medium-sized enterprises on Merseyside.

substantial amount of evidence from the workplace (see GSCC 2005, para 84). Manthorpe et al (forthcoming) see some convergence between policy initiatives in social work and those longer-established for nursing, notably a greater emphasis on practice-based learning.

#### *Control over the placement experience*

Institutions seem not to be able to control the general nature of all of the placements that form part of their curricula (see, for example, Schaafsma 1996, regarding the University of Technology, Sydney; Fowler and Tietze 1996, regarding business degrees at Sheffield Hallam University). The proportion of placements that accord with institutional expectations is, however, unclear. If there is uncertainty about the quality of placements, then consequentially this must extend to whatever assessment regime is in place.

Rickard (2002) noted that the occasional agency was unprepared for the placement and lacked experience in providing placements. Gammie and Hornby (1994) noted the problem for assessment of variation in quality of the workplace experience – a quarter of students on business degrees at their institution were faced with placements that gave them relatively little responsibility and challenge. Smith and Betts (2000), however, argue that the quality of the learning is not dependent on the quality of the experience but on the quality of the process of reflection in relation to the agreed learning outcomes: however, their argument is tenable only to the extent that the placement is able to offer experiences that are consistent with the intended learning outcomes.

#### *The award of credit*

The award of credit for achievement in respect of work placements has been problematic wherever the work placement has not been an integral part of the programme (as it is in nursing, social work and teacher education, for example). Sandwich course placements have presented particular difficulties, because the placement year has been construed as of very limited relevance to an academic assessment.

Gomez and Lush (2005) describe how they have attempted to enhance the credit status of sandwich degrees in science at the University of the West of England. The award of notional credit for the sandwich year reflected the fact that the placement experience fell largely outside institutional control, and hence could not be subjected to the rigorous quality assurance procedures operated in respect of the academic components of the degrees. The assessment expectations for the placement year were typical: a report from a visiting tutor; a report from a work supervisor; and a report written by the student: students were expected to pass on all three components. Weaknesses in the process are familiar:

- very limited communication between student and university
- participants on their 'best behaviour' when meetings took place at the placement site
- the work supervisor's report being produced too late to influence learning
- the final report tending to focus on 'highlights'.

Overcoming these weaknesses led to the production of a learning agreement which was in form tantamount to the specification for a module at Level 3, and hence could be brought within the university's modular framework. The evidencing of placement learning was realised through the development of an e-portfolio<sup>8</sup> owned by the student but which tutor and supervisor could be invited to read.

Johnson (2004) surveyed credit practice in England, Wales and Northern Ireland, and found a recent growth in the number of institutions awarding credit for learning achieved through work experience (between 1999 and 2003, from 55% to 68%: for in-company programmes, the rise was from 46% to 54%). Concomitantly, there was a growth in the number of institutions grading work-based learning, thus enabling it to figure in calculations of degree classification. The advent of foundation degrees has led to greater attention being paid to the issue of the assessment of work-based learning.

### *Straddling boundaries*

Brewer (1994, p.62<sup>9</sup>) raised an issue that continues to exercise higher education when he said:

The real question is, should performance on placement be counted toward the degree qualification, or to some separate form of certification ... The larger question which this leads to is what is a degree, what does it measure? ... It is clear that differing assessment methods all produce differing outcomes, some of which lend themselves more comfortably to differing kinds of accreditation.

The nature of the first degree remains an issue of debate in contemporary higher education, despite the development that has taken place in many curricula<sup>10</sup> in response to the policy expectation that (young) graduates should leave higher education with enhanced employability<sup>11</sup>. The first degree has undergone a slow transformation towards accepting that capabilities that relate to successful performance in the workplace align quite well with those that relate to success in academic terms. The shift in perspective has meant that first degree curricula, as well as being seen through the lens of academic study, are increasingly being seen through the lens of external expectations. In other words, they are beginning to straddle the divide perceived by writers on the development of practical knowledge (e.g. Eraut et al 2000; Eraut 2004a; Billett 2004a, b; Hager 1998, 2004a, b) between academic and practical knowledge or, from Sternberg's (1997) perspective, between academic and successful intelligence. Griffiths and Guile (2004), implicitly suggest moving further. They see learners as 'boundary crossers' between education and the workplace, and argue for what they term a 'connective pedagogy' to achieve this. What such a pedagogy might look like in practice is a matter for further thought and curricular development.

---

<sup>8</sup> Details at [www.profile.ac.uk](http://www.profile.ac.uk). This project is supported by FDTL4 funding.

<sup>9</sup> Quoted in Toohey et al (1996, pp.215-6) and Gerrish et al (1997, p.27).

<sup>10</sup> With the exception, of course, of subject areas which have a longstanding commitment to PBPL.

<sup>11</sup> Seen by the Enhancing Student Employability Co-ordination Team [ESECT] as 'a set of achievements — skills, understandings and personal attributes — that make graduates more likely to gain employment and be successful in their chosen occupations...' (Yorke 2004, p.7).

---

Following Griffiths and Guile would imply taking up the challenge thrown out by Eraut et al (2000) that simplistic assumptions about the dominance of the codified knowledge and learning derived from formal education and training should be abandoned. There are, of course, harbingers. Institutions have for a long time awarded honorary degrees to many who have made a significant contribution to the public services, the arts, and industry, yet have not taken the academic route to the development of their expertise. Business schools run MBA programmes, and many departments run master's and doctoral programmes, based on the practical demands of participants' workplaces. Degrees by independent study (latterly in decline) allowed individuals to base a programme of baccalaureate study on workplace interests. A foundation degree in management is run at the University of Central England in which the pedagogic approach is through action learning sets whose learning agenda are driven by problems thrown up in workplaces. Other foundation degrees (particularly part-time programmes) draw extensively on the workplace in order to blend practical experience with academic study. When the structure of student funding is changing, what is to stop more being done to enable students to develop themselves as professionals concomitantly with employment, using their employment situation as the stimulus to academic study?<sup>12</sup> How might the dominant, bifocal, view of higher education and workplace be developed? Could not stereoscopy offer a better perspective?

#### **4 Purposes of assessment**

Assessment is used for three main groups of reasons: to promote learning; to certify achievements; and to provide data that can be used for quality assurance (sometimes, quality control) purposes (Table 1).

There are obvious tensions within Table 1 – for example, between the formative and summative assessment of students, and between satisfying the expectations of the various interested parties. The tension is apparent at various points in the assessment of PBPL: for example, as many professional learning situations develop, assessors shift their primary role from that of provider of formative feedback to that of summative judge.

The assessment of practice-based learning is influenced by a variety of pressures, some philosophical as movements have gained sway, and some as a consequence of political pressures deriving from tragedies in the social arena such as the case of Victoria Climbié<sup>13</sup>. In the social arena, concerns relating to public safety have pressed assessments towards greater stringency.

---

<sup>12</sup> Yorke (2003) has argued the case for making part-time employment a component of a joint-honours degree in Subject X and work-based learning.

<sup>13</sup> Howe (1998) suggests that, in social work, the biggest influence has been the political reaction to social tragedies.

<b>Broad purpose</b>	<b>More detailed purpose</b>
<b>Learning</b>	<ul style="list-style-type: none"> <li>To motivate students</li> <li>To diagnose strengths and weaknesses</li> <li>To provide feedback</li> <li>To consolidate work done to date</li> <li>To help students develop their capacity for self-assessment</li> <li>To establish the level of achievement at the end of a unit of study</li> </ul>
<b>Certification</b>	<ul style="list-style-type: none"> <li>To establish the level of achievement at the end of a programme of study</li> <li>To pass or fail a student</li> <li>To grade or rank a student (with reference to norms and/or criteria)</li> <li>To underwrite a 'licence to practise'</li> <li>To demonstrate conformity with external regulations, such as those of a professional or statutory body</li> <li>To select for employment, further educational activity, etc.</li> <li>To predict future performance</li> </ul>
<b>Quality assurance</b>	<ul style="list-style-type: none"> <li>To assess the extent to which a programme's aims have been achieved</li> <li>To judge the effectiveness of the learning environment</li> <li>To provide feedback to teachers regarding their personal effectiveness</li> <li>To monitor levels of achievement over time</li> <li>To assure interested parties that the programme or unit of study is of an appropriate standard</li> <li>To protect the relevant profession</li> <li>To protect the public</li> </ul>

Table 1 Purposes of assessment

Note: This Table draws on Atkins et al (1993), Brown et al (1997, p.11), Yorke (1998, p.178) and Micklin and Kenworthy (2000, pp.108-9).

When students' performance is attested (typically by some form of certification), receivers of the attestation's 'message' can to some extent be forgiven for inferring, *contra* the warnings given regarding financial investment, that past success is a predictor of future performance. In more technical language, the assessment is seen as having both summative and predictive validity. In some fields, the blurring of the two may be of limited import: in others, it may literally be vital (a matter that is discussed below).

#### *Assessment drives what students do*

When discussing assessment, the reactive effect of the assessment regime on student behaviour cannot be ignored. It is widely appreciated that students' behaviour regarding assessments is strongly influenced by their perceptions of the demands of assessment and

of the implications that the assessment regime has for the grading of their performance. Miller and Parlett's (1974) study of student approaches to assessment was an early exemplification of the point, and Graham Gibbs has collected a range of comments from students that indicate the enduring nature of the issue.

An illustration of how assessors' intentions can be subverted is given by Newble and Jaeger (1983). They describe changes in the final year of the medical curriculum at the University of Adelaide, which took the form of internships with four-weekly ward-based assessments, together with a final examination which was predominantly multiple-choice in format. Unsatisfactory performance on the ward would have led to a viva which would determine whether the student passed or failed. This approach was believed to be a more valid method of assessing clinical practice than its predecessor. In practice, few students proved unsatisfactory on the wards (assessors were reluctant to rate students as 'unsatisfactory', preferring to pass the decision to the viva). Realising that academic success in reality depended on the final examination, students spent time in the library swotting up for the examination and demanding didactic lectures instead of widening their ward experience – exactly what the curriculum designers did not want. This led the curriculum designers to develop a clinical test with 'stations' and a final examination focusing on clinical content, thereby bringing the validity of the assessments into closer alignment with the assessors' original intentions.

## 5 Competence

### *Contested meanings*

A term widely used in the relevant literature is 'competence'. However, the meanings ascribed to it vary, which makes it difficult for the reader in this field. In some contexts, such as in North America and in management it refers to a personal attribute or quality<sup>14</sup>, whereas elsewhere it refers to social expectations (Eraut 2004b). In the UK, the latter can be further divided into the plethora of detailed competences of the type developed for national Vocational Qualifications [NVQs] and a broader interpretation in which components of performance are 'bundled together'. A number of writers (e.g. Chambers 1998; Watson 2002; Hager 1998, 2004a, b) have acknowledged the varied understandings associated with learning in the workplace and with competence: Hager (2004a) points to the need to differentiate between performance and its outcomes; the underpinning constituents of competence (i.e. capabilities, abilities and skills); and the development of people to be competent performers.

Whilst there is a general acceptance that 'competence' is a social construct (e.g. Lum 1999; Hager 2004a), and hence not value-free (e.g. Kemshall 1993), the literature bears witness to a tussle over its theoretical base. Following a spirited defence by Hager et al (1994) of competence-based assessment, Hager (2004a) takes issue with Hyland's argument that competence is based in behaviourism (e.g. Hyland 1994), arguing that

---

<sup>14</sup> The term 'competency' has also been used in this sense.

Hyland makes too strong a connection between the manifestation of competence (i.e. behaviour) and behaviourism. Indeed, Hager goes so far as to offer Jessup's (1991) approach to competence (which has a strong affinity with the behavioural objectives approach espoused by Mager 1962/1990 and others) the prospect of redemption by suggesting that, if Jessup's outcomes of performance and attainment are construed as performance descriptors, they have an abstractness that takes them some distance away from the narrowness that most critics of National Vocational Qualifications [NVQs] have perceived in them. Some might feel that Hager goes too far.

### *Complexity*

Worth Butler et al (1994, pp.226-7)<sup>15</sup> get close to the complexity underlying competence:

... mastery of requirements for effective functioning, in the varied circumstances of the real world, and in a range of contexts and organizations. It involves not only observable behaviour which can be measured, but also unobservable attributes including attitudes, values, judgemental ability and personal dispositions: that is, not only performance, but capability.

Whilst highly disaggregated competences, such as those that were introduced in NVQs, have value in helping people to understand the dimensions of workplace performance, and there has been the occasional suggestion that the NVQ framework might be used in the award of credit for achievements in work placements (Aslin et al 1995), there is widespread support for seeing 'competence' in much broader terms<sup>16</sup>. Eraut (2004a, p.804), for example, writes:

... treating [required competences] as separate bundles of knowledge and skills for assessment purposes fails to recognize that complex professional actions require more than several different areas of knowledge and skills. They all have to be integrated together in larger, more complex chunks of behaviour.

Others have acknowledged that competence or competency frameworks derived from functional analysis are inadequate for assessment purposes, missing some of the subtleties of performance (e.g. Owens 1995, in respect of social work; Jones 2001 and Coll et al 2002, in respect of teaching; Jones 1999 in respect of vocational education and training; Lang and Woolston 2005, in respect of policing in Australia).

### *Competence and performance*

Some writers (e.g. Hays et al 2002; Schuwirth et al 2002) make a distinction between 'competence' and 'performance', the former representing a person's achievement under test conditions, knowing that they are being challenged to demonstrate knowledge, attitudes and skills (and implicitly taken to be the best that they can achieve<sup>17</sup>), whereas

---

<sup>15</sup> Quoted by Redfern et al (2002, p.53).

