

ASSESSMENT OF WORK-BASED REPORTS: AN ANALYSIS OF ASSESSMENT FRAMEWORKS

John Monk

j.monk@open.ac.uk

Faculty of Technology, Open University, UK

Abstract—In Britain engineering professional development has traditionally been seen as a three phase process consisting of a period of engineering formation, a period of training and a period during which engineering responsibilities are demonstrated. An individual could submit evidence of these activities and become registered as a Professional Engineer. Increasing numbers of people employed in the role of engineer do not have formal engineering qualifications and a part or all their engineering formation is carried out within engineering companies or organizations. These people therefore do not have the academically authenticated credentials to register as professional engineers but if they are ignored then the pool of registered engineers will cease to be representative of the profession. The Engineering Council, the body responsible for registering engineers in the UK, has acknowledged the changes in the structure of the profession and has introduced an alternative route for assessing the knowledge and understanding that underpins the competence of a professional engineer. Individual engineers can demonstrate that they have an adequate engineering formation through any combination of academic qualifications and a technical report on some aspect of their professional engineering work. The introduction of the technical report requires the Professional Engineering Bodies to carry out an assessment outside the traditional assessment framework of the Universities. This paper reviews and analyses the requirements of assessment systems and derives the components of such a system that will ensure that the results of the assessment of a work-based technical report will be respected and be seen as assuring comparable standards to the academic routes to engineering formation. By examining assessment separately from the processes of teaching and learning, the paper also reveals the extent of an assessment process and its costs.

Key Words—Assessment, Work-based learning, Engineering

INTRODUCTION

Commonly, aspiring engineers are assessed on their knowledge and understanding during the course of their formal education. Educational institutions provide a succession of assessment exercises and certify their students' achievements. Employers and other groups use the evidence of certification as an indicator of the knowledge and understanding of individuals.

Knowledge and understanding are not enough, and it is not widely expected that new graduates will be fully competent as engineers; some time is likely to be spent in initial professional development before graduates can be given the full responsibilities as professional engineers. Once graduates have demonstrated their engineering competence, they can be registered as Professional Engineers by the Engineering Council in the UK (ec^{uk}). Registration is useful for employers who are recruiting or promoting staff since it attests to the competence of an individual; registration of employees is also helpful when an engineering enterprise itself is faced with demonstrating its competency to shareholders, the law, politicians or the public.

The assessment of the engineering competence of individuals is carried out by Professional Engineering Institutions (PEIs), such as the Institution of Electrical Engineers (IEE) and the British Computer Society (BCS) who

act under license on the ec^{uk}'s behalf. The ec^{uk} issues a set of regulations for licensees and audits the licensees' processes and procedures. The regulations are occasionally reviewed and a major revision was issued in 2004. One effect of the new and more demanding regulations was that fewer people would have the more stringent benchmark qualifications. The regulations, though, recognised that

Formal education is the usual, though not the only, way of demonstrating the underpinning knowledge and understanding for professional competence. [1]

Within this statement is an acknowledgement that many people are practising with high levels of competence without the traditional formal qualifications expected of an engineer. In software engineering, for example, there are many individuals with unusual formal qualifications who demonstrate high levels of engineering competence in their work.

To cater for those who had developed their engineering knowledge and understanding outside of a formal education system and thus lacked formal endorsement of their knowledge and understanding, an alternative was specified. The assumption is that some experienced but formally unqualified people working as engineers have had adequate learning opportunities in the course of their work but they lack

credentials and a procedure for the assessment of their knowledge and understanding.

WORK-BASED LEARNING

Frequently, the presumption is that engineers begin their careers by taking a degree. Engineers Australia announces:

The benchmark Stage 1 qualification for professional engineers is the four-year Bachelor of Engineering degree [2]

And immediately under the heading “So, you want to be an engineer” on the American Society for Engineering Education Engineering K12 Center web site comes the heading “Getting a degree in engineering” [3]. A study however discovered that amongst engineering practitioners in the UK

... the number of people with engineering qualifications who do not meet those required by the Engineering Council/Engineering Institutions ... are estimated to be about 1,000,000, there being an estimated further 500,000 people practising engineering with no qualifications[4]

On the one hand this could be a cause for concern, but on the other presumably a proportion of these people are competent engineers and this competence could be formally recognised if it could be assessed. While this cohort of competent engineers is not recognised, the register of engineers is not representative of the profession, and a swathe of competent practitioners is formally indistinguishable from those who would fail a competency test. It is in the interest of the profession, therefore, to find ways of assessing and registering competent engineers who have not followed a conventional route into the profession.

