*Global J. of Engng. Educ., Vol.7, No.1*
*Published in Australia*



# The Assessment of Complex Learning Outcomes*

**Peter T. Knight**
*Centre for Outcomes-Based Education, The Open University*
*Walton Hall, Milton Keynes, MK7 6AA, England, United Kingdom*
**William M. Banks**
*University of Strathclyde, 16 Richmond Street, Glasgow, G1 1XQ, Scotland, United Kingdom*

The Engineering Professors' Council (EPC) produced an *output standard* in 2000 containing a set of 26 generic statements of what an engineering graduate should have an *ability to* tackle. In addition, Higher Education (HE) is concerned with the promotion of complex or advanced understanding of subject matter. This leads to *complex learning* outcomes, which need to be adequately assessed. Changing demands mean changing assessment practices. While good practice is being used in many cases, there is a need to ensure assessment stimulates complex learning. The article seeks to address these issues.

## INTRODUCTION

The output standard of the Engineering Professors' Council (EPC) comes out of consultation within Higher Education (HE) and with employer organisations and accrediting bodies. It describes what is expected of all engineering graduates in terms of 26 generic statements of graduates' *Ability to* tackle an engineering process [1]. Insofar as it is based on an analysis of what engineers *do*, it fits well with Haug and Tauch's comment that *enhanced employability seems to be the strongest source of change and reform in* [European] *higher education* [2].

However, these *Ability to* statements are insufficiently informative on their own so they have been exemplified by statements from providers of degree courses in particular engineering disciplines. Such statements are referred to as exemplar *benchmarks*. The standard and methodology were validated by nine pilot universities who developed benchmark statements for a range of their engineering programmes in the main engineering disciplines. This illustrates one of the fundamental strengths of the EPC output standard: the generic *Ability to* statements provide a framework describing what *all* engineering graduates must be able to do, which individual programmes can then benchmark to describe and communicate the intended threshold level. In fact, realistically, it may be that this *framework* is the most valuable result of the output standard project: providing a common language that different stakeholders can use to describe their desires or attainments at whatever level may be of concern.

*A revised and expanded version of a presentation given at the 3rd *Global Congress on Engineering Education*, held in Glasgow, Scotland, UK, from 30 June to 5 July 2002.

## COMPLEX LEARNING

European higher education is obviously concerned with the promotion of complex, or advanced, *understandings* of subject matter, as is the EPC's output standard.

If one aspect of this sort of thinking is then explored, namely the view that HE should contribute to student employability, complex learning outcomes can be found aplenty. A glance at some of the research on employability shows that, among other things, HE is expected to foster: willingness to learn; self-management skills; communication skills; effective learning skills; exploring and creating opportunities; action planning; networking; coping with uncertainty; transfer skills; self-confidence; teamworking; managing others; critical analysis; being able to work under pressure; and imagination/creativity [3-5]. This calls on Higher Education Institutions (HEIs) to complicate what they have been doing. For instance, helping students to make strong claims to being highly employable people implies some preparation for: participating in problem-solving, consultative committees and quality circles; formal and informal on-the-job training; flexible teamworking; and understanding the sorts of identities that are valued in workplaces and appreciating how to take them on [6].

So far, this article has focused on describing what it is that makes HE engineering programmes complex. This idea can be clarified by identifying some things that are not regarded as complex. Complicated learning, for example, is different. It can be complicated to memorise procedures, formulae, sequences and plots, especially if several sources have to be used in the process. That is not complex because the outcome can be defined in fairly convergent, fixed or determinate ways. Nor is formal operational thinking, the highest epistemological level identified by Piaget, complex, because, at least in most of his examples, it is about the application of mathematical and scientific reasoning to solve determinate, convergent problems. It may be tough – for most of us, it *is* tough - but there are answers that are generally recognised to be the right answers and known procedures for getting them.

It is known that complex learning takes time: Norman says up to 5,000 learning hours [7]. Even if Norman's figure is disputed, complex learning usually takes a lot longer than a single module allows, sometimes appearing unexpectedly weeks, months or years after the stimulus that got it started [8]. While information and inert knowledge can, in principle, be fixed in some form of memory in a fairly short time, and while the convergent use of formulae can also become quite quickly routinised (how long does it take to learn how to do $c^2$ tests on a calculator?), complex social and academic practices can take years. That has profound implications for the design of student learning environments and for the assessment of their learning.

This learning, which characterises higher education and which suffuses the output standard, is *fuzzy* learning, but it has not been widely researched. Even without research, it can be quickly seen that there are profound issues to address and informed guesses need to be made about how best to do so.

## PROGRAMME DESIGN ISSUES

There is not space to say a great deal about the issues involved, the main one being, of course, that slow learning means programme-level, not module-level, thinking. Three less obvious design issues are described below.

### The Need to Stimulate Practical Intelligence

Employability, and the output standard more generally, might be construed as a mix of emotional and practical intelligences [9][10]. There is a body of research on these constructs that is suggestive about the development pathway of those achievements that make for employability in early adulthood. There is also a useful working knowledge of the extent to which interventions in the non-cognitive domain may be successful and about the characteristics of successful interventions [11][12].