<sup>16</sup> Some of the debate over competence has arisen because the protagonists have not made clear the level of analysis that they were applying in respect of the term. Hager et al (1994) offer a spirited defence of competence-based assessment.

<sup>17</sup> However, not everyone gives their best performance under the stress of formal testing.

the latter is what the person routinely achieves on a day-to-day basis. Put another way, performance on test is not a perfect predictor of performance in real life.

*How much competence?*

A critical issue in some areas (e.g. health-related professions; social work; education) is the need to assess with an eye to the implications of the assessment outcome for public safety. In these areas the consequences of passing a student who becomes a bad practitioner can be severe. The assessment regime needs to ‘play safe’ by minimising (preferably, but probably unrealistically, eliminating) ‘false positives’ – those who pass but in reality should not. Cowburn et al (2000) are at pains to point out that the assessment of students on social work programmes has to safeguard the public directly, and indirectly through maintaining professional standards, and state: ‘It is essential that courses do not avoid the difficult issue of failing inadequate students’ (p.630).

The cost of minimising false positives is, however, to increase the chances of failing students who might well become competent practitioners. This can be represented as in Table 2, in which the outcome of the institution’s assessment approach is set against the subsequent performance in practice.

	Performance in practice	
Assessment outcome	Not yet competent	Competent
Fail	Accurate prediction	False negative*
Pass	False positive	Accurate prediction

Table 2 A matrix relating assessment outcome to competence in practice.

\* If the student failed their assessment, then they would of course not be permitted to enter actual practice.

Newble et al (1994), acknowledging the public safety aspect of medical training, remark that the primary need is to differentiate reliably between competent and incompetent students, rather than to worry about finely grading the competent (which might serve the purpose of determining academic honours). They suggested a filtering approach to assessment, in which cheap and efficient tests (such as multiple-choice questions) might differentiate those who were well above threshold levels (and who could therefore move on without further ado) from those who might need a more extensive testing process to establish their achievement level. Whilst there is a logic to such a ‘decision-tree’ approach to assessment, the approach may not align well with the formalised structure of assessments embedded in validated curricula.

‘Competence’ is a term that gains meaning from the circumstances to which it is being applied. A beginning professional who is judged to be competent is adjudged competent

---

as an entrant to the profession, and is expected to develop their competence through further learning, in the workplace and through other educative experiences. Hence the concept of 'good enough' is relevant to determining who should become a beginning professional. Kemshall (1993) and Furness and Gilligan (2004) have pointed to uncertainty as to what 'good enough' actually means in practice, the latter pointing to difficulties in interpreting the National Occupational Standards that apply to social work. Holloway and Haggerty (2005), in working to develop the skills of assessors in nurse education in New Zealand, found reinforcement for their understanding that competence related to a safe standard of practice, and not to expert standard. Quinn (1995), however, had previously argued that safety was a necessary but insufficient condition for competence.

A particular problem may lie in the lack of specification of the criteria against which 'good enough' can be assessed. Stones (1994) argued that practical teaching lacked consensus as regards criteria against which performance could be judged. Redfern et al (2002) noted that workplace assessors were uncertain about their role and about what constituted an appropriate level of practice, and Murrell (1993) referred to the presence of biases in views of what constituted good practice. Reynolds (1992) had earlier been more sanguine about the existence of a broad consensus on what characterised competent beginning teaching, which can be summarised as planning; practising in the classroom; and reflecting in order to develop as a professional<sup>18</sup>.

The obverse of the genuinely 'good enough' beginning professional is the person who is assessed as 'competent' yet who is ill-equipped to undertake the further professional development needed to grow in the job (Wright and Bottery 1997). In a similar vein, Furness and Gilligan (2004) note that some students can pass all the course components without dispelling doubts about their fitness to practise as a beginning professional.

Eraut (2004b) suggests that one should conceive of competence in terms of the relatively short term (developed during the span of a first degree), the medium term (for example, basic training following graduation), and the long term (as in continuing professional development). This 'learning trajectory bundle' would be consistent with a lifelong learning perspective.

---

<sup>18</sup> Stones and Reynolds may have been approaching the issue at different levels of analysis.

## 6 Levels and criteria

Embedded in the notion of competence are two facets of assessment that are linked: the first being the level of the performance, and the second the criteria against which the performance is being assessed.

### *Levels*

A difficulty with the term ‘level’ is that it carries two different meanings which are not always clearly differentiated: level in terms of gradation of expectation (as used in qualification frameworks), and level in terms of actual performance at a particular level of expectation (e.g. good v. poor performance at Level 3)<sup>19</sup>. Burchell et al (1999) made an attempt to distinguish between the two meanings in respect of radiography, seeing a gradation between ‘basic’ competences in 1<sup>st</sup> year studies and greater complexity by the end of the programme – i.e. a level that enabled the student to be declared ‘competent to practise’. The *manner* in which the radiography practice was carried out (i.e. the exercise of professional attitudes and behaviour) distinguished the excellent from the (merely) satisfactory student.

The distinction between the two meanings can be depicted schematically in the manner of Table 3.

		Level of performance				
		Fail	Satisfactory	Good	Very good	Excellent
Level of expectation	Level 3					
	Level 2					
	Level 1					

Table 3 A schematisation of the two meanings of ‘level’.

Winter and Maisch (1996, p.91) described the use of levels as being ‘confused’. Canham (2002) asserted that there was no agreed strategy in nursing for describing performance at different academic levels.

Schutz et al (2004, p.52) note a lack of clarity regarding what constitutes reflection, and as to whether reflection can be said to have levels and, if it does, how these might develop over time. Later (p.67) they claim that the assessment of reflection is problematic for staff who are uncertain as to what they are assessing, and whose uncertainty extends to the relationship of grading to levels of reflection. In an earlier edition, Palmer et al (1994) had proposed a 3-level scale for reflective learning for demonstrating competence in reflection:

<sup>19</sup> See, for example, Winter (1994a, b; 2001).

1. mainly descriptive (though this hardly merits the label 'reflection')
2. awareness of personal values and beliefs; their relation to action; provision of explicit rationales for action
3. acknowledgement of wider influences (e.g. ethical; political).

There are hints in this scale of the kinds of personal developmental trajectory described by Kohlberg (1964), Perry (1970/1998), and King and Kitchener (1994).

Gerrish et al (1997, ch4) discuss the confusion inherent in levels, and the misapplication in the nursing field of taxonomies such as those of Bloom (1956) and Benner (1984). Ashworth et al (1999) note that the problem of defining the level of academic and practice-based achievements has been around for a long time. Their sorting exercise with lecturers, using statements derived from degree-level and diploma-level studies in nursing, showed that discrimination between the two levels of academic expectation was far from clear, leading them to state: 'We have begun to take the cynical view that [rationales for level descriptors] generally amount to no more than rhetoric' (p.166). They suggested three reasons why levels were problematic:

1. The language being used was insufficiently precise to allow differences in level to be identified
2. There was insufficient sharing of meanings to communicate reliably specifications for levels of attainment
3. The actual phenomena of nursing practice did not lend themselves to verbal categorisation in terms of discrete levels (e.g. interpersonal sensitivity).

Following Eraut (1994) they argued that academic knowledge ('knowing that') tended to be structured in lectures and textbooks, whereas the practical knowledge of the workplace ('knowing how') tended to lack such structuring. The HEQC's (1997) report on 'graduateness' showed the multidimensionality of the concept, but (probably very wisely) did not attempt to break down the various dimensions into levels of expectation. In practice, communities develop normative – and to some extent tacit – understandings of what the stated criteria mean at different levels but, as Wolf (1995) pointed out, abstract statements of understandings lack practical meaning until they are 'fleshed out' with examples.

Hence the ascription of expectation level would seem to present a considerable challenge – one that reaches well beyond the boundaries of nursing. The issue of level of expectation is treated with a surprising degree of insouciance, give its problematic nature.

Ashworth et al (1999) suggest that a way forward might be to start from what nurses actually do, and to decide whether there are real discriminatory levels of practice, instead of trying to specify levels on a 'top-down' basis. The danger is that this could turn out rather like the functional analysis that was used as the rationale in the development of NVQs (Jessup 1991), and hence be open to the risk of overspecification and practical unwieldiness.

Shumway and Harden (2003) approached the concept of level from the perspective of practice when drawing on a model rather casually introduced into medical education by Miller (1990) in which performance is characterised in terms of four levels (basic knowledge; knowledge of how to use such knowledge; demonstrating the ability to use the knowledge; and being able to practise as a professional). Avoiding the ‘complexity trap’, they set 12 general learning outcomes from three broad categories of a medical doctor’s performance (what s/he is able to do; how s/he approaches practice; how s/he maintains professionalism) against the four hierarchical levels, and suggested assessment methods that might be particularly appropriate (Figure 2).

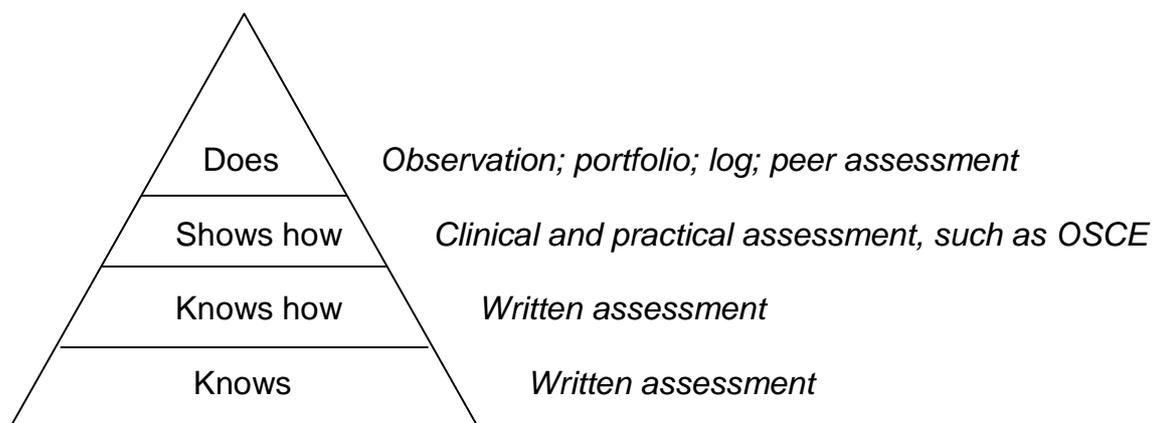


Figure 2 Miller’s (1990) hierarchy as presented by Shumway and Harden (2003). (OSCE refers to the objective structured clinical examination.)

What is important is that Figure 2 suggests that different assessment methods are appropriate for different kinds of performance, and – implicitly – that the assessment of professional competence requires a cluster of assessment methods whose outcomes may be expressed in a variety of ways (not necessarily in numerical terms, or even in gradings), and whose combination may not be straightforward.

### *Criteria*

Aviles (2001) put forward an argument for criterion-referenced assessment in the field of social work. Whilst the intention is understandable, its enactment presents some formidable challenges.

If the concept of ‘level’ is more problematic than some might perceive, the same can be said of assessment criteria. Sadler (1987) had long ago pointed to the inherent fuzziness of criteria expressed in linguistic terms, and Wolf (1995) subsequently showed that criteria were interpreted variably unless the criteria were supported by exemplifications that showed what they actually meant in practice. Stones (1994), writing of the assessment of teaching practice, pointed to the lack of underlying theory to provide a

rationale for the assessment criteria, and noted that the criteria that were actually used in respect of judgements about competence were ill-defined. On the latter point, Winter (1994a) pointed out the ambiguity and circularity of gradings whose practical meanings were embedded in phrases like ‘a high degree of...’, and ‘what would usually be expected of a student at this level’, or in words such as ‘outstanding’, ‘average’, and ‘satisfactory’. Blake and Laing (2000) illustrate some of the problems associated with assessment criteria in their analysis of the standards of professional practice adopted by the erstwhile Institute for Learning and Teaching in Higher Education and in the management of IT projects, as do Cope et al (2003) regarding the assessment of teaching practice. Hussey and Smith (2002, p.232) go so far as to suggest that the ‘alleged explicit clarity, precision and objectivity [of written learning outcomes] are largely spurious’, since they can only be understood with reference to a particular context.

O’Donovan et al (2004), who attempted to construct an ‘assessment grid’ in respect of business studies, found that there were problems in articulating clearly the expectations to be laid on students, and that the meanings were not necessarily understood by the students. Attempts to be more explicit led towards ever-increasing specification, its concomitant unwieldiness, and the problem that the wording used moved away from normal language. This fits with the doubt expressed by Hinett and Weeden (2000) that students understood the sub-texts of the assessment expectations.

## **7 Philosophical and methodological issues**

### *Models of educational assessment*

Hager and Butler (1996) drew, as others have broadly acknowledged, a distinction between ‘scientific measurement’ and ‘judgemental’ models of educational assessment, arguing that these were manifestations of two different epistemological approaches. The models are presented as Weberian ‘ideal types’ whose juxtaposition is intended to help reflection on some of the challenges posed by assessment. The models, whose essence is shown in Table 4, contrast an impersonal, theory-led approach with one in which context and human engagement are acknowledged as key influences. One is again reminded of the distinction made by Gibbons et al (1994) between ‘Mode 1’ and ‘Mode 2’ approaches to knowledge generation, the former being driven through separate disciplines; the latter by multidisciplinary endeavour. However, it should not be overlooked that a high proportion of assessments (even under the ‘scientific measurement’ model) involve at least some inference, which is tantamount to making a judgement regarding the student’s achievement.