A primary source of evidence for the engineering knowledge and understanding that such people have is in their work. But, the principal goal of a work task is rarely to promote learning or to assess it. Nevertheless it is to be expected that work environments should offer rich opportunities for learning since research and theories of learning are now presenting learning as

a process of knowledge construction (rather than of knowledge reproduction) [5].

so that a learner’s achievement is related primarily to engagement [6], which involves working on tasks that promote interaction within a community of practice [7]. The theory of situated cognition, for example, sets learning

within a community of practice in which knowledge is ...created, sustained and changed through interaction [8]

Learning can then be seen to be a “cognitive apprenticeship” in which learners employ authentic tools in authentic activities. Theories of cognitive development, similarly, stress the importance of experience and focus on

the qualitative and stable changes that occur in thinking over time and through experience [8]

The theory notes that the environment can promote or hinder learning, so for those assessing learning in the workplace, a first step might be to find out about the candidate’s work

environment to gain an assurance that the candidate’s situation has not inhibited learning.

Professional engineers in engineering enterprises often work on a succession of projects — for example, a product design, a process improvement, a testing regime, the commissioning of plant or a fault diagnosis — and it is reasonable to assume that projects will have provided substantial opportunities for engineers to gain knowledge and understanding that underpin their engineering tasks. Learning through project related activity in formal education has a seventy year old pedigree. Kilpatrick described a “project method” of instruction in 1918[9] and Albery in 1927 described this project method as one that secured learning as a by-product of activities that aimed at a concrete result or accomplishment.[10] Learning is thus instrumental in working towards a project’s goal and the outcomes of projects can provide a vehicle for assessment.

Prospective engineers can, therefore, develop their competences in activities and projects governed by the objectives of company or organisation where they work. In some instances the projects that they engage in will be formally linked to professional development plans, but in others their development will be an outcome of an unstructured career, which nevertheless can provide comprehensive opportunities for learning and experience of projects that can set a bounded context for assessment.

METHODS OF ASSESSMENT

In the ec^{uk} scheme, the PEI’s task is to evaluate the outcome of the candidate’s personal professional development. This process separates learning from assessment unlike formal education where assessment is frequently designed to support and guide learning. The usual conventions must be inverted and assessment must be devised after learning has taken place so that it draws on what the candidate practised in the work environment. As usual there is a need for

good alignment between the practice and the assessment [8]

and this requires the assessors to be or become familiar with the context of the candidate’s work.

But, the assessment must also align with the competences that are expected of an engineer and it is conceivable the candidate’s experience renders this alignment impossible. As a first step a PEI must check that a candidate does have relevant engineering experience and that he or she has had the opportunity to gain the knowledge and understanding that underpins engineering competence

There is a spectrum of assessment methods that might be employed to provide a demonstration of knowledge and understanding ranging from exams through to practical tests. Competence, for example, can be assessed by observing people conducting a task, but the failing of such a method is that it offers no warrant for future performances. To provide assurances about his or her likely performance in new circumstance, a candidate will have to demonstrate an understanding of what he or she has rehearsed. Understanding has been described as

having a grasp of the structure of a discipline, seeing how things are related, using the ideas in novel situations, and evaluating even challenging, the knowledge claims embedded in the discipline [11]

and a demonstration of that understanding requires the candidates to explain and rationalise their actions and to offer some reflections on their experience, and how things might have been done differently. Thus, for the candidate, the assessment involves more than gaining experience.