Significantly, there is also evidence that these *intelligences* scarcely correlate with academic intelligence [13]. The implication is that employability and similar complex outcomes of learning will not be stimulated by the routines that have been used to enhance student scores on tests of academic achievement. As was noted earlier, it is also understood that definite arrangements will need to be made to increase the chances, which are normally slim chances, of students applying or transferring learning from one place and time to others. The conclusion is that fresh learning, teaching, assessment and curriculum strategies, such as the ones being piloted in the Skills *plus* project, will be needed if the output standard is to achieve its promise.

### The Need to Design for Non-Formal Learning

It is increasingly appreciated that most professional learning is non-formal learning [14-18]. Two conclusions are:

- Any strategy that tries to enhance complex learning will be limited if it relies on formal learning. The reason is that formal learning may have its place but it is not authentic (in the sense that it is not the main shape that authentic learning takes) and it may be too de-contexted, even artificial, to be of much use in the workplace.
- By definition, the course and outcomes of non-formal learning engagements cannot be pre-specified. If non-formal learning is to be taken seriously, new approaches to curriculum need to be developed in order to move from programme design and determinate learning outcomes to the design of learning environments rich in affordances for complex learning. One of the most significant features of learning environments is the nature of the workgroups and communities in which learning happens [19][20]. Consideration needs to be given to social environments in which engineering teachers and students work as much as, if not more than, the design of their physical working environments. When it comes to Web-based and networked learning, it is all the more important to think carefully about the ways in

which online learners might become vibrant learning communities [21-23].

### Learning from Work Done on the Design of Environments for Online Learning

To echo the previous point: scholars interested in effective online learning (for example [24][25]) are concerned with the design of whole learning environments that encourage complex achievements. This literature has produced metaphors and principles that can inform the design of whole learning environments, which should favour the emergence of complex outcomes of learning, such as those captured by the output standard [26][27]. The Skills *plus* project has begun to apply them to the work of enhancing existing face-to-face curricula in a wide range of subject areas, although not in engineering [28].

In assessing the output standard, it has been observed that:

> *The single, strongest influence on learning is surely the assessment procedures … even the form of an examination question or essay topics set can affect how students study … It is also important to remember that entrenched attitudes which support traditional methods of teaching and assessment are hard to change* [29].

The theme of this section is that assessment practices need to be changed, perhaps quite dramatically, in order to support the output standard. There is some comfort in the finding that engineering teachers are using a good range of appropriate techniques, although some may be disconcerted to realise how much needs to be done to get them in a coherent relationship that can stimulate complex learning.

## EVIDENCE OF GOOD PRACTICE

This section reports the findings of a 2001 survey of EPC members, which was designed to get a better understanding of what works well in present assessment practices and what is proving problematic. It was kept simple in the hope that more engineering teachers would then complete it. Forty-eight usable responses to a semi-structured questionnaire were received. No claim can be made that the findings are representative but it is believed that they identify the main features of assessment practices and points of stress in them.

The survey found that UK engineers were already adapting their assessment methods to developments in engineering curricula by adopting a good range of assessment methods. Specifically:

- All informants were using examinations, emphasising their importance in providing secure judgements of individual attainments (there are lively concerns about plagiarism in coursework).
- Time-constrained tests, often done in lectures, were reported by almost half the informants.
- Virtually all informants used projects work and reports of project work to assess students.
- Three quarters referred to presentations.
- Just over half of the informants mentioned using laboratory reports for assessment purposes.
- Design studies were specifically identified as a powerful assessment method by about a quarter of respondents.
- About a quarter praised *viva voce* examinations or other oral investigations as searching appraisals of understanding and good safeguards against plagiarism. A similar number valued assessment by poster presentation.

One conclusion is that a good range of assessment methods is in use. In the words of one EPC member:

> *The methods employed currently are perfectly adequate. They provide for a variety of assessments and allow both formative and summative feedback. The methods have evolved over a number of years and are still being enhanced and improved. I would expect to be looking continually at what we do and how we do it and developing new strategy's as we move along* (Informant #37).

This conclusion is strengthened by responses to a question, which asked whether these assessment approaches seem to satisfy employers. Almost three-fifths thought they did and while another third had suggestions for improvement, they thought existing approaches broadly satisfied them.

Even so, there are unanswered questions about the quality of these practices. Diversity of practice is not a guarantee of diversity of good quality practices and there is a problem understanding how the potential contained in good, diverse practices can be realised across the system of undergraduate engineering as a whole.

The next section considers the survey data on what more might be needed to align these promising assessment practices with the EPC output standard, followed by an analysis of the difficulties that might be anticipated.

## LIMITING FACTORS

The 2001 survey established that extra demands on engineering teachers, such as the demands of revising programme assessment practices so as to align them with the authentic *Ability to* statements, would test a system already in tension.