There is perhaps a danger of seeing assessment in over-polarised, Manichaeian terms. The scientific measurement model serves some purposes well: students need, say, to understand the principles underlying the production of business accounts, to be able to detect errors in provided accounts, and to construct accounts on an approved basis. There is an essential corpus of knowledge and application that has to be acquired. But the preparation of accounts is set in a human context, and dealing with related management

issues moves the focus of attention towards an area where appropriate action is not necessarily susceptible to formulaic resolution. Indeed, the professional practitioner is often in the position of achieving the best outcome possible in the prevailing circumstances and not the best possible outcome in the abstract. As Cope et al (2003) observe in connection with the assessment of teaching practice, success, in situations of this kind, is more a matter of judgement than of measurement, and as Ashworth et al (1999, p.167) observe: ‘the very skills that are at the core of nursing are the most resistant to measurement’. The ‘scientific model’, with its technical rationality<sup>20</sup>, is of limited utility.

<b>Scientific measurement model</b>	<b>Judgemental model</b>
Practice derived from theory	Practice and theory (loosely) symbiotic
Knowledge is a ‘given’ for practical purposes	Knowledge is understood as provisional
Knowledge is ‘impersonal’ and context-free	Knowledge is a human construct and reflects context
Discipline-driven	Problem-driven
Deals with structured problems	Deals with unstructured problems

Table 4 Two contrasting models of educational assessment (after Hager and Butler 1996).

The assessment of PBPL can be undertaken by a variety of methods. No one method will deal adequately with the complexity of performance<sup>21</sup>, and so the logic is that a variety of methods needs to be used, and the assessment outcomes triangulated in order to develop a rounded picture of achievement. Further, the context determines what assessment methodology is most appropriate for the prevailing circumstances. Generic forms of competences are coloured by context-specific circumstances (e.g. Stasz et al 1996).

#### *Approaches to assessment*

Brennan and Little (1996) provided a useful tabulation of a number of methods which is elaborated in Table 5, and which contains implicit connections to each of Hager and Butler’s (1996) models. Some methods clearly have a stronger formative potential than do others.

<sup>20</sup> Richardson (1995) criticises the technical rationality embedded in Mezirow’s (1991) schema of reflective action, and notes that nursing is a social activity that is historically located and intrinsically political. The same can be said of other professions.

<sup>21</sup> In a study of rating in medical education Dannefer et al (2005) found that interpersonal attributes did not correlate highly with performance measures – these were essentially two different components of professional behaviour. The inference is that each needs its own approach to assessment.

Method	Used for...	Amplification	Comment
Direct observation	Assessing various aspects of professional competence	Reliability enhanced by use of checklists or well-designed rating scales	Need for observer to be trained and to have 'practice credibility'. Resource-intensive. Vulnerable to interpersonal factors such as the effect of the observer on the observed; reluctance to fail. Observation schedules may undervalue 'low-visibility' skills, and may be difficult to score.
Interview about practice.	Obtaining an appreciation of understanding for actions taken.		Reliability can be compromised by lack of structure, interpersonal factors.
Performance appraisal.	Developmental purposes (ideally).	Establishment of employee's targets for future.	Appraisee may have a different perspective than appraiser, and behave accordingly.
Opinions of others ('surrogate assessment').	Seeking a rounded picture of performance in the workplace.	Less resource-intensive than trying to achieve comprehensive observation.	Validity and reliability problematic. Evidence can be open to challenge as hearsay. Possibility in some situations of obtaining client reactions; also of peer-assessment.
Simulation.	Used in clinical work.	In clinical work, the use of 'simulated patients' (e.g. in OSCEs) allows practice in a 'safe' environment.	In some circumstances, the workplace has to be simulated (e.g. when a placement cannot be arranged).
Logbook, work diary or portfolio.	Development of reflective capacity.	If assessed, could attest to coverage of intended work experiences.	If assessed, could be over-concerned with 'presentation of the self'.
Preparation of report of PBPL.	Assessment of reflective and written communication capability.	Often used in sandwich courses.	May be a gloss on what really happened. Attention may be given to the report at the expense of the learning that has taken place, perhaps to the detriment of the latter <sup>22</sup> .
Written test or assignment.	Testing 'knowing that' and 'knowing how'	Widely used: more easily assimilated into academic award mechanisms than other methods.	Possibility of oral testing. Knowing is not the same as doing.

Table 5 Some assessment methods which might have a role in PBPL (adapted from Brennan and Little 1996).

<sup>22</sup> See, for example, Gammie and Hornby (1994).

Norcini (2005), writing about performance assessment in the medical profession, adds to the methods featured in Table 5 the use of clinical practice records and administrative data in record systems. Whilst these offer the possibility of accessing extensive data, the costs and time that analysis would involve are likely to be formidable.

Gleeson (1997) argued that the 'long case' (in which medical students take a patient's history, conduct an examination, and are subsequently examined by a pair of assessors) gave rise to unreliable marks, and that the unreliability was not greatly improved by training the assessors. Whilst the Objective Structured Case Examination [OSCE] was appropriate for assessing component parts of practice, it did not address the totality of the trainee practitioner's performance. The Objective Structured Long Examination Record [OSLER], developed as a result, was graded separately by two examiners with reference to ten broad criteria, using three performance bands (P+ = very good/excellent; P = pass/borderline pass; P- = below pass) and with the difficulty of the case being taken into account.

Ottewill and Wall (1996) argued that, whilst traditional methods of examining were inappropriate for the assessment of sandwich placements, the (then) newer competency-based methods were controversial, difficult to apply and extremely time-consuming.

There is always a trade-off to be made between desirable and undesirable characteristics of assessment methods (Norman et al 1991), and a need to appreciate the limitations inherent in any assessment method. Even when competences are specified quite tightly, as are those of the Australian Nursing Council Incorporated [ANCI], the latitude for interpretation has attracted comment<sup>23</sup>, and the potential variability in interpretation has been increased by the course accreditation processes adopted in individual States and Territories (Heath 2002, para 7.1.1). Against this, however, Walkington (2004) indicated that feedback from mentoring teachers regarding the competency-based approach to teacher education used in the Australian Capital Territory was too prescriptive and did not allow contextual considerations to be included.

### *Multiple assessments*

There is a widespread appreciation in the literature that the assessment of performance in the workplace requires multiple approaches to assessment if a rounded picture is to be obtained<sup>24</sup>. In a way, there is something of a parallel with the so-called 360° approach to appraisal, where the intention is to build up a picture from data collated from a range of

---

<sup>23</sup> It may be no coincidence that Australian graduates reported a low level of satisfaction with 'assessment' in university nursing courses. For the years 1996-2000, the average 'broad satisfaction' score for 'appropriate assessment' in the Course Experience Questionnaire was between 33 and 39. This compares with a broad satisfaction ratings of 84 and 85 from the general university student population for the same years. While this single aspect of student experience was rated poorly by nursing graduates, their broad satisfaction scores in respect of other CEQ scales ranged from 83 to 91 for the same period (Heath 2002, para 7.1.4).

<sup>24</sup> See, for example, Shardlow and Doel (1993); Schwartz et al (1997); Norman et al (2000); Schuwirth et al (2002); Knight and Yorke (2003).

sources, including clients<sup>25</sup>. Critical issues here are what should be included in the assessment, and how the various contributions should be weighted<sup>26</sup>.

The assessment of performance in a work environment ideally requires

- sufficiently large samples of practice across a range of relevant situations
- sufficient variety in assessment method, blending the apparently objective with the subjective (i.e. drawing appropriately from both of the assessment models outlined by Hager and Butler, 1996)
- a focus on outcomes (bearing in mind that some outcomes may relate to processes from which final outcomes emerge).

The ideal is unlikely to be fully realised in practice, and hence there is a need to identify the optimal assessment regime for the circumstances. This means, *inter alia*, determining those aspects of performance that are vitally important, those where some sort of judgement is desirable, and those which are of lesser importance or which can be assessed as by-products of the more important aspects. However, the expectations of some professional bodies may militate against the subsumption of some aspects of performance within others.

Examples of multiple assessment (there are a myriad) include the following.

1. Hager and Beckett (1995) describe an assessment approach used by Law Society of New South Wales in which the student is videotaped whilst undertaking an interview with a person adopting the role of client<sup>27</sup>: the videotape is assessed by examiners. Other parts of the assessment include dealing with tasks in a mock file; referees' reports and examinations testing legal knowledge.
2. Wood (1997), writing about the use of undergraduate students as tutors in schools, described assessment in terms of three components: a formative logbook including a personal action plan; a summative formal written report; and a summative oral presentation.
3. Rickard (2002), writing about a short (6-week) work-based learning placement in the field of health that was interposed between two blocks of formal teaching, describes assessment in terms of a short reflective piece on the placement, accompanied by a portfolio; and a longer critical discussion of health issues in East London, drawing on theory and the placement.
4. Dennis (2003) describes the 'Client Consultancy Project' in which final year business studies students are required, in groups, to find and carry out a consultancy project during two semesters, which is then peer-assessed, as is a presentation related to this work. Dennis presents some evidence that students

---

<sup>25</sup> Edwards (2003) pointed to the involvement of service users in assessment, and some of the ethical issues that were thereby raised.

<sup>26</sup> Shardlow and Doel (1993, p.75ff) made an initial foray into this territory.

<sup>27</sup> This is similar in some respects to the taking of a medical history. The videorecording of practice is unlikely to be feasible save where the professional is practising in a single location, such as the doctors' surgeries covered in a study by Ram et al (1999).

- developed their capability for self- and group assessment, and that this was of value in the process of job-seeking.
5. Tapper (2004) describes a module for third year students in the humanities at the University of Melbourne which is designed to help them come to terms with workplace expectations. The module is based on workplace projects. The assessment regime consists of four assignments: a collaborative project brief, an oral presentation, a business plan and a final evaluative report. Host employers grade students on a 5-point scale, and the grades contribute to the final subject grade.
  6. Teacher education students in Washington State in the US have, since Fall 2004, been assessed according to the provisions of Performance-based Pedagogic Assessment, in which they are observed in practice and also are required to submit evidence of pupil learning (see Bergeson 2004 for a description and assessment instrumentation).
  7. McCulloch (2005) describes Kajulu Communications, a student-operated agency in which students of advertising can spend their final year in what amounts to an 'internship on campus'. Kajulu is a standalone agency on the campus which fulfils commissions from external organisations. Assessment of students' performance consists of a combination of group, peer and client assessments.

### *Context*

Assessment methods have to take into account the practice-based learning situation and the extent to which the institution responsible for the assessment is involved in determining its nature. Toohey (1996) offered a typology of the practicum (i.e. the placement) ranging from the fairly *laissez-faire* to a negotiated experience (Table 6).

Griffiths and Guile (2004, p.19ff) provide a typology of work-based experience which shares some features with the categorisation described by Toohey et al. They do note that the relationship between actual work experience and their five models is blurred.

Duignan (2003) offered a dichotomy which seems to cover only the first, fourth and fifth of the types described by Toohey et al:

- Work environment model, in which the educational institution arranges the placement, prepares the student, and then lets go.
- Learning environment model, in which the placement is designed with educational learning objectives in mind, and the educational institution plays a role in the placement experience, following this with activities intended to embed the benefits<sup>28</sup>.

Cross-cutting the variations that are possible in placements is the nature of the workplace environment. Teacher educators, for example, appreciate well that the achievement of a

---

<sup>28</sup> The empirical study reported by Duignan in the same paper is not helped by his choice of the mark for the placement module as the criterion for comparisons between the gains of students on placement versus those did not undertake a placement.

good performance in a genteel school can be a much less challenging task than in a more 'difficult' school. An issue that may need further attention is how the nature of the workplace can be 'factored into' the assessment process. In a more extreme form, the issue is the extent to which it is actually possible to learn well (and hence perform well) when the placement is either extremely demanding on a day-to-day basis, or simply not demanding enough.

Model	Key features	Comment
Attendance	Attend placement, complete satisfactorily, attested by supervisor, graded pass/fail or ungraded	Valued for socialisation, but minimalist as far as assessment goes
Work history	Student documents, and perhaps reflects on, activities. Assessment certifies task completion	Unclear how much influence HEI has
Broad abilities	Broad capabilities assessed alongside specific technical skills, aiming to integrate academic and practical	Some prefer pass/not yet pass; others grading
Specific competences	Key roles and tasks specified; placement organised accordingly	Assessment of sample of performances, or full range; may or may not be graded. Knowledge component can be assessed by exam. Competences may be too narrowly defined, and limit development of handling complex problems. Would seem to 'work' only with a large employer.
Negotiated curriculum	Learning contract negotiated between HEI, workplace and student	Assessment focuses on fulfilment of contract. Setting up takes time, hence again likely to apply to larger employers where economies of scale are possible.

Table 6 The typology of the practicum suggested by Toohey et al (1996).

## 8 Portfolios for summative assessment

Given the importance attached to professionals' development of their capacities to operate with a high degree of autonomy, learning activities that focus their attention on their own learning (i.e. their metacognitive development) are important in PBPL. Earlier developments were concerned with the development of 'profiles' of learning that took place, often in sandwich-type placements: Assiter (1995, pp.53-131) contains a number of examples of profiling in different disciplinary areas.

