THE ASSESSMENT OF COMPLEX LEARNING OUTCOMES

A paper from the Engineering Professors' Council

Peter Knight [1]

Abstract *- The EPC's output standard is striking in many ways, not least in its decision to specify some complex, fuzzy and non-determinate outcomes of learning as defining characteristics of new engineering graduates. Programme leaders will be concerned about how to enhance what they do so that it has best aligns with these outcomes. Important though these design and delivery issues are, this paper approaches the implementation of the output standard from a different angle: what are the implications for assessment?*

The argument that follows is in four main sections. First, EPC survey data are used to show that engineers do use suitable assessment practices and that these practices are curbed by a number of difficulties that worry EPC members. Secondly, it is briefly claimed that our confidence in objective, reliable and accurate grading is misplaced. Third comes the suggestion that some of the difficulties identified in this analysis could be eased by taking a more differentiated view of assessment. The fourth section looks at the implications for assessment plans and practices, drawing on the report of the EPC's Assessment working group. It is noted in passing that what the EPC is doing could be a model for other subject areas which are also concerned to see complex outcomes emerge from stimulating learning.

Index Terms *- Complex learning, assessment in engineering degrees, formative assessment, summative assessment, differentiated assessment, programme assessment plans.*

The Argument

The authentic learning that the Engineering Professors' Council (EPC) values in its output standard [[11] and Appendix 1, below] takes time to develop. That directs curriculum designers' attention to *programme* issues. It also makes it necessary to re-examine the assessment of learning. Indeed, the success of the EPC's output standard may be closely related to the degree to which engineering teachers reject the assumption that assessment is measurement.

1. Introduction: the EPC's output standard

Description of the standard

The Engineering Professors Council (EPC) output standard comes out of consultation within Higher Education 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 [11]. Insofar as it is based on an analysis of what engineers *do*, it fits well with Haug and Tauch's [20] comment that 'enhanced employability seems to be the strongest source of change and reform in [European] higher education'.

These 'Ability to' statements are, however, 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 which different stakeholders can use to describe their desires or attainments, at whatever level may be of concern.

Complex learning

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

If we then explore one aspect of this sort of thinking, namely the view that HE should contribute to student employability, we find complex learning outcomes aplenty. A glance at some of the research on employability [1],[25] [30] shows that, amongst other things, higher education 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; team-working; managing others; critical analysis; being able to

[1] Centre for Outcomes-Based Education, the Open University, Milton Keynes, MK7 6HF, UK peter.knight@open.ac.uk

work under pressure; and imagination/creativity. 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 team-working; and understanding the sorts of identities that are valued in workplaces and appreciating how to take them on [6].

So far I have been trying to describe what it is that makes higher education engineering programmes complex. I can clarify the idea by identifying some things that I do not regard as complex. Complicated learning, for example, is different. It can be complicated to memorise procedures, formulae, sequences and plots, especially if we have to use several sources in the process. That is not complex because we can define the outcome 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 which are generally recognised to be the right answers and known procedures for getting them.

We do know that complex learning takes time — Norman [25] says up to 5000 learning hours. Even if we dispute Norman's figure, 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 routinized (how long does it take to learn how to do chi-squared 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 characterizes higher education and which suffuses the output standard, is 'fuzzy' learning but it has not been widely researched. Even without research, though, we can quickly see that there are profound issues to address and we can make informed guesses about how best to do so.

Programme design issues

The main design issue is that slow learning means programme-level not module-level thinking. Three less obvious design issues are:

• The need to stimulate practical intelligence

Employability, and the output standard more generally, might be construed as a mix of emotional (Goleman, 1998) and practical intelligences (Hedlund and Sternberg, 2000). 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 [7],[9] .