Informants were not confident that their conditions of work were conducive to the spread of existing good practices and suggested that fresh demands, such as those implied by trying to assess the EPC's output standard, could not be met. They identified a number of contributors to this state as follows:

- The prime contributor was the semester system: no one had anything good to say about it. Complaints included that it led to a bunching of assignments, that scripts had to be marked to tight deadlines, leading to what one person called severe time compression. Reference was also made to fragmentation and to the difficulties of scheduling complex and authentic assessments in semester-long courses (by the time students have learned enough to be able to tackle complex assignments, there is not enough time left for them to undertake them). Opportunities for formative assessment could be similarly restricted.
- Time was widely felt to be in short supply. Improved quality assurance procedures, tightening up double marking practices, for example, added to pressures on time.
- New assessment methods were valued but seen as costly, particularly in the sense of demanding a lot of time (for students to do them and for teachers to mark them).
- Large classes and rising student numbers have exacerbated tensions.
- More valid assessment methods often made it harder to detect plagiarism.

## ASSESSMENT: ALTERNATIVE CONCEPTUAL FRAMEWORKS

Before progress can be made with regard to devising assessment arrangements that are fit for the purposes implicit in the output standard, it is important to understand social measurement theory, that we as educators understand what can be assessed, how and with what certainty [30]. Yet one of the biggest challenges to the establishment of assessment regimes that serve the output standard well is the prevalence of common-sense notions of what assessment is. Carter says that:

> *It is a commonplace of Engineering that any statement of requirements (requirements specification) is incomplete without a test specification. The argument is that any requirement which is not capable of being tested or verified in some way is meaningless* [31].

This tends to produce the following conclusions:

- There must be objective and reliable measures of the requirements or specification.
- Any assessment procedure that falls short is therefore defective and a waste of time and effort.

Leaving aside the objection that where complex and indeterminate outcomes are concerned, the best that can be done is to ensure that good process standards are in place and trust that they will tend to have effects in the desired direction. Instead, consider the objection that all assessment, especially where human thinking and doing are concerned, rests on judgement of available evidence. There are a few cases where judgement may be akin to measurement but, in general, human thinking and doing are not susceptible to measurement, only to good judgements. Hamer stated that:

> *What much recent work on assessment has indicated is that the gold standard* [examining and testing techniques] *is not quite as refined as was commonly believed: that there are not quite as many things we can assess with certainty as was once thought, and that those that we can are not necessarily the most worthwhile or useful. This is helping to free up thinking* [32].

It follows that good practice in the assessment of engineering achievements depends on good understandings of the assessment of human achievements. The success of the output standard may be closely related to the degree to which engineering teachers reject the assumption that assessment is measurement.

## THE LIMITS OF SUMMATIVE, HIGH STAKES, HIGH-RELIABILITY ASSESSMENT

Most assessment in higher education is summative. It warrants or certifies students' achievements, which means that it is a high-stakes, graded judgement of achievement. When the purposes of assessment are summative - to provide *feedout* - reliability is at a premium. Some achievements can easily be reliably assessed. These assessments are called *low-inference* assessments and are typified by MCQ tests of information retention. Low-inference assessments may be reliable but they only work with determinate achievements where there is little ambiguity about the correct answer. EPC output standards put considerable emphasis on achievements that are far more complex, where credit could be given for a range of solutions and for the means by which the solutions were developed.

In general, reliability is costly, can be difficult to achieve, and is often to be bought by using artificial techniques that may be poor predictors of life-like performances. Complex processes are required to judge complex abilities and the more complex the abilities that the performance is supposed to show, the more samples are needed and the more complex is the assessment process. The process can be simplified but only by simplifying that which is to be assessed, but simplification is at the price of validity. For example, the ability to transform existing (complex and fuzzy) systems into conceptual models, which are then to be transformed into determinable models is a sophisticated set of problem-*working* abilities. It is not validly assessed by tasks in which parameters are set for the student so that standard methods can be routinely applied to *solve* the problem. This may make for more reliable assessment, but in the process, the abilities in question have become simplified: routine problem solving has been substituted for complex problem working. If validity is to be preserved, reliability costs soar.

In other words, there are sharp questions to be asked about their validity or worth. Where complex learning achievements are in question, there is a tension between the demands of reliable assessment and the requirements of valid assessment. A common response is to go for reliability. Understandable though that is, the ten points listed below should give pause for thought because they suggest that reliable assessment is something of a chimera.