More recently, the emphasis has shifted towards the use of portfolios of experience. Portfolios are generally seen as vehicles for student learning and as repositories of evidence on which students or graduates can draw for various purposes. Rather less is

---

made of the possibility of assessing them, for reasons such as their confidentiality to the student and the distorting effect on their compiling if they are to be subject to assessment (the latter point is acknowledged by writers such as Schuwirth et al, 2002). Wilkinson et al (2002) seem not to appreciate the tension between using portfolios to contribute to the improvement of the quality of medical practice and as a method that can contribute to the attainment of minimum acceptable standards. They do acknowledge the need for collateral validation through evidence and the reports of referees, but this does not eliminate the basic tension at the heart of using a portfolio as an assessment method.

The use of portfolios for summative assessment purposes was criticised by Gerrish et al (1997) and McCullan et al (2003). The criticisms of Gerrish et al included the following:

- portfolio construction was time-consuming, and distracted attention from actual practice
- the production of a portfolio involved writing about practice rather than actually practising
- the compiler of a portfolio needed to possess writing ability in addition to the capacity to reflect upon practice
- there was a risk that the compiler would write what they thought that the assessor wanted to read rather than record experience spontaneously.

Taylor et al (1999) indicated that portfolios in respect of social work (via the SAPHE Project) contained

- student evidence
- practice teacher evidence
- evidence provided by others.

They noted the dilemma for students of being both a critically reflective practitioner and a competency-driven social worker, and raised the question of what a student should include and exclude.

Harland (2005) noted that there were three key issues relating to the use of portfolios: who owns them; who reads them; and how they should be assessed. Whilst acknowledging that assessment of their portfolios would distort student behaviour, they found that all of their students allowed staff to read the whole portfolio, despite being given the option to retain items as private. This did not accord with tutors' desire that students should use the portfolio as a vehicle for private reflection, so they then asked students to provide summaries of their learning; later still, even the summaries were declared private documents. Harland found that students took time to acclimatise to the requirements of portfolios; not all engaged as was hoped, some portfolios – despite the expectation that they could be private – being written more for presentation of the self than as reflective documents. Coleman et al (2002) touch on assessment via portfolio and the associated tension between using the portfolio for personal growth and for assessment purposes. Regarding the latter, they also raise the question of the writer's honesty if the portfolio were used for assessment purposes.

---

Driessen et al (2005) studied the assessment of 233 portfolios from the perspective of economy of effort. A 'decision-tree' approach was adopted, in which failure of those involved to agree meant that the portfolio was referred for resolution to another judgemental situation, the final 'loop' in the process involving review by a committee of assessors. First, the mentor and student attempted to agree whether the portfolio passed (the vast majority did agree). Next, a separate rater read the portfolio and, if s/he agreed with the previous 'pass' decision, that became the final decision. Only where there was disagreement did the process 'loop' through other judgmental processes. Nine cases ended up for committee scrutiny: in these, the mentor was allowed to put an argument, but was not involved in the judgment. Most of the mentors' time was spent on providing formative feedback, rather than on summative assessment.

Inter-rater reliability in respect of portfolios is problematic (Pitts et al, 1999; Baume and Yorke, 2002) even when the portfolio is quite tightly structured. It is doubtful whether inter-rater reliabilities will reach those achievable in more 'objective' assessment methods. The use of multiple raters of portfolios is unlikely for reasons of practicability (Driessen et al 2005).

In an addendum to Canham's (2002) chapter, Joanne Bennett indicated that students at the University of Northumbria needed some guidance regarding the compilation of their portfolios, the initial idea of giving the students flexibility to compile portfolios in their own way having (perhaps not surprisingly) proved untenable. Compiling a portfolio is a sophisticated exercise, especially when it is expected to include reflective commentary, and the novice compiler is likely to need some advice regarding structure and content. One has to be careful not to over-standardise with the intention of enhancing reliability of assessment, since standardisation brings with it threats to richness and validity.

Johnston (2004) and Tigelaar et al (2005) argue that traditional approaches to the assessment of portfolios are inappropriate to unique productions derived from particular contexts, and propose instead an interpretive, hermeneutic approach. Both point towards the kind of constructivist, qualitative methodology associated with Guba and Lincoln (1989). This might be an attractive and fruitful way of approaching the assessment of portfolios, but of course assumes that the portfolio is produced for assessment purposes (and hence is not a document private to the student), with all the inherent threats to its validity. Although these authors touch on the 'cut' between pass and fail, an issue that needs further attention is the extent to which assessment via portfolio can address the issues of public safety that were noted earlier in this Report. A further issue, not addressed by Johnston, is the elasticity of Guba and Lincoln's (1989) 'credibility' and 'transferability' when (almost inevitably) there are discrepancies between the multiple sources of evidence.

The focus in the literature on the use of the portfolio for formal assessment purposes diverts attention away from the value a portfolio may have in terms of self-development and self-assessment. Taken in conjunction with personal development planning, the portfolio offers the opportunity for the student to self-assess and to reflect on progression in their learning and professionalism. This ipsative approach could contribute to a formal

assessment process, but would tend to draw the portfolio away from direct assessment by others.

In post-16 education, in contrast, the complexity of issues relating to portfolios seems not to arise. Portfolios are widely interpreted as summative documents, to be simply 'signed off' for the purposes of gaining the qualification (Torrance and Coultas 2004).

## 9 Self-assessment

The preceding section touched on the potential of the portfolio for self-assessment, and the expectation that, through the compilation of a portfolio and reflection on the learning that has taken place, the student will be better equipped to assess their own progress and to self-regulate in respect of expected professional standards. One of the trajectories expected – sometimes implicitly – of students is that they should shift away from dependence on a teacher for judgements about their achievements to the autonomy to undertake a considerable amount of self-assessment. In a nutshell, this is the kind of shift from acquiescence to autonomy envisaged by writers like Kohlberg (1964), Perry (1970/1998) and King and Kitchener (1994).

It is generally understood that success usually depends on more than academic capability: Sternberg (1997) argues convincingly for the need for 'practical intelligence' as well. One who seems to lack the latter is the graduate who solicited advice from the column penned by Jeremy Bullmore (2005):

I am at my wits' end. I'm a recent business studies graduate with a first class degree, but I've been fired from my third job in a year. It's always the same pattern: I ace the interviews, get a position and never make it past the three-month probation period. My bosses tell me that I have the skills, but not the attitude. I don't understand this – I get to work on time, put in the hours, am charming to everyone and am a model worker. How can I make them see I'm their dream candidate?

The response is what any reader with semi-functioning antennae might make, and is summed up in an excerpt. Inviting the writer to imagine that he was reading the letter as for the first time, and from a third party, Bullmore wrote:

Bosses don't hire and fire just for the fun of it. Yet, not for one second does this person entertain the possibility that any part of the blame might lie with him. Don't you deduce that this is someone whose confidence level has risen from the enviable to the intolerable? Who carries such a load of self-certainty that the concept of personal inadequacy is never remotely entertained.

The point of this anecdote is that much of what counts as success outside the higher education environment is difficult – if not impossible – to 'measure' with the instruments that higher education usually employs. Business Studies programmes typically incorporate one or more periods of work placement, yet in this particular case the student seems not to have been alerted to the danger of feeling that he (Bullmore assumes that the person is male) is God's gift to any enterprise. There could have been a variety of reasons why the person's arrogance was not picked up: the supervision may have been

inadequate; the assessment of the workplace performance may have been deficient; the student may have only been required to submit a report on his workplace experience (and it is evident from the anecdote that self-presentation is unlikely to cause the presenter any difficulty); and the weighting given to (presumed) workplace performance in the degree classification process may have been minimal.

### *Unheard voices?*

As Brackenreg (2004) notes in relation to nursing, there is very little in the literature on students' perceptions of assessment on work placements. Yet students' perceptions of assessment, and of how the assessment process influences their behaviour, are potentially significant contributions to a broader understanding of assessment in PBPL. They are also of significance in respect of students' and practitioners' self-assessment: to what extent does the assessment regime encourage the metacognitive activities of reflection and self-regulation? There are rare examples in the literature. Adams (2003) provides an example (albeit not particularly reflective) of a student's experience of the OSCE, and Hillier (1999) gives a distillation of 16 students' experiences on higher level NVQs.

## **10 Assessing more than the individual**

Of greater importance to professionals in practice, rather than to individuals on placements, is the functioning of the teams and systems in which they are situated. Farmer et al (2002) suggested screening the medical team at what they termed 'low power' in order to establish whether it satisfied threshold expectations regarding processes and outcomes. If the team 'passed' this scrutiny, it would be presumed to be operating at an acceptable level. If it did not pass, then further scrutiny would be needed to establish whether the problem lay with the doctor(s), the team or the system. Whilst there is a logic to looking at levels above that of the individual professional, it is possible for a team as a whole to work at an acceptable level when individuals do not.

The assessment of teams and systems, however, leads discussion towards quality systems and assurance, which lie outside the scope of the present Report.

## **11 Some technical issues relating to assessment**

Assessment methods have to respond to a number of considerations whose salience varies with the purpose. Hays et al (2002) suggest the need to consider various aspects of the medical practitioner's role (Figure 3), pointing out that assessment in respect of practising professionals<sup>29</sup> is to some extent different from that for those intending to enter the profession.

---

<sup>29</sup> This also varies with medical specialism.

They also provide a tabulation in which six characteristics of assessment methods

- sensitivity (in the case of doctors, high technical sensitivity is needed to minimise ‘false positive’ passes)
- reliability
- validity
- educational impact
- cost-effectiveness
- acceptability

are rated against a variety of assessment purposes, such as screening; improvement of performance and validation by the external world (ibid, p.913).

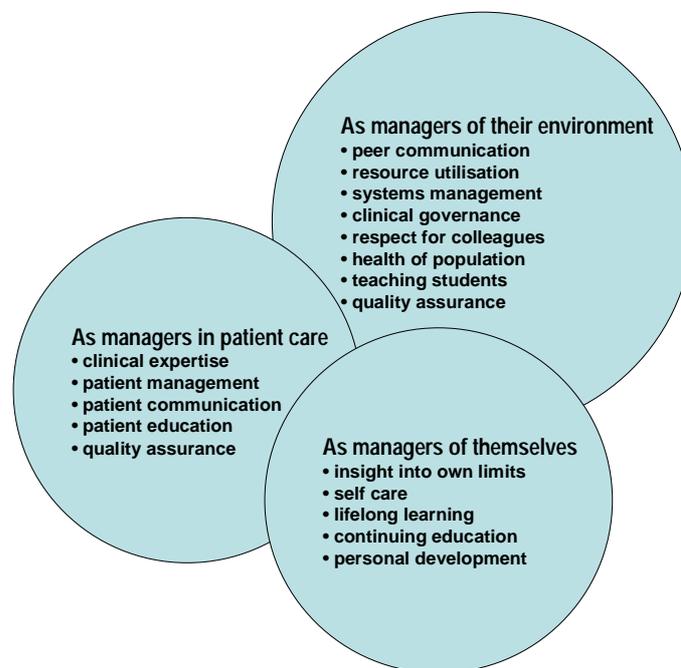


Figure 3 Aspects of the professional medical practitioner's role. (Hays et al 2002, p.912).

Assuming that the assessment methods chosen are valid for the intended purposes (in test language, they actually test what they are intended to test), the issue of reliability has to be addressed. The discussion above of ‘false positives’ regarding competence sharpens the point. The optimisation of reliability is discussed in Knight and Yorke (2003, ch 8), with particular attention being given to two themes that are relevant to PBPL – the assessment of work-based learning and of portfolios.

‘Objectification’ – rendering the assessment process less susceptible to ‘measurement error’ – is not necessarily followed by an improvement in reliability. There is a common misperception that subjective, judgemental methods of assessment are unreliable, and that

'objective', 'scientific' methods are reliable (Schuwirth et al 2002). Van der Vleuten et al (1991) showed that scores from subjective assessment methods can be at least as reliable as those derived from 'objectified' methods. In other words, subjective judgements should not be rejected simply because they are subjective: however, a corollary is that subjective judgements need to be supported by justification (a point that applies equally to more 'objectified' assessments, even though these are less likely to be subject to challenge). Guba and Lincoln (1989) use a constructivist language to address this kind of issue, referring to the 'dependability' and 'confirmability' of an evaluation.

Ladyshevsky (1999) found that the reliability of ratings varied with the focus of interest. The reliability of ratings of medical students' performance for established procedures ranged from 0.68 to 0.93, whereas for 'patient education' and skill in communication the range was lower, at 0.50 to 0.77.

Martin and Jolly (2002) showed that the OSCE in the first year of clinical studies predicted well further achievements in clinical work later in the study programme. Of interest, perhaps, for some assessment situations is their use of 'receiver operating characteristic' [ROC] curves<sup>30</sup> to provide cut-off points on assessment scores that best discriminate the competent aspiring practitioner from the not yet competent.

Norcini (2005) notes some threats to the validity and reliability of assessment of practitioners which are applicable to other professional contexts.

- Some doctors treat patients with particularly poor prognoses; some with better prognoses. Patients vary in the extent to which other factors outside the doctor's specialism may affect their life-chances. More broadly, the 'mix' of professional and client group may have a bearing on the professional's performance.
- Where a professional forms part of a team, how can an individual's contribution be singled out in the assessment process? One might have a successful doctor in a less successful team, or vice versa. This is an issue that has been acknowledged more broadly in the assessment of group work in higher education, but to which no unequivocal 'best buy' resolution has been found.