Significantly, there is also evidence that these 'intelligences' scarcely correlate with academic intelligence [29]. 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. Furthermore, the research into the transfer of learning from one context and time to others implies 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. This issue will be familiar to many engineers, especially those taken with the suggestions that George King was making some twenty years ago. However, it has not always been systematically addressed in the design of engineering programmes. For example, fresh learning, teaching, assessment and curriculum strategies 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 [27],[17],[2],[13],[22]. Two conclusions are:

1. 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 and it may be too decontexted, even artificial, to be much use in the workplace.
2. By definition, we cannot pre-specify the course and outcomes of non-formal learning engagements. If we take non-formal learning seriously, we need to develop new approaches to curriculum, to move from programme design and determinate learning outcomes to the design of learning environments rich in opportunities for complex learning. One of the most significant features of learning environments is the nature of the workgroups and communities in which learning happens [10],[3]. We need to think about 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 [26], [24],[28] it is all the more important to think carefully about the ways in which on-line learners might become vibrant learning communities.

• Learning from work done on the design of environments for on-line learning

To echo the previous point: scholars interested in effective on-line learning [14], [15] are concerned with the design of whole learning environments that encourage complex achievements. This literature has produced metaphors [15] and principles [23] that can inform the design of whole learning environments that should favour the

emergence of complex outcomes of learning, such as those captured by the output standard. The Skills *plus* project[http://www.open.ac.uk/vqportal/Skills-Plus/home.htm] 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.

2. Assessing the output standard

> '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' [12].

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 use 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 methods 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' .

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 and the one after it looks at 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:

- The prime contributor was the semester system. No-one had anything good to say about it. Complaints were 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 we can make progress with devising assessment arrangements that are fit for the purposes implicit in the output standard, we need to be sure that we understand social measurement theory [3], that we understand what can be assessed, how and with what certainty. 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 [5] 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.'

This tends to produce the conclusions (a) that there must be objective and reliable measures of the requirements or specification and (b) that any assessment procedure which falls short is therefore defective and a waste of time and effort. Let us leave to one side 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.

TABLE 1
Problems with high-stakes assessments of complex achievements

1. *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, generalizable 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 with no explanatory powers.

2. *The limits of reliability :* Plainly, fictional objects of assessment cannot be assessed 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 (self-motivation, for example).

3. *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.

4. *The transferability of achievement:* The achievements that grades 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 warrants describe achievements that the learner can readily and independently transfer to fresh settings.

5. *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.

6. *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. First, things covered by high stakes assessment get serious attention, others don't. 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.

7. *The misuse of number:* Summative assessment data are usually presented numerically but they really ought not to be treated 'numerically'. Beware of the numbers created by summative assessment and mistrust conclusions based on the transformation or manipulation of those numbers.

8. *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.

9. *Process-blindness:* Scores and grades are silent about the learning processes involved. This matters because if you tell me that someone has repeatedly shown that they can solve problems and I find that problem-solving has been taught and learned as the manipulation of numbers according to learned algorithms, I may be less impressed than if I hear that it has been developed through engagement with a series of 'fuzzy', authentic tasks.

10. *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.

(From Knight, P. T., 2002, *H850 file 3, Assessment for learning: practices and programmes*. Milton Keynes: The Open University.)

As Hamer [18] puts it:

'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'.

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 output standard says that successful students should have 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 the validity or worth of reliable tasks that assess achievements that are so simplified and detached that it is hard to know what value there is in the information they produce. 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 contained in Figure 1 should give pause for thought because they suggest that reliable assessment is something of a chimera.

The Engineering Professors Council'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 be neither as useful nor as meaningful as is sometimes assumed. 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 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 clearly recognises that complex learning depends upon information, recall, command of algorithms, and such like. If we wish to stimulate complex learning outcomes in higher education in general and in engineering education in particular, then we need to look away from the 'assessment-as-measurement' paradigm as we search for ideas to help us to create assessment systems that are fit for the output standard.

Figure 2 suggests that we make more use 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. This point is worth emphasising because student would be *required* to take formative assessment seriously. It is not seen as something that keen students can opt into and others can cut.
- Feedback should then be fast, focused, relevant to the assessment criteria, 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 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.
- Students should therefore have the programme criteria from the first, regularly use them, share them, and practise applying them. They would have a clear interest in using them because the criteria would be the points of reference against which their work would be summatively judged. In this sense, formative use of the criteria prepares the way for successful performance on summative tasks that will be judged with reference to elements of the same set of criteria.
- 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 to 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 [11]. 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. Students, though, 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 [11 pp. 15-25], 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 which 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 2 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.)