The ec^{uk} in its assessment scheme therefore requires candidates first to demonstrate that they have suitable experience and then asks them to write a technical report based on that experience. To place a reasonable bound the assessment task candidates select an area of their experience and prepare a report, or select pre-existing reports and add a commentary. In this way the scheme is similar to assessment using portfolios in which the candidates make a selection from the products of their work. In a conventional educational setting the assessors sets specific tasks to match the assessment criteria. In contrast in an assessment based on portfolios, the candidate makes the selection but this leads to the criticism that there is frequently a poor link between the expressed competencies and the criteria for assessment. [12]

Assessment of work-based learning is also similar in style to what is known as “*alternative assessment*” and critics warn that such assessments privileges the candidate’s “*conceptualizations of their experiences*” and while this makes it suitable for supporting learning, its weaknesses are, with the candidate in control, that the exercise may not cover understandings that are relevant to the assessment. A second observation is that the assessors may not be entirely familiar with the elements of the candidate’s field of work and consequently the consistency of judgement amongst assessors is poor. [13] Similarly when compared to other methods work-based learning is reported to generate great variability unless supervisors and mentors are trained in the use of criteria of assessment. Both reflective practice and project work are also said to introduce variability in assessment which is again reduced by the use of criteria. Aside from these difficulties, work-based learning provides for all-round testing and can provide high “*learning gains*”. [14] Overall then the ec^{uk} technical report based on learning in the workplace is a viable form of assessment, but the technique is only satisfactory when candidates have had relevant experience and attention needs to be paid to the consistency of judgement. In the ec^{uk} scheme the candidate has to choose the experience that is to be reported, but to prevent wasted effort it is worthwhile for the candidate and the assessing Institution to negotiate over which experiences and which aspects of experience should be the focus of assessment.

STAGES

There are a number of participants in any assessment process and a number of tasks to be completed. There are four basic steps

- Set and disseminate the criteria for assessment
- Candidates complete assignments which generate or collate the evidence for their assessment

- Collect the evidence relating to the work that individual candidates have performed.
- Judge the standard implied by the evidence against the criteria
- Decide upon the credentials to be issued and publish them giving the identity of the candidate and his or her level of achievement

Academic assessments are rarely completed in a single stage. Assessments are broken up into stages with the aim of

- Covering different criteria with different assessment components
- Allowing for variations in the candidates performance
- Taking remedial action early
- Assessing variations in the candidate’s performance
- Pacing learning
- Reducing the possibility of certain kinds of fraud or error
- Assessing different degrees of detail
- Allowing different pools of expertise to be employed in doing the assessment

Staged assessment introduces new sets of criteria for

- Making judgements about progression to the next stage
- Judging how the candidate’s performances are to be aggregated and recorded in the credentials that are issued.

The ec^{uk} assessment procedure is divided into three stages; first a regulation requires a check to be sure that candidates’ career histories indicate “*they have sufficient engineering experience*” [15]. The review of the career history also helps to determine the size and broad content of the subsequent assessment tasks. A candidate, for example, with a significant amount of reliably certified learning would be asked to provide less evidence of work-based learning than a candidate with no certified learning; or certified learning in one sphere would shift the emphasis in the later assessment to other areas.

Candidates whose knowledge and understanding has developed in an engineering workplace inevitably work as members of teams and much of their understanding and knowledge will come from their observations of others in the team rather than their own direct experience. Thus their accounts of projects will only be, in part, an account of their personal experience. Nevertheless such accounts can show an understanding of processes and roles without providing evidence those candidates are fully competent in exercising those roles or participating in those processes. Judgements must therefore give credit for observation as well as reasoned accounts of participation, but only insofar as reasoned observations demonstrate understanding. Satisfactory explanations alone are not evidence of competence; ultimately evidence of participation is required.

The assessment is thus divided into two further parts involving, first, evidence of the ability to provide satisfactory explanations for action and, secondly, evidence of participation in the action. The Engineering Council regulations recognise these two facets and allow for the development and assessment of knowledge and understanding that underpins competence followed by the development and assessment of competence. In some instances, where development has taken place entirely in the workplace, learning and practice are inseparable nonetheless a thorough assessment of knowledge and

understanding is completed before proceeding with the final assessment of competence.