One might dispute Norcini's claim that 'infrequent but important patient problems are not amenable to assessment nor are areas of medicine where change is occurring in the nature of diagnosis or treatment' (ibid, p.887) – unless the word 'reliable' (or 'dependable') were placed before 'assessment'. Performance with respect to infrequent events can be assessed validly: it is the *reliability* or *dependability* of the assessment that is the difficulty.

The more one is able to practise in a profession, the greater the chances of successful outcomes – another way of expressing the adage that practice makes perfect. Halm et al (2002) reviewed hospital-based studies covering a 20-year period, finding that around seven in ten showed a statistically significant positive association between the volume of

---

<sup>30</sup> A useful introduction to ROC curves can be found at <http://gim.unmc.edu/dxtests/ROC1.htm> et seq.

---

work and outcomes. The ‘message’ for the assessor of workplace performance is that some account may need to be taken of the number of opportunities the assessee has had to demonstrate the acquisition of the capabilities at the focus of interest.

### *The rating of performance*

There is an optimum number of subdivisions when judging performance, though this optimum probably varies with circumstances. Too many categories for assessment, and assessment becomes unwieldy (the case of NVQs looms large here)<sup>31</sup>; too few, and the assessment outcome becomes relatively uninformative (honours degree classifications on their own, for example, tell nothing of a graduate’s pattern of strengths and weaknesses). Stones (1994, p.239) writing of teacher education (but with a potentially broader reach), says: ‘The flawed nature of assessment should be recognized and the ... common practice of awarding finely graded assessments should be abandoned’.

Canham (2002, p.97) describes an approach to assessment at Manchester Metropolitan University in which nursing practice attracts 20 credits at Level 3:

A percentage mark is applied to the student’s performance as it stands during the very last week of the course. [...] Applying a precise mark to practice enables all parties to identify at what point the student is functioning before being allowed to re-enter the world as a specialist practitioner.

Note the weight apparently being carried by the mark here. After referring to nine overarching practice assessment criteria, she remarks:

Although the skills, knowledge and attributes are the same for each student, the exact content and the way that they will be applied in practice will vary dependent on both the student and the practice.

– which rather undermines the claim at the end of the previous excerpt.

An ‘inverted U’ curve schematically relates the number of categories of performance used in assessment to their practical utility-value (Figure 4).

---

<sup>31</sup> Cooper (2000) calculated that there were 8008 indices of evidence relating to the judgement of competence as a child care social worker, which he described as ‘formidably dispiriting’ (p.114).

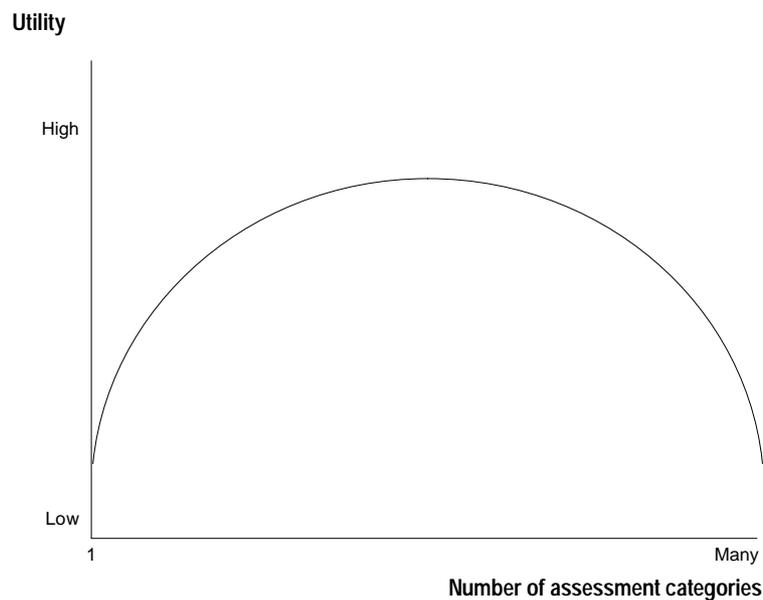


Figure 4 The number of assessment categories related to their practical utility.

One of the arguments for a larger number of assessment categories is the identification of aspects of performance on which the student must pass, and hence the minimising of the possibility of a failure on such aspects being buried by some form of compensation mechanism. The problem is implicit in a rating tool devised by Donoghue and Pelletier (1991) for the assessment of nurses' clinical practice. Specified activities were rated on four dimensions dealing with knowledge and safety; psychomotor performance; level of independence in action; and interpersonal performance<sup>32</sup>. The expansion of scale points reveals plenty of scope for compensation: for example, one scale point descriptor is

Occasionally communicates effectively by: active listening; assertiveness if appropriate; congruent nonverbal behaviour; empathy; therapeutic touch (p.357).

Similarly insecure anchor points can be found in Young's (1999, pp.129-30) scales for the assessment of portfolios for Level 2 in health sciences at Thames Valley University.

The complexity inherent in workplace performance makes the psychometric validation of assessment instruments highly problematic, and it is no surprise to find Redfern et al (2002) remarking that few competence methods in nursing have been adequately validated for their psychometric properties.

#### *Combining assessment outcomes*

Foulkes et al (1994, pp.139-40) remark that measurements may be combined only if they are from identical levels of measurement and if they are expressed in identical units of

<sup>32</sup> The actual aspects of assessment seem to be rather loosely specified, given the descriptions in the article, which raises questions about validity.

measurement. Their assertion seems limited to interval or ratio measures, since one cannot validly combine ordinal or categorical measures, save by some form of profile. It does not clearly cater for differential variances between measures, which can have a marked effect on overall rankings (though the point is later acknowledged on p.144)<sup>33</sup>. Further, it fails to make explicit that some items being measured may be much more important than others.

### *Representativeness*

There is a further issue that has to be noted regarding rating: that of the extent to which the observed and assessed behaviour of the person is representative. Alexander (1996) noted that students' performance could be misjudged because of presumptions of the likelihood of a behaviour occurring and its presumed representativeness. The 'sampling problem' could be compounded by the effects of pressure on workplace assessors and difficulties that such assessors may have in setting the performances of their assessees against some normative standards<sup>34</sup> (Redfern et al 2002). The tension between the roles of mentor and assessor (noted earlier) may also influence rating.

### *Prioritising*

The literature is often insufficiently clear about the inherent prioritising of aspects of performance. Yet students must pass on some aspects of performance (or else fail their programme) whereas weaknesses in other aspects can simply trigger some form of remedial or compensatory activity (i.e. they are not necessarily fatal to the student's progress). A failure to respect clients in social work, for example, is a more serious matter than is a weakness in written communication. An example from a different field of practical experience demonstrates the issue. In setting up procedures for allocating FDTL project funding, a set of criteria were provided by HEFCE against which bids were to be judged. However, no guidance was given as to how the criteria were to be used. Most judges scored bids against each criterion and gave the total: others took the view that, if the idea behind the bid was of little merit, this could not be redeemed by well-structured procedures for dissemination, management and evaluation. There was considerable debate as to the merits of proposals for which the ratings were discrepant in this way.

### *Defensibility*

As well as the 'public interest' aspect of the need for assessments to withstand scrutiny and possible challenge, the interests of the assessee likewise have to be taken into account. The word 'defensible' and its derivatives have become noticeable in the recent literature, reflecting concerns about both the assurance of professional standards and potential challenges mounted by the assessee. Needham (1988), though writing with

---

<sup>33</sup> Analyses of students' academic records (Yorke et al 2005) and of institutional 'league tables' (Yorke and Longden 2005) demonstrate the point.

<sup>34</sup> Redfern et al (2002) note that practice-based assessors may be more confident about assessing some aspects of performance than others. The same might be true of academics.

respect to the then emergent system of NVQs, remarked that the assessment of competence often left no solid proof of a candidate's performance that could be seen by a moderator. Where licences to practise are at stake, as in medicine, the importance of having defensible assessments is obvious (see, for example, Lew et al, 2002, Driessen et al 2005).

### *Authenticity*

'Authentic assessment' (more accurately, the assessment of authentic performance) has particular attraction in respect of PBPL. However, assessing an authentic performance may lack validity if the purpose of the assessment is not aligned with assessment practice. For example, the apparently 'more authentic' patient management problem may be less valid for problem-solving than apparently 'less authentic' methods such as extended matching questions or 'key feature problems' (Schuwirth and van der Vleuten 2004).

### *Pressures on assessors*

Validity in assessment is compromised wherever experienced assessors find themselves unable to give the assessment process the attention it requires. In some cases, the logistical effects of academic staff travelling to placements act against the interests of assessment (e.g. Aston et al 2000). In others, the demands being made on the senior officers' time are the problem: Lang and Woolston (2005) found this to be a real difficulty in a study of the training of probationary police constables in New South Wales.

More broadly, the involvement of workplace personnel in the assessment of placement students (assuming their willingness, which is not always the case) raises questions of validity and reliability where they have not been trained for the role (Young 2004). The same applies to academic staff who may be assessing in arenas for which they lack training or experience: doubts have been expressed about the ability of some academic staff to make summative judgements in competency-based education (Jones 1999) and regarding performance on placement in professional practices (Goldenberg and Waddell 1990; Redfern et al 2002).

## **12 Practitioners as assessors**

Resource considerations lead to the assessment of workplace performance often being largely a matter for practitioners<sup>35</sup>. They observe students working throughout a period of placement and hence are in a position to judge how well they are measuring up to the demands made upon them. The tutor from the institution is likely to visit the student occasionally to discuss progress, and to meet the supervising practitioner. In general, the tutor is not in a position to make an assessment of the student, but the engagement with

---

<sup>35</sup> Brackenreg (2004) noted the pressures on university teachers which made it difficult for them to engage in students' workplace learning.

the placement may enable the tutor to understand what is said in a report from the placement.

The validity and reliability of observation-based assessment depend on both the capability of the assessor and on the assessment instrument(s) being used. Many practitioners lack the training – in some instances the qualifications – necessary for assessment at the intended level. For example, research conducted by Culley and Genders (1999) found that five out of six Community Learning Disability Nurses doubted their ability to educate parents, which would suggest, by extension, that they would have doubted their capacity to assess students at degree level. Furness and Gilligan (2004) remark on the shortage of people with the Practice Teacher's Award, attributing this to a range of pressures on practitioners in the workplace. They note that the pressures may be greater for some staff (e.g. black staff), where they are particularly loaded with responsibilities. Brackenreg (2004) notes that, in Australia, summative judgement is often *de facto* delegated to clinical nurses, without moderation, thus placing considerable trust in the assessment process.

It is well understood that untrained observers may produce assessments whose validity can be questioned (perhaps because they focus on some aspects of performance and not others), and that they may be inconsistent from student to student (perhaps because of differences in the assessor/student relationship). The obvious response is to use a checklist and to train assessors in its use. However, checklists may fragment the desired performance into components which, whilst all being valid as components of the performance, nevertheless fail to capture the complexity that inheres in professional practice. The reductionist approach favoured for the introduction of NVQs based on functional analysis (see Jessup 1991) showed that, at Level 5, managerial competences were inherently complex (and hence implicitly that the criteria for judgement contained inherent ambiguities): only at the lowest level were competences and criteria of limited ambiguity.

#### *The double role: mentor and assessor*

The tension between formative and summative assessment in PBPL is particularly apparent in connection with mentorship, where the developmental mentor can become the summative judge at the end of the placement. In some situations, the duality of role seems to be understood by students as a part of the normal educational package, as in teacher education<sup>36</sup>. When the mentor and mentee are employees of the organisation, and have to remain in a working relationship, the role-relationship could have long-term consequences. Where a greater emphasis is being placed on practice-based learning, as in teacher education in England and Australia, and as in social work, the significance of mentoring is increased.

---

<sup>36</sup> Jones (2001) notes that, although 22 out of 25 mentors in state comprehensive schools saw assessment (presumably summative) as part of their role, only one out of 25 mentors in German hauptschulen took this view. Twenty-four of the 25 in each sample, however, acknowledged the role of adviser.

Mentoring is predicated on the capacity of the mentor to draw on expertise in supporting the mentee's learning. In some instances, the mentors' expertise has been called into question. Gerrish et al (1997) appear, at the time of their study, to have had some reservations regarding the expertise of practice assessors in nursing. Wellard et al (2000) surveyed 30 Australian universities and found that almost half of their clinical supervisors were not above bachelor's degree level, which, together with the limited preparation that they had had for their role, raised concerns about the quality and effectiveness of the students' clinical education experience<sup>37</sup>.

The literature bears continuing witness to a sense of unease by those involved on the provider side of mentoring<sup>38</sup>. The main concerns expressed are:

- the tension between nurturing and judgemental assessment
- the risk to trust of the (summative) assessor's power.

In its report *Preparation of teachers, practitioner teachers, mentors and supervisors in the context of Project 2000*, the ENB (1989) separated the role of assessor from that of mentor, though the subsequent literature suggests that this has not always happened in practice.

Where the mentor has a judgemental role, the mentee may be reluctant to ask for help or admit mistakes – hence prejudicing learning. Eraut et al (1995) noted that all of the nursing and midwifery students in their study had felt foolish on occasion because they had not known simple things, and that this had undermined their confidence. The students were afraid to ask for help in case they were labelled as dim or troublesome.

The mentor/assessor is (in theory) in a much better position to comment on the mentee's actual performance than an occasional visiting assessor from an educational institution. However, set against this is the socialisation implicit in the mentor/mentee relationship, with its potential for biasing summative assessment (see, for example, Jones 2001; Watson et al 2002). Further, there is a risk that untrained assessors may over-rate students' performances due to their delight at receiving the students' help (Hounsell et al 1996, p.69).