## PROBLEMS WITH HIGH-STAKES ASSESSMENTS OF COMPLEX ACHIEVEMENTS

Problems encountered with high-stakes assessments of complex achievements include the following:

- *Knowledge and knowing*: Assessment involves making assumptions about what exists, what it is like and how we might know about it. For example, if skills are nothing more than convenient terms for social practices that are decidedly situation-specific, hence changeable, then it will be frustrating to try and assess skills as if they were real, generalisable achievements. Again, what some take to be a psychological property, such as self-esteem, that is measurable and has explanatory powers may, in fact, be no more than a non-stable self-evaluation without any explanatory powers.
- *The limits of reliability*: Plainly, fictional objects of assessment cannot be evaluated with validity. Where validity is lacking, reliability is compromised. So, were skills to be fictions, there would be interesting validity and reliability issues attaching to all efforts to assess them. So, too, with other qualities that HEIs might claim to promote (eg self-motivation).
- *The stability of assessment judgements*: If a HEI wishes to warrant achievement, then it should be based on several assessors judging different instances of it. Programmes have widely been deconstructed by modularisation and increased student choice, which makes this desirable summative assessment practice rather elusive.
- *The transferability of achievement*: Achievements that grade or degree classes signify may not be very transferable. Many psychologists say that transfer is an achievement in its own right, not something that flows freely and easily, except in familiar settings where specific transfer heuristics have been routinised. So, we do not know whether degree classes or grades indicate a performance achieved with the help of plenty of scaffolding or with none, which makes it prudent to doubt whether it warrants describing achievements that the learner can readily and independently transfer to fresh settings.
- *Limitations to criteria-referencing*: Benchmarks, specifications, criteria and learning outcomes do not and cannot make summative assessment reliable, may limit its validity and certainly compound its costs. Difficulties are reported in getting agreement on criteria and their application in a subject and in a school. There remain significant variations between groups of HEIs and between subject communities.
- *Assessment and curriculum skew*: High stakes assessments have to be robust enough stand up to legal challenge, so they tend to rest on assessments of things that people believe can be judged reliably. This distorts the curriculum in two ways. Firstly, things covered by high stakes assessment get serious attention, while others do not. Secondly, achievements that are not warranted by high

stakes assessment are neither recorded nor celebrated. In such ways, the enacted curriculum becomes what high stakes judgements cover.

- *The misuse of number*: Summative assessment data are usually presented numerically but they really ought not to be treated *numerically*. Educators should be wary of numbers created by summative assessment and mistrust conclusions based on the transformation or manipulation of those numbers.
- *The opacity of number*: Some grades or classifications are based only on examinations, some only on coursework, and some on varying mixes of the two. Likewise, a degree classification may describe students' sustained performance across the programme, the level they reached at the end of it, or some unknown blend of the two.
- *Process-blindness*: Scores and grades are silent about the learning processes involved. This matters because if an educator is told that someone has repeatedly shown that they can solve problems and the educator then finds that problem-solving has been taught and learned as the manipulation of numbers according to learned algorithms, the educator may be less impressed than if he/she hears that it has been developed through engagement with a series of *fuzzy*, authentic tasks.
- *Utility*: Summative assessments may appear to speak reliably about some achievements at given points in the undergraduate years but be moderate or poor predictors of career achievement. Employers, who might be expected to rely on summative assessment data, often mistrust assessment data, probably for this reason [33].

The EPC's Assessment Working Group has accepted that there do need to be reliable assessments of some of the *Ability to* statements, even though the resultant scores may neither as useful, nor as meaningful as is sometimes assumed. In order to reach other outcomes of learning, the Working Group has capitalised on a well-established distinction between assessment that has summative purposes and that which has formative purposes.

The aim of formative assessment is to provide an opportunity for students to experiment in a *safe* environment and to identify their own level of performance and how they might improve their future performances. With formative assessments, the stakes are perceived to be lower; less is visibly at risk if there is error in the judgement. Any learning achievement can be the subject of low-stakes, formative assessment, even complex ones relating to ill-defined or *soft* skills. In such circumstances, it would be hard to claim that the assessor's judgement would be as reliable as, say, a score on a set of Multiple-Choice Questions (MCQs), but that need not matter. The purpose is conversational, the anticipated outcome is learning and learning often involves dialogue. Seen like that, the assessor's judgement is a starting point in a learning conversation. It is not a final judgement and, although it should obviously be a fair judgement, it does not have to be reliable in the same way as summative assessments.

The Assessment Working Group's view is that all of the *Ability to* statements *can* be assessed in some way. However, that does not mean that all can be summatively (reliably) assessed, let alone within the resources available to most departments. Engineering departments are advised to plan a differentiated, programme-wide approach to assessment if they are to cover all or most of the *Ability to* statements.

## THE POTENTIAL OF LOW STAKES, FORMATIVE ASSESSMENT

Taken together, the objections to trusting that valued learning outcomes can all be reliably assessed at an affordable cost suggest that the further we move from the assessment of simple achievements, like information recall, the less feasible it becomes to make reliable judgements. The EPC's output standard has little to say about simple achievements, although it is clearly recognised that complex learning depends upon information, recall, command of algorithms and such like. If this is the case, then ideas to help create assessment systems that are fit for the output standard need to be found elsewhere. Figure 1 suggests that more use be made of formative assessment so that:

- Many outcomes/abilities/achievements would be formatively assessed. This assessment would be low-stakes, designed to give learners useful feedback on how to improve performance against programme-wide criteria. It would be embedded in the learning activities. Student participation in formative assessment would be a requirement for progress through the programme.
- Feedback should then be fast, focused, relevant to the assessment criteria, and be developmental and personal to the student. Reliability would come second to plausibility of judgement because if a learner felt that a judgement was wrong, then it would be important in the interests of learning for there to be open dialogue about that. This could help to reduce the incidence of the undesirable *final language* of assessment and generally to

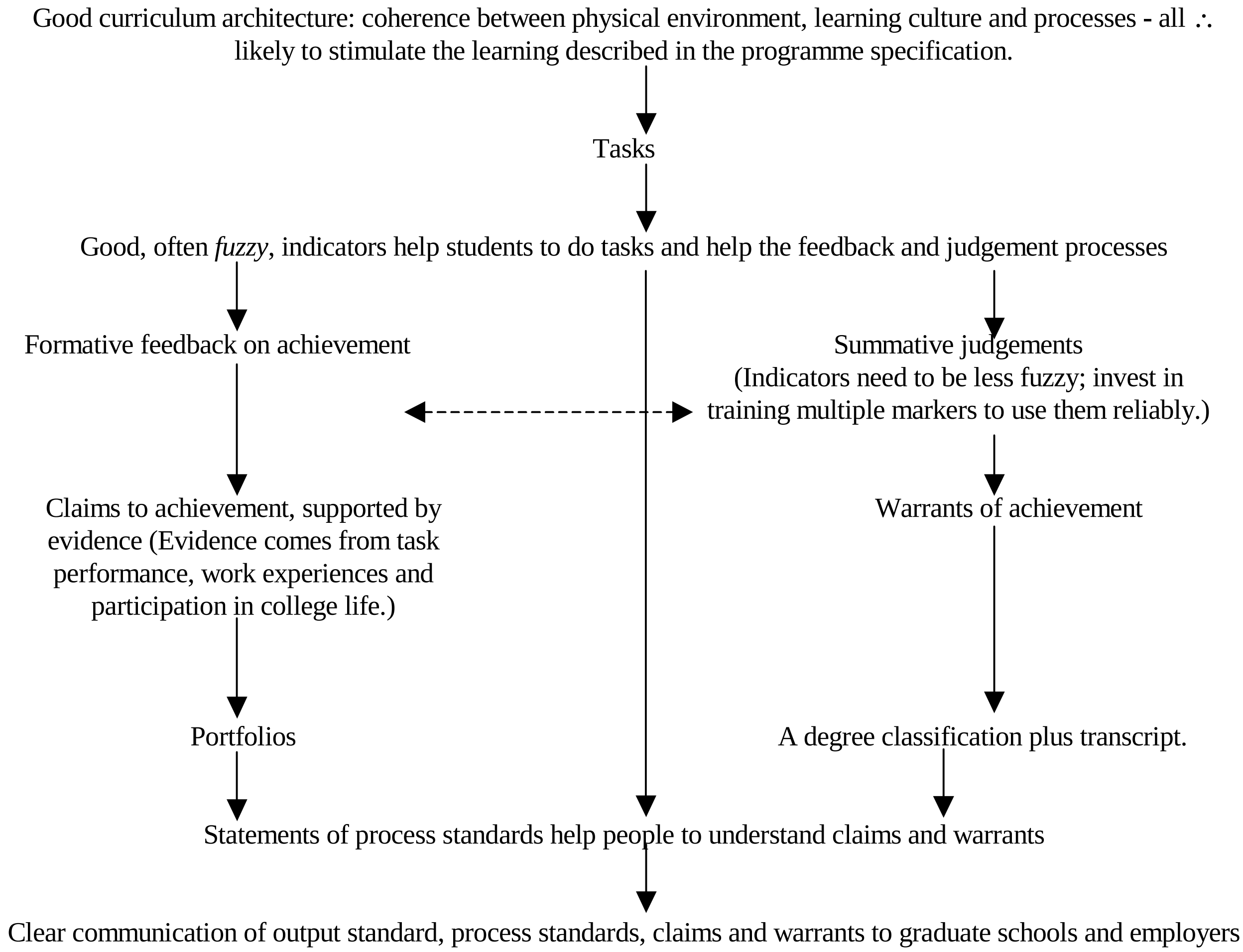


Figure 1: Differentiated programme-level assessment arrangements.

reduce the negative emotions associated with the assessment of learning.

- Authentic assessments would become easier to manage. The bugbear of authentic assessments has been getting reliability levels that are good enough for high-stakes purposes. Reliability is not such an issue when assessments are low-stakes and the main intention is to promote learning dialogues that inform future work.
- Each programme learning outcome should then be complemented by grade indicators, including threshold descriptors, which would give teachers and students a better idea of what would be rewarded.
- Therefore, students should have the programme criteria from the first, regularly use them, share them and practise applying them.
- Peer- and self-assessment should be embedded in programmes. Both save teachers time (which can then be used on high-stakes assessment) and help learners become familiar with programme grade indicators. There have been heroic attempts to devise summative self- and peer-assessment systems, but the position here is that they are best kept for formative purposes.
- Information and communications technology would support the on-demand self-assessment that can provide feedback and even coaching on points of difficulty.