The second stage therefore requires the assessment of the technical report by informed assessors who gauge the degree of knowledge and understanding that the candidate can demonstrate. This stage includes an interview of the candidate by the assessors who explore the technical report and record their judgement on the candidate's performance. If the report or the interview is unsatisfactory the candidate might be asked to revise the report or add to it, or the recommendation may be to reject the candidate's application or to produce a technical report on a different topic.

Individual PEIs have added steps to the first stage to help the candidate produce a suitable report. The candidate is commonly asked to produce a synopsis of his or her proposed report and will receive feedback after the synopsis has been informally examined. Once the synopsis is approved, the candidate is expected to write and submit the report

A satisfactory result in the second stage allows a candidate, after working as a practitioner at a professional level, to participate in the third stage which involves the examination of all the evidence about a candidate's career and an additional interview by interviewers who are trained in competence assessment. The result is a record of the interviewers' views on the competence of the candidate.

TYPES OF ASSESSMENT

Assessment systems can be loosely divided into formative assessments which are designed to give feedback and help students to learn and summative assignments that give a measure of a student's performance and that often provide backing for credentials certifying the student's learning. Since it is assumed in the PEIs' assessment process that learning has already taken place, then, unless the candidate fails, the assessment is primarily about supplying credentials and is summative. [16]

Within the formative and summative categories there are normative assessments, which judge the student's overall performance in relation to statistics collected about a student population, and criterion-based assessments that judge the student against a set of criteria and that often provide backing for credentials certifying the candidate's learning. Normative assessment is relevant where a sub-group of a population is to be selected to satisfy, for example, a capacity constraint but the PEIs welcome new members and are not especially constrained. The PEIs aim is to judge whether or not candidates satisfy the requirements to be a member of the profession. Criterion referenced assessment is therefore more appropriate.

Besides, the projects that are being assessed in the ec^{uk} scheme are quite varied and this

poses problems with a standard norm-referenced assessment strategy because like is not being compared with like [17]

Fortunately, in appraisals of all kinds, including the evaluation of projects, there is strong support in the literature for criterion-referenced assessment.

For summative assessment the overall criterion requires candidates to demonstrate a skill in creating material or

performance that satisfies the authorities who stand as the guardians of standards. But candidates will require more specific guidance. And the public may require evidence of a rationale for judgements beyond an appeal to the opinions of experts.

Criteria can be conveyed orally, made explicit in writing or be inferred from the archives of candidates' work and the judgements made by the assessors. The research literature suggests that explicit criteria are to be preferred and it is reported that:

Improvements in project assessment can take place if clear guidelines and explicit criteria are provided [17]

firstly because it helps to offer consistency of assessment; secondly the criteria help candidates to understand the nature of the assessment task [18], thus the use of criteria is endorsed by candidates as 'a well-conceived assessment tool' [19] that provides 'guidance as to what is valued' [20]. Thirdly

For the external stakeholders such as employers or outside evaluators, assessment criteria may be used to gain a more detailed sense of the academic rigour of a particular programme [20]

The criteria for assessment, therefore, direct assessors in making their judgements, guide candidates in assembling their evidence and inform the public when questioning the relevance and standing of awards.

GRADING

It is common in criterion based assessments to attach a grading scale to each criterion in order to differentiate between the performances of candidates. Frequently a grid is constructed listing aspects of the assignment in one dimension and grades in the other with entries comprised of statements of the individual criteria for awarding the grade. For example the aspect might be "presentation" and the criteria for the different grades would extend from "Disorganised/incoherent" through criteria such as "Carefully and logically organised" which would warrant a higher grade. [19] Applied to engineering such a scheme would identify the aspects of professional engineering work that are valued and give implicitly an indication of what is considered a good standard of work.

In the assessment of professional engineers such an elaborate scheme may not be required since judgements are being made about whether someone meets the criterion or not. However it could be argued that engineers in different settings require different degrees of understanding. Production engineers, for example, might have a more sophisticated understanding of statistics than software engineers who would have greater understanding of programming languages. Thus a lower standard may be acceptable in one context because the significance of the criterion is lower than in other contexts. Thus every criterion might have a scale associated with it and an assessment involves associating an individual's performance with a grade for each criterion and then aggregating the results in some way to come to a conclusion. In this way a candidate might be compensated for a lower performance surrounding one criterion by a performance judged as good against another criterion.