Where a student succeeds on placement, the inherent role ambiguity may not be a problem. In a study of students on foundation degree programmes, Yorke found that 13 of 87 who had had a mentor said that the mentor 'doubled up' as an assessor: in only one instance had this caused a problem. (The respondents indicated that they had had a much greater level of concern about the extent to which there was co-ordination between the HEI and the workplace regarding the placement.)

---

<sup>37</sup> See earlier comment on the extent to which practitioners in the health arena might lack expertise in respect of the assessor's role.

<sup>38</sup> See, for example, Andrews and Wallis (1999); Heilbronn et al (2002); Redfern et al (2002); Bennetts (2003); Brackenreg (2004); Walkington (2004). Surprisingly, however, Pawson (2004) made no mention of the role-tension in his review of mentoring.

---

*Generosity in assessment*

Giving students the benefit of the doubt is a persistent theme in the literature<sup>39</sup>. This is done for a variety of understandable and perhaps sometimes interlinked reasons:

- a desire to encourage students' growth by awarding a pass
- a nurturing rather than a judgmental academic climate, in part, perhaps, reflecting that students often perceive grades as indices of their personal worth
- providing students with a second chance
- affirmative action in respect of disadvantaged groups
- a culture of strong support for colleagues<sup>40</sup>
- avoidance of the hassle associated with the failing of students
- to reduce the possibility of litigation.

Where there is more than one placement, giving the benefit of the doubt tends to arise at first placement, where the assessor has to decide between an early fail and leaving the matter to be resolved at a subsequent placement. In a caring profession, where support for development is strongly embedded in the pedagogic approach, it is particularly difficult for an assessor not to 'give the student a second chance'. The consequence is, though, that the student can reach a similar point at the next placement with the weaknesses unresolved, where the decision to fail becomes even more difficult because the student has already completed a substantial part of the programme.

There is evidence that assessors may 'bend the rules' of the assessment process in order to produce the more favourable outcome that they believe is justified (Hawe 2003; Baume et al 2004).

Brandon and Davies (1979) sought to investigate the assessment of marginally-performing students in social work programmes, and explain how difficult it is to address such a sensitive issue<sup>41</sup>. Whilst their eventual sample of 35 such students exhibited a range of weaknesses in their performance (the modal number of weaknesses being four), 30 passed the fieldwork component of their programme, and a further two were awarded delayed passes<sup>42</sup>. The default position seems to have generally been that, if there was no evidence of actual incompetence, the student should pass<sup>43</sup>: an issue left hanging was whether there was sufficient time in the placements being studied for the students to demonstrate actual incompetence.

---

<sup>39</sup> See for instance Lankshear (1990); Baird (1991); Ilott and Murphy (1997); Hawe (2003); Furness and Gilligan (2004).

<sup>40</sup> Lang and Woolston (2005) note that 'mateship' is (understandably) a strong part of police culture, which makes negative assessment difficult.

<sup>41</sup> Their paper is essential reading for anyone seeking to study the assessment of marginal performance, irrespective of the subject area.

<sup>42</sup> The assessments of the academic component of their programmes produced a higher proportion of failures.

<sup>43</sup> Lankshear (1990), writing about nursing, suggests assessors do not fail incompetent learners unless there is very clear evidence of unsafe practice. However, Goldenberg and Waddell (1990), who surveyed a convenience sample of 70 nurse educators on baccalaureate programmes, found that more than half of their respondents failed students whom they deemed to be unsafe in practice.

Failing students is stressful for assessors: Ilott and Murphy (1997) refer to the mixture of emotions that can arise, including anxiety, guilt and relief. Failing a student is especially stressful when it has to be done face-to-face. As Ilott and Murphy note: therapists 'look for the positive' (p.311). An examination result is more distanced, psychologically, and the failing candidate can blame a variety of extraneous factors for the outcome. In addition to the possibility of challenges of various sorts, failing a student results in a loss of income to the institution.

A while ago, and by chance, a student on a CQSW programme who had produced good notes of his placement interviews with clients was found to have created these as a complete fiction. The interviews had not taken place. When challenged, the student was unabashed, saying that he would have conducted the interviews along the stated lines had they really taken place (Baird 1991). He failed: sometimes there is no scope for generosity.

### **13 The workplace as a learning opportunity**

Work environments generally, and placements particularly (given the context of this Report), ideally offer their members opportunities to develop their capabilities – i.e. they offer formative opportunities. An appropriate organisational policy is a necessary condition for the sustained success of developmental activity, which cannot sensibly be built on *ad hoc* arrangements and goodwill. It is an insufficient condition since, without the engagement of those in positions to offer formative support, success on any scale will not ensue.

Support, in which formative assessment is very strongly implicit, is one of three influences on learning, the other two being the confidence and commitment of the learner and the challenge and value of the work<sup>44</sup>. Confidence to tackle tasks was seen by Eraut and colleagues to be overwhelmingly important, but depended on the extent of the support that students perceived themselves to be receiving. A low degree of challenge and/or insufficient support may lead to a decline in both confidence and motivation to learn (see Eraut 2004b) – an issue also addressed by Cameron-Jones and O'Hara (1999: see below).

Eraut (2004b, p.21, web version) reported that he and colleagues had found that learners needed 'both short-term, task-specific, feedback and longer-term, more strategic, feedback on general progress'. Further, 'good short-term feedback on performance was often accompanied by an almost total absence of strategic feedback, giving even the most confident workers an unnecessary sense of uncertainty and lowering their commitment to their current employers' (ibid). A problem, highlighted by Eraut et al (2000), that could affect feedback is that people enter and leave workgroups, making them more labile than

---

<sup>44</sup> Eraut (2004b) suggests a similar three-way connection between the allocation and structuring of work; encounters and relationships with people at work; and expectations of each person's role, performance and progress. These contribute to the context within which the work is undertaken.

is implied in the notion of 'communities of practice' (Lave and Wenger 1991; Wenger 1998). Eraut (2004, pp.21-2, web version) also observes:

The half-life of working groups is decreasing in many contexts; so few groups are sufficiently stable and coherent to develop a positive learning climate quickly and spontaneously. Our evidence suggests that management styles and local workplace climates affect learning, retention and quality improvement in similar ways. Hence, managers have to be educated and supported for this role. Surprisingly, this aspect of management capability is very rarely found in management development programmes.

Caballero (2005) noted that feedback was really important to newly qualified nurses as they began their employment, and it mattered less to them whether it was good or bad so long as they knew how they are doing – something that they found really difficult to judge for themselves. She pointed to shortages of nurses appropriately experienced to act as mentors. Interviewed by Caballero, many mentors to newly qualified nurses did not see themselves as 'real' mentors as they did not possess the mentorship qualification. The consequent 'hit and miss' aspect of mentoring also emerged from a study by Gray and Smith (2000) and from an unpublished study, conducted by Yorke in 2003-4, of students' experiences of foundation degrees. In the latter, some mentees were unable to discuss their progress because their mentors were otherwise engaged, whereas others received mentoring that they described as excellent.

It appears that those involved in placements find that some topics are easier to discuss than others. Lopez-Real et al (2001), looking at two teacher education programmes in Hong Kong, noted that supervisors and students found it easier to discuss 'classroom delivery skills' than personal aspects of performance. They stressed the need for trust and gentleness in the supervisor/student relationship: however, there are risks that this could restrict the student's growth. Similarly, Wright and Bottery (1997) found, in their study students and mentors, that the instrumental aspects of lesson planning and delivery were regarded as the most important on which to focus attention. They also found self-development being construed in broadly instrumental terms, with reflection not figuring prominently. Williams (1993) obtained responses from 101 teachers (the majority were heads of department or equivalent) from eight different subject areas regarding their needs as mentors. Pragmatic instrumental needs regarding teaching figured prominently amongst the responses, as did helping students with university assignments, counselling students and providing written feedback on observed lessons.

Some students are quick to develop the skills necessary for effective work performance, as Carlisle et al (1999) found with student nurses, despite the concerns of senior and clinical managers that they lacked core skills. There were some difficulties regarding the assimilation of the students into teams, but placement periods may perhaps be too short for this.

Whilst the majority of 31 supervisors in community mental health surveyed by Bailey (2004) felt able to support students in compiling portfolios, more training would be welcomed by around half of them.

Cameron-Jones and O'Hara (1999) offer a way of approaching the extent to which a placement assists student growth, implicitly pointing to the role that can be played by formative assessment. They adapted a matrix from Daloz (1986) to characterise placements according to two dimensions – level of challenge and level of support. The matrix is represented in Table 7, together with the characterisations they gave to its four cells.

		Level of support	
		Low	High
Level of challenge	High	Retreat	Growth
	Low	Stasis	Confirmation

Table 7 Placements characterised according to the levels of support and challenge (source: Cameron-Jones and O'Hara 1999, p.93).

They devised a rating instrument of eight items and administered it to 578 students on four initial teacher training courses. Whilst 38% of responses fell in the 'growth' cell, the 32% falling in the 'stasis' cell and the 13% falling in the 'retreat' cell are surely triggers for concern. (A possible weakness in the design of the matrix is that the challenge could be set at an unreasonably high level, such that even a high level of support could not engender growth: if so, then the 'retreat' cell would encroach on the 'growth' cell.)

## 14 Concluding comment

The assessment of practice-based professional learning is, on the analysis undertaken in this Report, a complex matter that involves formidable challenges in respect of conception, practice and reporting. Achievements are often highly complex, and much more so than they are sometimes understood to be. Some aspects of achievement are fairly readily amenable to a grading system whereas others are not. The combination of achievements from the different milieux of the higher education institution and the workplace merely compounds the complexity. Hence it is a very demanding task to communicate achievements to a diverse range of stakeholders in such a way that they can appreciate and use the information is a very demanding task, and probably more demanding than the second 'Burgess Group' acknowledged in its recent Consultation Paper on the UK honours degree (UUK and SCoP 2005). Higher education may be making summative assessments with reference to a paradigm that is becoming decreasingly defensible as the complexity of assessment becomes apparent.

There is, manifestly, plenty of work to be done.

---

## References

- Adams J (2003) Observing objective, structure, clinical examinations (OSCE) [sic]. *McGill Journal of Medicine* 7(1), unpagged. Available at <http://www.medicine.mcgill.ca/mjm/issues/v07n01/commentaries/commentaries.htm> (accessed 21 October 2005)
- Alexander HA (1996) Physiotherapy student clinical education: the influence of subjective judgements on observational assessment. *Assessment and Evaluation in Higher Education* 21 (4), pp.357-366.
- Andrews M and Wallis M (1999) Mentorship in nursing: a literature review. *Journal of Advanced Nursing* 29 (1), pp.201-207.
- Arnold J et al (1999) Students' perceptions of competence development in undergraduate business-related degrees. *Studies in Higher Education* 24 (1), pp.43-57.
- Ashworth PD et al (1999) "Levels" of attainment in nursing practice: reality or illusion? *Journal of Advanced Nursing* 30 (1), pp.159-168.
- Aslin PP et al (1995) The role of work-based learning in engineering degree programmes. *Capability* 1 (3), pp.31-38.
- Assiter A, ed, (1995) *Transferable skills in higher education*. London: Kogan Page.
- Aston L et al (2000) An exploration into the role of the teacher/lecturer in practice: findings from a case study in adult nursing. *Nurse Education Today* 20 (3) pp.178-188.
- Atkins M et al (1993) *Assessment issues in higher education*. Sheffield: Employment Department.
- Aviles C (2001) Grading with norm-referenced or criterion-referenced measurements: to curve or not to curve, that is the question. *Social Work Education* 20 (5), pp.603-608.
- Bailey D (2004) The contribution of work-based supervision to interprofessional learning on a masters programme in Community Mental Health. *Active Learning in Higher Education* 5 (3), pp.263-278.
- Baird P (1991) The proof of the pudding: a study of client views of student practice competence. *Issues in Social Work Education* 10 (1&2), pp.24-41.

Baume D and Yorke M (2002) The reliability of assessment by portfolio on a course to develop and accredit teachers in higher education. *Studies in Higher Education* 27 (1), pp.7-25.

Baume D et al (2004) What is happening when we assess, and how can we use our understanding of this to improve assessment? *Assessment and Evaluation in Higher Education* 29 (4) pp.451-477

Benner P (1984) *From novice to expert: excellence and power in clinical nursing practice*. Menlo Park: Addison-Wesley.

Bennett R (2002) Employers' demands for personal transferable skills in graduates: a content analysis of 1000 job advertisements and an associated empirical study. *Journal of Vocational Education and Training* 54 (4), pp.457-475.

Bennetts C (2003) Mentoring youth: trend and tradition. *British Journal of Guidance and Counselling* 31 (1) 63-76

Bergeson T (2004) *Performance-based pedagogy assessment of teacher candidates*. Olympia, WA: Office of Superintendent of Public Instruction. Available at <http://www.k12.wa.us/certification/profed/pedagogy.aspx> (accessed 28 September 2005).

Billett S (2004a) Learning in the workplace: reappraisals and reconceptions. In Hayward G and James S (eds) *Balancing the skills equation: key issues and challenges for policy and practice*. Bristol: The Policy Press, pp.149-170.

Billett S (2004b) Learning through work: workplace participatory practices. In Rainbird H et al (eds) *Workplace learning in context*. London: Routledge, pp.109-125.