The value of this formative approach to assessment can best be shown by reference to pages 11-14 of the *Interim Report* [1]. The civil engineering *Ability to* statements say graduates should have experience in relation to ten statements and awareness in relation to six. Expressed in these terms, these are *Ability to* statements that resist summative assessment. However, students should benefit from plenty of opportunities for formative feedback on work related to these 16 statements. Both teachers and students should benefit from using fuzzy learning criteria or indicators to organise their assessment conversations.

As for the other nine *Ability to* statements, departments might wish to invest quite heavily in systematic, programme-wide summative assessment of knowledge (one statement) and ability (eight). So, too, with the other three engineering disciplines that contributed examples to the Report, where the different verbs in the *Ability to* statements (discuss, construct, use, make, recognise, carry out, write, appreciate, identify, assess, produce, choose, experiment, derive, test, plan, implement) call for differing approaches to assessment.

Plainly, departments could not warrant student achievement in respect of *Ability to* statements that were mainly subject to formative assessment. However, these formative assessment arrangements, combined with a careers/employability support programme, should enable students to lay powerful claims to achievement that they could substantiate with material drawn from the learning portfolios they would keep (this meshes with the QAA's recommendations on progress files). Where reliable summative assessments allow departments to warrant achievement, valid formative assessment helps students to lay claim to achievement.

## DIFFERENTIATED ASSESSMENT PLANS

Figure 1 sketches an approach to differentiated assessment that centres on the distinction between formative and summative purposes, relating them to warrants, student claims to learning and the process standards that lie behind both warrants and claims.

The programme assessment plan will also need to show that a range of assessment methods is used, differentiating between those most suited to the assessment of some learning and those best suited to the assessment of others.

A third form of differentiation will be between the amount of scaffolding to support assessment tasks in the first and final years.

The underlying point is simple and radical. The simple idea is that the assessment of complex learning outcomes, such as the EPC's output standard describes, demands a programme-wide approach (the same is true for teaching and learning arrangements as well).

It is a radical idea for at least three reasons: firstly, it breaks with a tradition of concentrating on modules and assuming that the programme will look after itself. Secondly, it suggests that teachers may find themselves being strongly encouraged to design teaching, learning and assessment sequences in order to help the programme. In this sense their pedagogical freedom is liable to be attenuated. Thirdly, the idea that some outcomes of learning should be assessed formatively can seem to be novel and challenging. If it is to work, it demands that programme teams put a lot of care into creating *knowing students*.

## PRACTICES AND PLANS

### Disseminating Examples of Good Practice: Assessment Toolkits

The 2001 survey of EPC members' assessment practices identified a lot of good methods and there is obviously value in disseminating them and brief descriptions of assessment methods in common use can be found elsewhere [18][34][35]. Although the survey was not designed to get detailed examples, some contributions showed that there are plenty to be collected. For example:

- Communication exercises: *Oral or written or visual presentations usually encountered in the context of other civil engineering activities and seen as valuable transferable skills* [output standard 1.2.1] … *Such exercises are time consuming for staff and students, especially marking of written work. Objectivity of marking is not easy to guarantee. We have attempted to produce a graded performance scale … by giving a clear description of the qualities one would expect to associate with any particular band of marks. In principle this can provide an opportunity for self/group/peer/staff criticism and be very positively formative.*
- Design project: *Students work in groups of 3 or 4 and are asked to indicate the distribution of effort among the group to aid eventual award of* [individual] *marks … the projects are very open ended, allowing students to apply a subset of the technical skills they have acquired over the previous three years. Assessment is through a preliminary written report, an oral presentation, a final written report and a poster presentation … grading criteria are provided … Each project has two supervisors and there are usually two assessors. This activity is time consuming and the assessment is time consuming* [but] *it counts heavily towards the final degree.*
- Interview: … *eliciting and clarifying clients' true needs* [output standard 1.2.2(a)] *might best be assessed by observing performance in a simulated interview; whereas the ability to produce detailed specifications of real target systems could be assessed in a written examination.*

By itself, disseminating examples of good practice will not be enough to align assessment regimes with the demands of the output standard. In part, this is because teachers want help to work out how to adapt good practice to their particular situations, but it is also because they are short of time, juggling multiple roles and operating in departmental and institutional environments that may not be conducive to fresh assessment practices. Anything that simplifies the burden of innovation will be a welcome contribution to the hard-pressed potential innovator, although a toolkit of assessment methods suited to the output standard is not enough.

## MAKING ASSESSMENT PLANS

It is suggested that departments start to make programme-level assessment plans by mapping what already happens. For example, the Assessment Working Group's Spring 2001 survey found that the eight most common assessment practices are: examinations, time-constrained (class) tests, project reports, presentations, laboratory reports, design studies, *vivas* or orals, and poster presentations. An assessment mapping exercise might consider each of the common assessment methods in turn and determine their effectiveness in measuring a student's achievement against each of the seven *Ability to* statements. This might lead to a 7x8 matrix as shown in Table 1.