Simpler schemes employ numeric grades, a numeric weighting for each criterion that is graded and a linear relationship between a total and the grades awarded to criteria. Sometimes combinations of criteria may be more important than criteria in isolation and this requires a more elaborate scheme. Often criteria represent facets of individual performance that are incommensurable and defy rational aggregation. The aggregation of performance measures is thus problematic.

Frequently in the simpler schemes, grading scales represent a reformulation of normative criteria. For example, the generic and vague criterion “*depth of understanding*” accompanied by a scale that ranges from “*no understanding*” to “*exceptional*” [17] does not offer a great deal of help to candidates since it does not tell them specifically what has to be in their report. And terms such as “*average*” or “*exceptional*” are normative expressions that camouflage normative assessment as criterion-based.

ENTRY

Crucial judgements are made in permitting people to become candidates; in schemes with several stages judgements are made about how the candidates evidence matches the criteria; there are judgements about whether or not a candidate should progress to a following stage and judgements about the candidates overall performance. These judgements give rise to the need for criteria for entry to the first component of assessment, plus criteria for progression and criteria for the final award. In the ec^{uk} scheme the entry criterion requires candidates to have been engaged in engineering projects at a suitable level with opportunities for learning, progression relies on meeting the assessment criteria at each stage, and the final award demands comprehensive evidence of competence.

Ritter reports that entry criteria to courses (and hence to the assessment system) are becoming extremely varied and that amongst other things leads to certain desirable aspects of assessment that include broad and open-ended assignments that permit candidates to select their own written formats which cater for the variety of objectives that the candidates have in their work. [21] The EC process too is aimed at assessing people with widely differing backgrounds. The criteria for judging the technical report therefore need to stress, not the form or process of presentation, but its effect.

THE TROUBLE WITH CRITERIA

Candidates in any assessment have to provide evidence. In preparing evidence they will look for the guidance that criteria provide. Criteria for assessing projects tend to be a mixture of candidate characteristics (such as “*comprehension*”, “*initiative*” or “*ability to problem solve*”), broad references to content and its qualities (such as “*background information*”, “*literature review*”, “*organisation*” or “*clarity*”) and references to the process of conducting project work (such as “*planning*”) and do not provide specific or direct guidance on what a candidate might write on the page.

Some criteria do focus on the mode of expression. Take for example, the criterion that candidates should

include an evaluative perspective on their own work and the work of others [20]

This criterion indicates an attitude to be taken towards project work but it does not specify in any way details of the subject that is to be reported on. It also begs the question, “What is an evaluative perspective?”. The explicit expression of criteria proves to be problematical. For example,

formulations are of necessity couched in the epistemology of the discourse which students are learning [20]

However the ec^{uk} test is applied after the candidate has completed the necessary learning. The candidates should therefore be able to follow criteria that elaborate on the required knowledge and understanding. Unfortunately criteria are often expressed using an educational jargon which may baffle candidates who have had little contact with formal education and

research has shown that students do not have the same understandings as their tutors [22]

The skills brought to bear on an engineering project by a professional engineer are many and varied hence assessment of projects and the specification of criteria have their own peculiar difficulties, for example, completing a project involves

a wide range of subtle and interacting skills ... which can be difficult to identify and isolate [17]

Criteria that are aimed at the fair treatment of candidates, in principle, have to cover every conceivable valid expression and text that candidates might produce. Detailing the whole gamut of knowledge and understanding and their valid modes of expression is usually impractical because of the sheer amount of detail, especially where the scope is defined by the attributes of a wide ranging profession. Inevitably, criteria are expressed in generic terms that have to be translated into the specific context that each candidate has been immersed in. This translation is a task for the candidates and the assessors who, without any contact, may reach different conclusions. Assessment is thus seen as an exercise that includes the interpretation of criteria and reaching agreement on their interpretation and not simply collecting evidence to demonstrate learning.