Good curriculum architecture: coherence between physical environment, learning culture and processes

- all likely to stimulate the learning described in the programme specification.

Tasks

Good, often 'fuzzy', indicators

help students to do tasks and help the feedback and judgement processes

Formative feedback on achievement ---------- Summative Judgements

(Indicators need to be less fuzzy; invest in training multiple markers to use them reliably

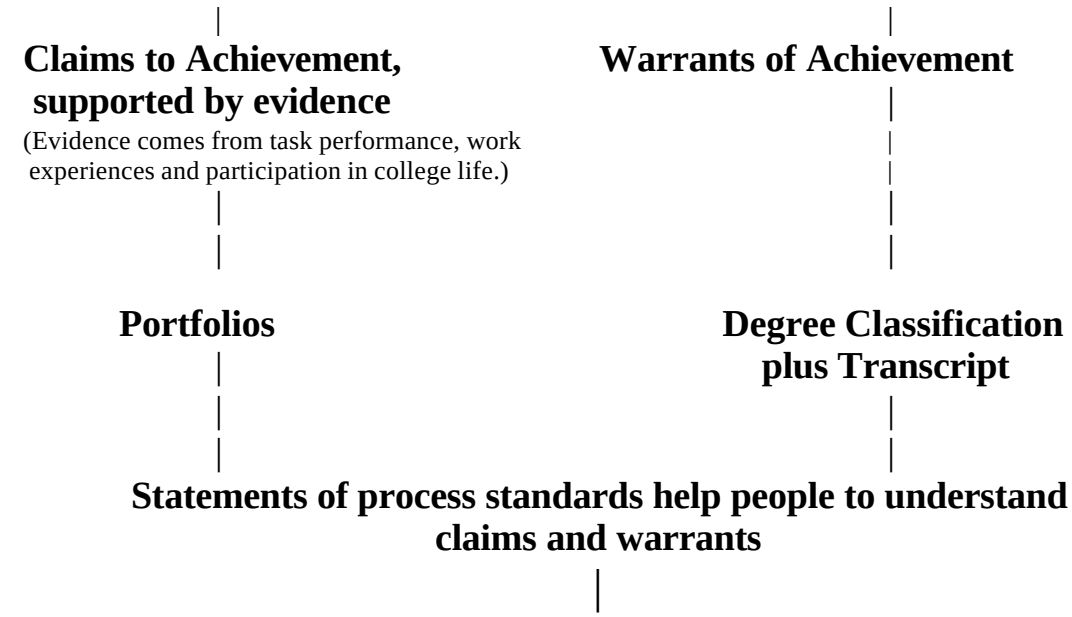

FIGURE 2.
Differentiated Programme-level Assessment Arrangements

It is a radical idea for at least three reasons: first, 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. attenuated.

Thirdly, the idea that some outcomes of learning In this sense their pedagogical freedom is liable to be 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'.

3. 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. (See [21],[22] for brief descriptions of assessment methods in common use.) 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

I suggest that departments start to make programme-level assessment plans by mapping what already happens.

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, lab 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.

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, critically assess results and improve performance.) 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 which are effective across a large range of the 'Ability to' statements and to distinguish them from methods which 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 e.g. cost and time demands.

Once we have information on existing practice we can begin to compare what *is*, with what *ought to be*. We 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 (a) that all of the 'Ability to' statements are assessed and (b) that none of them is over-assessed. It is also likely to lead to a more uniform learning and assessment environment for the student – but it *may* require large changes in practice from the status quo and therefore meet resistance from hard-pressed academic staff.

The Skills *plus* project (http://www.open.ac.uk /vqportal / Skills-Plus/ home.htm) approach has been to use the analysis of how things could be in order to identify the most urgent points for attention in how things are. 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.

4. A model for other areas

As a non-engineer - a historian and then a social scientist - I am struck by how much teachers of engineering have contributed to thinking about good practice in higher education. I find the output standard a remarkable enterprise. Its production has, I believe, shown other subjects processes they should consider adopting, I particularly like:

- The 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.
- The recognition that if we are to assess such complex achievements well, then we need a model of assessment that is itself complex and subtle.
- The 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 me to be an admirable model for others.