For example, project reports might be effective in assessing set 4, a student's ability to use determinable models to obtain system specifications. This includes mathematical modelling, use of standard software platforms, sensitivity analysis, critical assessment of results and performance improvement. Design studies might be effective in assessing a student's ability across all of the *Ability to* statements.

It is now possible to identify assessment methods that are effective across a large range of the *Ability to* statements and to distinguish them from methods that are only effective for a small range of statements. In this way, the analysis may well identify redundant assessment methods. The analysis could also be extended to consider other criteria for determining effective assessment methods, eg cost and time demands.

Once information is obtained on existing practice, comparisons can be made between what *is* with what *ought to be*. Analysis might begin with the output standard for a programme of study and go on to consider how the student might be given the opportunities necessary to:

- Develop these abilities.
- Provide evidence of having achieved these abilities.

This leads to a top-down, systematic and systemic approach to both programme design and to an assessment strategy. The first bullet point (development of abilities) gets a programme team thinking about the modules that need to be in a programme and how programme learning outcomes will be distributed so as to support the output standard. The second bullet point (provision of evidence) leads the team to the identification of an assessment strategy, which operates across the full set of modules. This improves the chances of ensuring:

- That all of the *Ability to* statements are assessed.
- That none of them is over-assessed.

It is also likely to lead to a more uniform learning and assessment environment for the student, although it *may* require large changes in practice from the status quo and therefore meet resistance from hard-pressed academic staff.

Table 1: A template for mapping assessment methods against elements of the output standard.

| | • Key skills | • Systems to models | • Conceptual to determinable models | • Obtain system specifications | • Physical models | • Create real target systems | • Critically review performance |
|---|---|---|---|---|---|---|---|
| Exams | | | | | | | |
| Class tests | | | | | | | |
| Project reports | | | | | | | |
| Presentations | | | | | | | |
| Lab reports | | | | | | | |
| Design studies | | | | | | | |
| Vivas/orals | | | | | | | |
| Posters | | | | | | | |

The Skills *plus* project approach has been to use the analysis of how things could be to identify the most urgent points for attention in how things are [28]. Departments then try to *tune* their programmes by making small, feasible and powerful changes so that the programme, over several years, edges towards the ideal state.

## A MODEL FOR OTHER AREAS

As a non-engineer - a historian and then a social scientist – the first author is struck by how much teachers of engineering have contributed to thinking about good practice in higher education.

The output standard can be considered a remarkable enterprise. Its production has, it is believed, shown other subjects processes they should consider adopting. In particular:

- Authenticity of the standard; it describes what engineers do and is not afraid of complexity. This is surely what higher education is about: complexity, a mix of cognitive and non-cognitive achievements and emphasis on the world of practice.
- Recognition that if such complex achievements are to be assessed well, then a model of assessment that is itself complex and subtle is needed.
- Provision of a five-day summer workshop to help programme leaders to grapple with the implications for their programmes and practices of the output standard.

This seems to be an admirable model for others.

## REFERENCES

1. Engineering Professors Council (EPC), The EPC Engineering Graduate Output Standard. Interim Report of the EPC output standards project. London: Engineering Professors Council (2000).
2. Haug, G. and Tauch, C., Trends in Learning Structures in Higher Education (II). Follow-up Report prepared for the Salamanca and Prague Conferences (2001).
3. Assoc. of Graduate Recruiters, Skills for Graduates in the 21st Century. London: Assoc. of Graduate Recruiters (1995).
4. Harvey, L., Moon, S. and Geall, V., *Graduates' Work*. Birmingham: Centre for Research into Quality (1997).
5. Yorke, M., *The Skills of Graduates: a Small Enterprise Perspective.* In: O'Reilly, D., Cunningham, L. and Lester, S. (Eds), Developing the Capable Practitioner. London: Kogan Page, 174-183 (1999).
6. Chappell, C., Farrell, L., Scheeres, H. and Solomon, N., *The Organization of Identity: Four Cases*. In: Symes, C. and McIntyre, J. (Eds), Working Knowledge: the New Vocationalism and Higher Education. Buckingham: Society for Research in Higher Education and Open University Press, 135-152 (2000).
7. Norman, D., *Notes towards a Complex Theory of Learning.* In: Lesgold, A. et al (Eds), Cognitive Psychology and Instruction. New York: Plenum (1978).
8. Claxton, G., *Hare Brain, Tortoise Mind*. London: Fourth Estate (1998).
9. Goleman, D., *Working with Emotional Intelligence*. London: Bloomsbury (1998).
10. Hedlund, J. and Sternberg, R.J., *Too Many Intelligences? Integrating Social, Emotional, and Practical Intelligence*. In: Bar-On, R. and Parker, J. (Eds), The Handbook of Emotional Intelligence. San Francisco: Jossey-Bass (2000).
11. Dweck, C., *Self-theories: Their Role in Motivation, Personality and Development.* Philadelphia: Psychology Press (1999).
12. Cherniss, C., *Social and Emotional Competence in the Workplace*. In: Bar-On, R. and Parker, J. (Eds), The Handbook of Emotional Intelligence. San Francisco: Jossey-Bass, 433-458 (2000).
13. Sternberg, R.J. and Grigorenko, E.L. *Practical Intelligence and its Development*. In: Bar-On, R. and Parker, J. (Eds), The Handbook of Emotional Intelligence. San Francisco: Jossey-Bass, 215-243 (2000).
14. Rosenholtz, S., *Teachers' Workplace*. New York: Macmillan (1989).
15. Guile, D. and Young, M., Apprenticeship as a conceptual basis for a social theory of learning. *J. of Vocational Educ. and Training*, 50, **2**, 173-192 (1998).
16. Becher, T., *Professional Practices: Commitment and Capability in a Changing Environment*. New Brunswick: Transaction Publishers (1999).
17. Eraut, M., Non-formal learning and tacit knowledge in professional work. *British J. of Educational Psychology*, 70, 113-136 (2000).
18. Knight, P., *Being a Teacher in Higher Education*. Buckingham: Society for Research in Higher Education and Open University Press (2002).
19. Engeström, Y. (2001) Expansive learning at work: towards an activity theoretical reconceptualization. *J. of Educ. and Work*, 14, **1**, 133-156 (2001).
20. Brown, J.S. and Duguid, P., *The Social Life of Information*. Cambridge: Harvard University Press (2000).
21. Palloff, R.M. and Pratt, K., *Building Learning*