Billett S et al (2003) Small business pedagogic practices. *Journal of Vocational Education and Training* 55 (2), pp.149-167.

Blake M and Laing S (2000) Assessment and standards. In Bourner T et al (eds) *New directions in professional higher education*. Buckingham: SRHE and Open University Press, pp.176-199.

Bloom BS (1956) *Taxonomy of educational objectives, Handbook 1: cognitive domain*. London: Longman.

Boud D and Solomon N, eds (2001) *Work-based learning: a new higher education?* Buckingham: SRHE and Open University Press.

Boud D and Symes C (2000) Learning for real: work-based education in universities. In Symes C and McIntyre J (eds) *Working knowledge: the new vocationalism and higher education*. Buckingham: SRHE and Open University Press, pp.14-29.

Brennan J and Little B (1996) *A review of work based learning in higher education*. London: Quality Support Centre, The Open University.

Brackenreg J (2004) The brave new university world: factors affecting quality assurance of nursing graduates in Australian undergraduate programs. *Collegian 11* (1), pp.28-34.

Brandon J and Davies M (1979) The limits of competence in social work: the assessment of marginal work in social work education. *British Journal of Social Work 9* (3), pp.295-347.

Bridges P et al (2002) Coursework marks high, examination marks low: discuss, *Assessment and Evaluation in Higher Education 27* (1), pp.36-48.

Brown, G. with Bull, J. and Pendlebury, M. (1997) *Assessing student learning in higher education*. London: Routledge.

Bullmore, J. (2005) Dear Jeremy. *The Guardian (Work Supplement)* 17 September, p.4.

Burchell H et al (1999) Assessment of competence in radiography education. *Assessment and Evaluation in Higher Education 24* (3), pp. 315-326.

Burgess H et al (1999) Developing Self-Assessment in Social Work Education, *Social Work Education, 18* (2) pp.133-146.

Caballero C (2005) Nurses taking action. Paper presented on behalf of the LiNEA Project team, AERA Conference, Montreal, 15 April. Available at [www.tlrp.org/dspace/retrieve/584/Nurses+Taking+Action+Montreal+Conference+paper+Final+Version.htm](http://www.tlrp.org/dspace/retrieve/584/Nurses+Taking+Action+Montreal+Conference+paper+Final+Version.htm) (accessed 24 October 2005).

Canham J (2002) Assessment of specialist community practice. In Canham J and Bennett J *Mentorship in community nursing: challenges and opportunities*. Oxford: Blackwell Science, pp.85-101.

Carlisle C et al (1999) Skills competency in nurse education: nurse managers' perceptions of diploma level preparation. *Journal of Advanced Nursing 29* (5), pp.1256-1264.

Chambers MA (1998) Some issues in the assessment of clinical practice: a review of the literature. *Journal of Clinical Nursing 7* (3), pp.201-208.

Coleman H et al (2002) Using portfolios to stimulate critical thinking in social work education. *Social Work Education 21* (5), pp.583-595.

Coll RK et al (2002) Assessment of work based learning: some lessons from the teaching profession. *Asia-Pacific Journal of Co-operative Education 3* (2), pp.5-12.

- 
- Cooper B (2000) 8008: The measure of a competent child care social worker? *Journal of Social Work Practice* 14 (2) 113-124
- Cope P et al (2003) Grading the practice of teaching: an unholy union of incompatibles. *Assessment and Evaluation in Higher Education* 28 (6), pp.673-684.
- Cowburn M et al (2000) Assessment of social work students: standpoint and strong objectivity. *Social Work Education* 19 (6), pp 627-637.
- Culley L and Genders N (1999) Parenting by people with learning disabilities: the educational needs of the community nurse. *Nurse Education Today* 19 (6), pp.502-508.
- Daloz L (1986) *Effective teaching and mentoring*. San Francisco, CA: Jossey-Bass.
- Dannefer EF et al (2005) Peer assessment of professional competence. *Medical Education* 39 (7), pp.713-722.
- Dennis J (2003) Order from chaos – the experience of final year business students undertaking consultancy projects. Paper presented at the BEST Conference, Brighton. Available at <http://www.business.heacademy.ac.uk/resources/reflect/conf/2003/dennis/dennis.pdf> (accessed 28 September 2005)
- Donoghue J and Pelletier SD (1991) An empirical analysis of a clinical assessment test. *Nurse Education Today* 11 (5), pp.354-362.
- Driessen E et al (2005) The use of qualitative research criteria for portfolio assessment as an alternative to reliability evaluation: a case study. *Medical Education* 39 (2), pp.214-220.
- Duignan J (2003) Placement and adding value to the academic performance of undergraduates: reconfiguring the architecture – an empirical investigation. *Journal of Vocational Education and Training* 55 (3), pp.335-350.
- Elton L (1998) Are UK degree standards going up, down or sideways? *Studies in Higher Education* 23 (1), pp.35-42.
- ENB (1989) *Preparation of teachers, practitioner teachers, mentors and supervisors in the context of Project 2000*. London: ENB.
- Eraut M (1994) *Developing professional knowledge and competence*. London: Falmer.
- Eraut M (2004a) A wider perspective on assessment. *Medical Education* 38 (8), pp.803-804.

Eraut M (2004b) Informal learning in the workplace. *Studies in Continuing Education* 26 (2), pp.247-273<sup>45</sup>.

Eraut M et al (1995) *Learning to use scientific knowledge in educational and practice settings: an evaluation of the contribution of the physical, biological and social sciences to pre-registration nursing and midwifery programmes [Research Report No.3]*. London: ENB.

Eraut M et al (2000) Development of knowledge and skills at work. In Coffield F (ed) *Differing Visions of a Learning Society, Vol 1*, Bristol, The Policy Press, pp.231-262.

Evans K (2002) The challenges of 'making learning visible': problems and issues in recognizing tacit skills and key competences. In Evans K et al (eds) *Working to learn: transforming learning in the workplace*. London: Kogan Page, pp.79-94.

Farmer EA et al (2002) Assessing the performance of doctors in teams and systems. *Medical Education* 36 (10) 942-948.

Foulkes et al (1994) Combining components of assessment. In Newble D et al (eds) *The certification and recertification of doctors: issues in the assessment of clinical competence*. Cambridge: Cambridge University Press, pp.134-150.

Fowler G and Tietze S (1996) A competence approach to the assessment of student placements. *Education and Training* 38 (1), pp.30-36.

Furness S and Gilligan P (2004) Fit for purpose: issues from practice placements, practice teaching and the assessment of students' practice. *Social Work Education* 23 (4), pp.465-479.

Gammie E and Hornby W (1994) Learning contracts and sandwich education: the accreditation of work-based learning. *Capability* 1 (2), pp.46-58.

Gerrish K et al (1997) *The assessment of practice at diploma, degree and postgraduate levels in nursing and midwifery education: literature review and documentary analysis*. London: English National Board for Nursing, Midwifery and Health Visiting.

Gibbons M et al (1994) *The new production of knowledge: the dynamics of science and research in contemporary societies*. London: Sage.

Gleeson F (1997) Assessment of clinical competence using the Objective Structured Long Examination Record (OSLER). *Medical Teacher* 19 (1), pp.7-14.

---

<sup>45</sup> Electronic version consulted: available via [www.tlrp-archive.org/cgi-bin/search\\_oai\\_all.pl?pn=12&no\\_menu=1&short\\_menu=1](http://www.tlrp-archive.org/cgi-bin/search_oai_all.pl?pn=12&no_menu=1&short_menu=1) (accessed 22 October 2005). Other publications from the LINEA Project can be accessed from this site.

---

Goldenberg and Waddell J (1990) Occupational stress and coping strategies among female baccalaureate nursing faculty. *Journal of Advanced Nursing* 15 (5), pp.531-543.

Gomez S and Lush D (2005) Combining novel pedagogic and IT approaches to align the assessment of workplace learning with criteria for academic credit. Paper presented at the First International Conference on Enhancing Teaching and Learning Through Assessment, Hong Kong.

Gray MA and Smith LN (2000) The qualities of an effective mentor from the student nurse's perspective: findings from a longitudinal qualitative study. *Journal of Advanced Nursing* 32 (6), pp.1542-1549.

Griffiths T and Guile D (2004) *Learning through work experience for the knowledge economy [CEDEFOP Reference Series No.48]*. Luxembourg: Office for Official Publications of the European Communities.

GSCC (2005) *Post-qualifying framework for social work education and training*. London: general Social Care Council.

Guba EG and Lincoln V (1989) *Fourth generation evaluation*. London: Sage.

Hager P (1998) Recognition of informal learning: challenges and issues. *Journal of Vocational Education and Training* 50 (4), pp.521-535.

Hager P (2004a) The competence affair: or why vocational education and training urgently need a new understanding of learning. *Journal of Vocational Education and Training* 56 (3), pp.409-433.

Hager P (2004b) The conceptualization and measurement of learning at work. In Rainbird H et al (eds) *Workplace learning in context*. London: Routledge, pp.242-258.

Hager P and Beckett D (1995) Philosophical underpinnings of the integrated conception of competence. *Educational Philosophy and Theory* 27 (1) pp.1-24.

Hager P and Butler J (1996) Two models of educational assessment. *Assessment and Evaluation in Higher Education* 21 (4), pp.367-378.

Hager P et al (1994) General issues about assessment of competence. *Assessment and Evaluation in Higher Education* 19 (1), pp.3-16.

Halm EA et al (2002) Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. *Annals of Internal Medicine* 137 (6) 511-520.

Harland T (2005) Developing a portfolio to promote authentic inquiry in teacher education. *Teaching in Higher Education* 10 (3), pp.327-337.

---

Harvey L et al (1997) *Graduates' work : organizational change and students' attributes*. Birmingham: Centre for Research into Quality, University of Central England.

Hawe E (2003) 'It's pretty difficult to fail': the reluctance of lecturers to award a failing grade. *Assessment and Evaluation in Higher Education* 28 (4), pp.371-382.

Hawkins P and Winter J (1995) *Skills for graduates in the 21<sup>st</sup> century*, Cambridge, Association of Graduate Recruiters.

Hays RB et al (2002) Selecting performance assessment methods for experienced physicians. *Medical Education* 36 (10) 910-917

Hays R et al (1994) Methods of assessment in recertification. In Newble D et al (eds) *The certification and recertification of doctors: issues in the assessment of clinical competence*. Cambridge: Cambridge University Press, pp.187-200.

Heath P, chair, (2002) *National review of nursing education 2002: our duty of care*. Canberra: Department of Education, Science and Training. Available at [http://www.dest.gov.au/archive/highered/nursing/pubs/duty\\_of\\_care/default.html](http://www.dest.gov.au/archive/highered/nursing/pubs/duty_of_care/default.html) (accessed 27 September 2005).

HEFCE (2000) *Foundation degree prospectus*. Bristol: HEFCE. Available at [www.hefce.ac.uk/pubs/hefce/2000/00\\_27.htm](http://www.hefce.ac.uk/pubs/hefce/2000/00_27.htm) (accessed 22 October 2005).

Heilbronn R et al (2002) School-based induction tutors: a challenging role. *School Leadership and Management* 22 (4) 371-388.

HEQC (1997) *Graduate Standards Programme: final report*. London: Higher Education Quality Council.

Hillier Y (1999) Higher level National Vocational Qualifications: the candidates' experience. *Journal of Vocational Education and Training* 51 (2), pp.199-228.

Hinett K and Weeden P (2000) How am I doing? Developing critical self-evaluation in trainee teachers. *Quality in Higher Education* 6 (3), 245-257.

Holloway K and Haggerty C (2005) Focusing the lenses: competency assessor workshop development. Paper submitted for publication in *Nurse Education Today*.

Howe D (1998) Relationship-based thinking and practice in social work. *Journal of Social Work Practice* 12 (1), pp.45-56.

Hounsell D et al (1996) *The ASSHE inventory: changing assessment practices in Scottish higher education*. Edinburgh: The University of Edinburgh and Napier University, in association with UCoSDA.

- 
- Hussey T and Smith P (2002) The trouble with learning outcomes. *Active Learning in Higher Education* 3 (3), pp.220-233.
- Hyland T (1994) *Competence, education and NVQs: dissenting perspectives*. London: Cassell.
- Ilott I and Murphy R (1997) Feelings and failing in professional training: the assessor's dilemma. *Assessment and Evaluation in Higher Education* 22 (3), pp.307-316.
- Jessup G (1991) *Outcomes: NVQs and the emerging model of education and training*. London: Falmer.
- Johnson R (2004) *Higher education credit practice in England, Wales and Northern Ireland*. Derby: EWN Credit Forum.
- Johnston B (2004) Summative assessment of portfolios: an examination of different approaches to agreement over outcomes. *Studies in Higher Education* 29 (3), pp.395-412.
- Jones A (1999) The place of judgement in competency-based assessment. *Journal of Vocational Education and Training* 51 (1), pp.145-160.
- Jones M (2001) Mentors' perceptions of their roles in school-based teacher training in England and Germany. *Journal of Education for Teaching* 27 (1) 75-94
- Kemshall H (1993) Assessing competence: scientific process or subjective inference? Do we really see it? *Social Work Education* 12 (1), pp.36-45.
- King PM and Kitchener KS (1994) *Developing reflective judgment: understanding and promoting intellectual growth and critical thinking in adolescents and young adults*. San Francisco, CA: Jossey-Bass.
- Knight P and Yorke M (2003) *Assessment, learning and employability*. Maidenhead: SRHE and Open University Press.
- Kohlberg L (1964) *The philosophy of moral development: moral stages and the idea of justice*. San Francisco, CA: Harper and Row.
- Ladyshevsky R (1999) Simulated patients and assessment. *Medical Teacher* 21 (3) pp.266-269.
- Lang K and Woolston R (2005) Workplace learning and assessment of probationary constables in the New South Wales Police. Paper presented at The First International Conference on Enhancing Teaching and Learning through Assessment, Hong Kong.
- Lankshear A (1990) Failure to fail: the teacher's dilemma. *Nursing Standard* 4 (20) pp.35-37.