References

1] Association of Graduate Recruiters (1995) *Skills for Graduates in the 21st* Century. London: Association of Graduate Recruiters.

2] Becher, T. (1999) *Professional Practices: commitment and capability in a changing environment* New Brunswick, NJ: Transaction Publishers.

3] Brown, J. S. and Duguid, P. (2000) *The Social Life of Information*. Cambridge MA: Harvard University Press.

4] Campbell, D. and Russo, (2001) *Social Measurement* Thousand Oaks, CA: Sage Publications.

5] Carter, R. (2001) A framework for the specification of degree programmes in engineering. Lancaster: unpublished paper.

6] Chappell, C., Farrell, L., Scheeres, H. and Solomon, N. (2000) The organization of identity: four cases, in: C. Symes and J. McIntyre, (Eds.) *Working Knowledge: the new vocationalism and higher education*, Buckingham: Society for Research in Higher Education and Open University Press, 135-152.

7] Cherniss, C. (2000) Social and emotional competence in the workplace, In: R. Bar-On, and J. Parker (Eds.) *The Handbook of Emotional Intelligence*. San Francisco: Jossey-Bass, 433-458.

8] Claxton, G. (1998), *Hare Brain, Tortoise Mind*. London: Fourth Estate.

9] Dweck, C. (1999) *Self-theories: their role in motivation, personality and development*. Philadelphia: Psychology Press.
10] Engeström, Y. (2001) Expansive learning at work: towards an activity theoretical reconceptualization, *Journal of Education and Work*, 14(1), 133-156.
11] Engineering Professors Council (2000) *The EPC Engineering Graduate Output Standard - the interim report of the EPC output standards project*. London: Engineering Professors Council.
12] Entwistle, N. (1996) Recent research on student learning in: J. Tait and P. Knight (Eds.) *The Management of Independent Learning*. London: Kogan Page, 97-112.
13] Eraut, M. (2000) Non-formal learning and tacit knowledge in professional work, *British Journal of Educational Psychology*, 70, 113-136.
14] Ganesan, R., Edmonds, G. and Spector, M. (2002) The changing nature of instructional design for networked learning, in: C. Steeples and C. Jones (Eds.) *Networked Learning: perspectives and issues*. London: Springer-Verlag, pp. 93-110.
15] Goodyear, P. (2000). Environments for lifelong learning: ergonomics, architecture and educational design. In J. M. Spector & T. Anderson (Eds.), *Integrated and Holistic Perspectives on Learning, Instruction & Technology*: *understanding complexity*. Dordrecht: Kluwer Academic Publishers, 1-18.
16] Goodyear, P. (2002) Psychological foundations for networked learning, in: C. Steeples and C. Jones (Eds.) *Networked Learning: perspectives and issues*. London: Springer-Verlag, pp. 49-76.
17] Guile, D. and Young, M. (1998) Apprenticeship as a conceptual basis for a social theory of learning, *Journal of Vocational Education and Training* 50(2) 173-192.
18] Hamer, J. (2001) Assessing output standards. Paper presented to the EPC Congress April 2001.
19] Harvey, L. *et al.* (1997) *Graduates' Work*. Birmingham: Centre for Research into Quality.
20] Haug,G. and Tauch,C. (2001) *Trends in Learning Structures in Higher Education (II). Follow-up Report prepared for the Salamanca and prague Conferences*
21] Hounsell, D., McCulloch, M. and Scott, M. (Eds.) (1996) *The ASSHE Inventory* (Edinburgh: University of Edinburgh and Napier University).
22] Knight, P. (2002) *Being a Teacher in Higher Education*. Buckingham: Society for Research in Higher Education and Open University Press.
23] Knight, P. T. (2001) Complexity and curriculum: a process approach to curriculum-making, *Teaching in Higher Education* 6(2), 371-383.
24] Laurillard, D., Stratfold, M., Luckin, R., Plowman, L and Taylor, J. (2000) Affordances for learning in a non-linear narrative medium, *Journal of Interactive Media in Education*, 2, 1-17.
25] Norman, D. (1978) Notes towards a complex theory of learning, in: A. Lesgold et al., (eds.) *Cognitive Psychology and Instruction*. New York: Plenum.
26] Palloff, R. M. and Pratt, K. (1999) *Building Learning Communities in Cyberspace*. San Francisco: Jossey-Bass.
27] Rosenholtz, S. (1989) *Teachers' Workplace*. New York: Macmillan.
28] Salmon, G. (2000) *E-moderation. The key to teaching and learning online*. London: Kogan Page.
29] Sternberg, R. J. and Grigorenko, E. L. (2000) Practical intelligence and its development, In: R. Bar-On, and J. 30]Parker (Eds.) (2000) *The Handbook of Emotional Intelligence*. San Francisco: Jossey-Bass, 215-243.
30]Yorke, M. (1999) 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