*Communities in Cyberspace*. San Francisco: Jossey-Bass (1999).
22. Laurillard, D., Stratfold, M., Luckin, R., Plowman, L. and Taylor, J., Affordances for learning in a non-linear narrative medium. *J. of Interactive Media in Educ.*, 2, 1-17 (2000).
23. Salmon, G., *E-moderation: the Key to Teaching and Learning Online*. London: Kogan Page (2000).
24. Ganesan, R., Edmonds, G. and Spector, M., *The Changing Nature of Instructional Design for Networked Learning*. In: Steeples, C. and Jones, C. (Eds), Networked Learning: Perspectives and Issues. London: Springer-Verlag, 93-110 (2002).
25. Goodyear, P., *Psychological Foundations for Networked L*earning. In: Steeples, C. and Jones, C. (Eds), Networked Learning: Perspectives and Issues. London: Springer-Verlag, 49-76 (2002).
26. Goodyear, P., *Environments for Lifelong Learning: Ergonomics, Architecture and Educational Design*. In: Spector, J.M. and Anderson, T. (Eds), Integrated and Holistic Perspectives on Learning, Instruction and Technology: Understanding Complexity. Dordrecht: Kluwer Academic Publishers, 1-18 (2000).
27. Knight, P.T., Complexity and curriculum: a process approach to curriculum-making. *Teaching in Higher Educ.*, 6, **2**, 371-383 (2001).
28. http://www.open.ac.uk/vqportal/Skills-Plus/home.htm
29. Entwistle, N., *Recent Research on Student Learning*. In: Tait, J. and Knight, P. (Eds), The Management of Independent Learning. London: Kogan Page, 97-112 (1996).
30. Campbell, D.T. and Russo, M.J., *Social Measurement*. Thousand Oaks: Sage Publications (2001).
31. Carter, R., A framework for the specification of degree programmes in engineering. Lancaster (unpublished paper) (2001).
32. Hamer, J., Assessing output standards. Paper presented at the Engineering Professors' Council, April (2001).
33. Knight, P.T., Assessment for Learning: Practices and Programmes. H850 file 3, Milton Keynes: The Open University (2002).
34. Hounsell, D., McCulloch, M. and Scott, M. (Eds), *The ASSHE Inventory.* Edinburgh: University of Edinburgh and Napier University (1996).
35. Gibbs, G., *Using Assessment Strategically to Change the Way Students Learn*. In: Brown, S. and Glasner, A. (Eds), Assessment Matters in Higher Education: Choosing and Using Diverse Approaches. Buckingham: SHRE and The Open University Press (1999).

## BIOGRAPHIES

Peter T. Knight works in the Open University advising on issues in the development of curriculum and associated assessment arrangements. He was previously at Lancaster University, where he taught Education as a social science subject from undergraduate to doctoral levels. He has a long-standing interest in assessment issues, on which he has published, and on related learning, teaching and design issues. He was a part of the Engineering Professors' Council's Assessment Working Group, contributing to its 2002 report *Assessment of Complex Outcomes*. He is also a member of HEFCE's Enhancing Student Employability Coordination Team.

Professor William M. Banks is currently the Chairman of the Engineering Professors' Council (EPC). He holds the Chair of Advanced Materials at the University of Strathclyde. His interest in Higher Education spans the spectrum from developments in teaching to interaction with Government as well as a keen interest in research. In this latter respect, he focuses on the load bearing properties and design of composite structures. He lectures internationally, serves on many national bodies and has published around 130 papers.