Lave J and Wenger E (1991) *Situated learning: legitimate peripheral participation*. Cambridge: Cambridge University Press.

Lew SR et al (2002) Procedures for establishing defensible programmes for assessing practice performance. *Medical Education* 36 (10) 936-941.

Lopez-Real F et al (2000) Supervisory experiences: an exploration of some difficult topics. *Journal of Education for Teaching* 27 (2), pp.161-173.

Lum G (1999) Where's the competence in competence-based education and training? *Journal of Philosophy of Education* 33 (3), pp.403-418.

Lyons F and Bemell M (2001) Setting the standards: judging levels of achievement. In Boud D and Solomon N (eds) *Work-based learning: a new higher education?* Buckingham: SRHE and Open University Press, pp.167-183.

Mager RF (1962/1990) *Preparing instructional objectives* [2<sup>nd</sup> ed]. London: Kogan Page. Originally published at Palo Alto: Fearon.

Manthorpe J et al (2005) The evolution of social work education in England: a critical review of its connections and commonalities with nurse education. *Nurse Education Today* (forthcoming).

Marshall IS and Cooper LSM (2001) Earning academic for part-time work. In Boud D and Solomon N (eds) *Work-based learning: a new higher education?* Buckingham: SRHE and Open University Press, pp.184-199.

Martin IG and Jolly B (2002) Predicted validity and estimated cut score of an objective structured clinical examination (OSCE) used as an assessment of clinical skills at the end of the first clinical year. *Medical Education* 36 (5), pp.418-425.

McCullan M et al (2003) Portfolios and assessment of competence: a review of the literature. *Journal of Advanced Nursing* 41 (3), pp.283-294.

McCulloch R (2005) From the classroom to Kajulu and beyond: authentic assessment within an industry-professional context. Paper presented at the First International Conference on Enhancing Teaching and Learning Through Assessment, Hong Kong.

Melnick DE et al (2002) Conceptual challenges in tailoring physician performance assessment to individual practice. *Medical Education* 36 (10) pp.931-935.

Mezirow J (1991) *Transformative dimensions of adult learning*. San Francisco, CA: Jossey-Bass.

- 
- Miller CML and Parlett M (1974) *Up to the mark: a study of the examination game [Monograph 21]*. London: Society for Research into Higher Education.
- Miller GE (1990) The assessment of clinical skills / competence / performance. *Academic Medicine* 65 (9: Supplement), pp.S63-S67.
- Murrell J (1993) Judgement of professional competence: bags of bias. *Social Work Education (Special publication: 'Assessment of competence in social work law')*, pp.5-19.
- Needham R (1988) Work-based learning and assessment. *Forum for the Discussion of New Trends in Education* 30 (2), pp.46-7.
- Newble DI and Jaeger K (1983) The effect of assessments and examinations on the learning of medical students. *Medical Education* 17 (3), pp.165-171.
- Newble D et al (1994) *The certification and recertification of doctors: issues in the assessment of clinical competence*. Cambridge: Cambridge University Press.
- NCIHE (1997) *Higher education in the learning society*, Report of the National Committee of Inquiry into Higher Education, chaired by Ronald Dearing [Main Report]. Norwich: HMSO.
- Nicklin PJ and Kenworthy N (2000) *Teaching and assessing nursing practice: an experiential approach [3<sup>rd</sup> ed]*. London: Ballière Tindall.
- Norcini JJ (2005) Current perspectives in assessment: the assessment of performance at work. *Medical Education* 39 (9), pp.880-889.
- Norman GR et al (1991) Pitfalls in the pursuit of objectivity: issues of validity, efficiency and acceptability. *Medical Education* 25 (2), pp.119-126.
- O'Donovan B et al (2004) Know what I mean? Enhancing student understanding of assessment standards and criteria. *Teaching in Higher Education* 9 (3), pp.325-335.
- Ottewill R and Wall A (1996) Sandwich education. *Capability* 2 (2), pp.18-24.
- Owens C (1995) How the assessment of competence in DipSW is changing the culture of practice teaching. *Social Work Education* 14 (3), pp.61-78.
- Palmer A et al (1994) *Reflective practice in nursing: the growth of the professional practitioner*. Oxford: Blackwell Science.
- Pawson R (2004) Mentoring relationships: an explanatory review. ESRC UK Centre for Evidence Based Policy and Practice, Working Paper 21. Available at <http://www.evidencenetwork.org/Documents/wp21.pdf> (accessed 22 September 2005).

- Perry WG (1970/1998) *Forms of ethical and intellectual development in the college years*. San Francisco, CA: Jossey-Bass [Reprint in 1998 of the 1970 text, with a new introduction].
- Pitts J et al (1999) Educational portfolios in the assessment of general practice trainers: reliability of assessors. *Medical Education* 33 (7), pp.515-520.
- Powell JH (1989) The reflective practitioner in nursing. *Journal of Advanced Nursing* 14 (10), pp.824-832.
- Purcell K and Pitcher J (1996) *Great expectations: the new diversity of graduate skills and aspirations*. Coventry: Institute for Employment Research, University of Warwick.
- Ram P et al (1999) Assessment of general practitioners by video observation of communicative and medical performance in daily practice: issues of validity, reliability and feasibility. *Medical Education* 33 (6) 447-454.
- Redfern S et al (2002) Assessing competence to practise in nursing: a review of the literature. *Research Papers in Education Policy and Practice* 17 (1), pp.51-77.
- Reynolds A (1992) What is competent beginning teaching? A review of the literature. *Review of Educational Research* 62 (1), pp.1-35.
- Richardson R (1995) Humpty Dumpty: reflection and reflective nursing practice. *Journal of Advanced Nursing* 21 (6), pp. 1044-1050.
- Rickard W (2002) Work-based learning in health: evaluating the experience of learners, community agencies and teachers. *Teaching in Higher Education* 7 (1), pp.47-63.
- Sadler DR (1987) Specifying and promulgating achievement standards. *Oxford Review of Education* 13 (2), pp191-209.
- Sadler R (2005) Interpretations of criteria-based assessment and grading in higher education. *Assessment and Evaluation in Higher Education* 30 (2), pp.175-194.
- Schaafsma H (1996) Back to the real world: work placements revisited. *Education and Training* 38 (1), pp.5-13.
- Schutz S et al (2004) Assessing and evaluating reflection. In Bulman C and Schutz S (eds) *Reflective practice in nursing (3<sup>rd</sup> ed)*. Oxford: Blackwell, pp.47-72.
- Schuwirth LWT et al (2002) When enough is enough: a conceptual basis for fair and defensible practice performance assessment *Medical Education* 36 (10) 925-930
- Schwartz RW et al (1997) Problem-based learning and performance testing: effective alternatives for undergraduate surgical education and assessment of student performance. *Medical Teacher* 19 (1) 19-23.

Shardlow S and Doel M (1993) Examination by triangulation: a model for practice teaching. *Social Work Education* 12 (3), pp.67-79.

Shumway and Harden (2003) The assessment of learning outcomes for the competent and effective physician [AMEE Guide No.25]. *Medical Teacher* 25 (6), pp.569-584.

Smith R and Betts M (2000) Learning as partners: realising the potential of work-based learning. *Journal of Vocational Education and Training* 52 (4), pp.589-604.

Stasz C et al (1996) Workplace skills in practice: case studies of technical work. Santa Monica, CA: RAND/National Center for Research in Vocational Education, University of California. Available at [www.rand.org/publications/MR/MR722/](http://www.rand.org/publications/MR/MR722/) (accessed 22 October 2005).

Sternberg RJ (1997) *Successful Intelligence*. New York: Plume.

Stones E (1994) Assessment of a complex skill: improving teacher education. *Assessment in Education* 1 (2), pp.235-251.

Swanson DB et al (1997) Strategies for student assessment. In Boud D and Feletti GE (eds) *The challenge of problem-based learning (2<sup>nd</sup> ed)*. London: Kogan Page, pp.269-282.

Tapper J (2004) The challenge of meeting students' expectations when providing work placements for generalist students. Paper presented at the 6<sup>th</sup> International Conference 'Practical Experiences in Professional Education', Brisbane, 2-4 February. Available at [http://www.pepe2004.qut.edu.au/graphics/Web%20Papers/joanne\\_tapper.pdf](http://www.pepe2004.qut.edu.au/graphics/Web%20Papers/joanne_tapper.pdf) (accessed 28 September 2005).

Taylor I et al (1999) Portfolios for learning and assessment: laying the foundations for continuing professional development, *Social Work Education*, 18 (2) pp.147-160.

Tigelaar DEH et al (2005) Quality issues in judging portfolios: implications for organizing teaching portfolio assessment procedures. *Studies in Higher Education* 30 (5), pp.595-610.

Toohy S et al (1996) Assessing the practicum. *Assessment and Evaluation in Higher Education* 21 (3), pp.215-227.

Torrance H and Coultas J (2004) *Do summative assessment and testing have a positive or negative effect on post-16 learners' motivation for learning in the learning and skills sector? A review of the research literature on assessment in post-compulsory education in the UK*. London: Learning and Skills Research Centre.

UUK and SCoP (2005) *The UK honours degree: provision of information [Consultation Paper from the Measuring and Recording Student Achievement Steering Group, the "Burgess Group"]*. London: Universities UK and Standing Conference of Principals.

Van der Hem-Stokroos et al (2005) Reliability of the Clinical Teaching Effectiveness instrument. *Medical Education* 39 (9), pp.904-910.

Van der Vleuten CPM et al (1991) Pitfalls in the pursuit of objectivity: issues of reliability. *Medical Education* 25 (2), pp.110-118.

Walkington J (2004) Assessing professional experience learning: dilemmas for the partnership. Paper presented at the 6<sup>th</sup> International Conference 'Practical Experiences in Professional Education', Brisbane, 2-4 February. Available at [http://www.pepe2004.qut.edu.au/graphics/Web%20Papers/jackie\\_walkington.pdf](http://www.pepe2004.qut.edu.au/graphics/Web%20Papers/jackie_walkington.pdf) (accessed 28 September 2005).

Watson R et al (2002) Clinical competence assessment in nursing: a systematic review of the literature. *Journal of Advanced Nursing* 39 (5), pp.421-431.

Wellard S et al (2000) Staffing of undergraduate clinical learning programs in Australia. *Nurse Education Today* 20 (7), pp.548-554.

Wenger E (1998) *Communities of practice: learning, meaning and identity*. Cambridge: Cambridge University Press.

Wilkinson TJ et al The use of portfolios for the assessment of the competence of doctors in practice. *Medical Education* 36 (10) 918-924.

Winter R (1994a) The problem of educational levels, Part 2. *Journal of Further and Higher Education* 18 (1), pp.92-106.

Winter R (1994b) Work-based learning and quality assurance in higher education. *Assessment and Evaluation in Higher Education* 19 (3), pp.247-256.

Winter R and Maisch M (1996) *Professional competence and higher education: the ASSET Programme*. London: Falmer.

Winter R (2001) A challenge to assessment and quality assurance in higher education. In Boud D and Solomon N (eds) *Work-based learning: a new higher education?* Buckingham: SRHE and Open University Press, pp.155-166.

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

Wood J (1997) Developing key skills through the accreditation of work-based learning: undergraduate student tutoring in schools. *Capability* 3 (2), pp.29-34.

Worth Butler M et al (1994) Towards an integrated model of competence in midwifery. *Midwifery 10*, pp.225-231.

Wright N and Bottery M (1997) Perceptions of professionalism by the mentors of student teachers. *Journal of Education for Teaching 23* (3), pp.235-252.

Yorke M (1998) Assessing capability. In Stephenson, J. and Yorke, M. (eds) *Capability and quality in higher education*. London: Kogan Page, pp.174-191.

Yorke M (1999) The skills of graduates: a small enterprise perspective. In O'Reilly D et al (eds) *Developing the capable practitioner*. London: Kogan Page, pp.174-183.

Yorke M (2003) Going with the flow: first-cycle higher education in a lifelong learning context. *Tertiary Education and Management 9* (2), pp.117-130.

Yorke M (2004) *Employability: what it is – what it is not*. York: Learning and Teaching Support Network.

Yorke M and Longden B (2005) *Significant figures: performance indicators and 'league tables'*. London: Standing Conference of Principals.

Yorke M et al (2005) Mining institutional datasets to support policy making and implementation. *Journal of Higher Education Policy and Management 27* (2), pp.285-298.

Young G (1999) Using portfolios for teacher preparation and health sciences. In Brown S and Glasner A (eds) *Assessment matters in higher education: choosing and using diverse approaches*. Buckingham: SRHE and Open University Press, pp.122-131.

Young M (2004) Conceptualizing vocational knowledge: some theoretical considerations. In Rainbird H et al (eds) *Workplace learning in context*. London: Routledge, pp.185-200.

Zegwaard K et al (2003) Assessment of workplace learning: a framework. *Asia-Pacific Journal of Co-operative Education 4* (1), pp.10-18.