Appendix 1. The Engineering Professors' Council Engineering Graduate Output Standard

The Standard comprises the following Generic *Ability to..* Statements. Further information about its application and benchmarking to illustrate level can be referenced in *EPC Occasional paper Number 10 - The Interim Report of the EPC Output Standards Project*, at the EPC website **engprofc.ac.uk**

1. Ability to exercise Key Skills in the completion of engineering-related tasks at a level implied by the benchmarks associated with the following statements. Key Skills for Engineering are Communication, IT, Application of Number, Working with Others, Problem Solving, Improving own Learning and Performance.

2. Ability to transform *Existing Systems* into *Conceptual Models.*

Ability to:

2.1. Elicit and clarify client's true needs.
2.2. Identify, classify and describe *Engineering Systems.*
2.3. Define *Real Target Systems* in terms of *objective functions*, performance specifications and other constraints (i.e. define the problem).
2.4. Take account of risk assessment, and social and environmental impacts, in the setting of constraints (including legal, and health and safety issues).
2.5. Select, review and experiment with existing *Engineering Systems* in order to obtain a database of knowledge and understanding that will contribute to the creation of specific *Real Target Systems.*
2.6. Resolve difficulties created by imperfect and incomplete information.
2.7. Derive conceptual models of *Real Target Systems*, identifying the key parameters.

3. Ability to transform *Conceptual Models* into *Determinable Models*

Ability to:

4.1. Construct *Determinable Models* over a range of complexity to suit a range of *Conceptual Models*.
4.2. Use mathematics and computing skills to create *Determinable Models* by deriving appropriate constitutive equations and specifying appropriate boundary conditions.
4.3. Use industry standard software tools and platforms to set up *Determinable Models.*

4.4. Recognise the value of *Determinable Models* of different complexity and the limitations of their application.

4. Ability to use *Determinable Models* to obtain system *Specifications* in terms of parametric values

Ability to:

5.1. Use mathematics and computing skills to manipulate and solve *Determinable Models*. Use data sheets in an appropriate way to supplement solutions.
5.2. Use industry standard software platforms and tools to solve *Determinable Models*.
5.3. Carry out a parametric sensitivity analysis.
5.4. Critically assess results and, if inadequate or invalid, improve knowledge database by further reference to existing systems, and/or improve performance of *Determinable Models.*

5. Ability to select optimum *Specifications* and create *Physical Models*

Ability to:

7.1. Use *objective functions* and constraints to identify optimum specifications.
7.2. Plan *Physical Modelling* studies, based on *Determinable Modelling*, in order to produce critical information.
7.3. Test and collate results, feeding these back into *Determinable Models.*

6. Ability to apply the results from *Physical Models* to create *Real Target Systems*

Ability to:

9.1. Write sufficiently detailed specifications of *Real Target Systems*, including risk assessments and impact statements.
9.2. Select production methods and write method statements.
9.3. Implement production and deliver products fit for purpose, in a timely and efficient manner.
9.4. Operate within relevant legislative frameworks.

7. Ability to critically review *Real Target Systems* **and personal performance**

Ability to:

7.1 Test and evaluate real systems in service against specification and client needs.
7.2 Recognise and make critical judgements about related environmental, social, ethical and professional issues.
7.3 Identify professional, technical and personal development needs and undertake training and independent research.