In the ec^{uk} criteria, it is the *underpinning knowledge and understanding* that is being judged. But the term “*underpinning*” implies that what is sought is out of sight, and that the existence of the knowledge and understanding has to be inferred from the evidence that the candidate provides. Another author, similarly, in a set of criteria, refers to ‘*underlying constructs*’ [20] rather than identifying an explicit content. Throughout assessment processes much remains unsaid. Some authors therefore refer to the tacit knowledge brought to bear by assessors. In other cases authors refer to a hidden curriculum evinced by ‘*a lack of match between the published criteria and feedback received*’ by candidates [19] — a topic pursued in [23] and [24].

Attempts to provide more specific criteria have not always been successful since they “*may encourage ... a concentration on the mechanics of the task rather than a meaningful engagement*”. Adding more detail raises the candidate’s

anxiety and gives rise to a reductionism that leads them “to focus on sometimes quite trivial issues [such as] the word count” and to ask such mundane questions as “Am I allowed to use bullet points...?” [22] All this suggests that the creation of comprehensive explicit criteria is impractical.

One reaction to the difficulties of conveying criteria is to put more effort into their expression, but experience has shown that

experts are unable to make standards explicit after months of learned debate [19]

and the attempt may not bring about improvements. Also, the development of criteria by a clique of experts does little to disseminate the criteria or expose the experts to the practical difficulties of applying their formulation.

One study showed that there was a wide variation on the part of the assessors in the interpretation of written criteria. Even with terms that are commonly used in academic discourse the meaning is uncertain. For example, the differences in interpretation of the word “analysis” were likely to lead to different kinds of judgement by different assessors. On the positive side it was found that there was “more agreement than not” amongst tight-knit teams of teachers. [24] Such investigations suggest that criteria are socially constructed — meaning emerges from use [25] — and that this implies that if candidates are to assimilate the criteria then they need to be integrated into the tight-knit team so the meanings of criteria “emerge in local communities of practice” and “in the context of specific tasks” [7], which in this case is the task of creating and assessing a technical report.

The creation of a universally understood and explicit set of criteria has, in many instances, proved to be idealistic and criteria are revealed to be

of limited practical use if presented in isolation without the benefit of explanation, exemplars and the opportunity for discussion [19]

In addition to stating the criteria there is a “need to clarify the meaning of terms and phrases” [19].

It is unlikely that standards can be set that will cover every conceivable engineering report. Report production is not necessarily a regular requirement in all engineering jobs and different industry sectors and different companies will have different house styles. If the report is specially created for an assessment exercise then there is always the question about whether the familiar company style is suitable, and this introduces a basic uncertainty on the part of the candidates and the assessors. One way of helping candidates is to provide examples of reports. The assessors could provide exemplars, but it is unlikely that they will match the candidate’s house style and the candidate may feel they have to imitate the exemplar. Alternatively the assessors can be given an example of the candidate’s style and asked to comment, but this transforms what set out to be an assessment into a teaching and learning task.

Making sure that candidates and assessors broadly agree on what a good report is and what a good interview performance is a prerequisite for fair and consistent assessment and an over emphasis on explicit criteria does not necessarily help. More attention has to be paid to the

Socialisation processes ... necessary for tacit knowledge transfer to occur ... based on practice, imitation, feedback and discussion [19]

And in the context of the ec^{uk} assessment this implies that the candidates need support from mentors familiar with the assessment of project reports.

QUALITY

The characteristics of assessments for competence proposed by a number of authors have been collated [26], and in the resulting list of ten characteristics some are connected with the choice of task and others with the process of assessment.

Authenticity, for example, is a characteristic commonly associated with assessment carried out in educational establishments where an assessment involves simulated tasks representative of the future work of the candidate. Authenticity in work-based learning is not an issue providing the candidate is engaged in projects that suit the assessment criteria.

Directness is an indication of how elaborate and how far inferences must be stretched to reach from the evidence provided by a candidate to an assessment criterion. Indirect assessments require greater expertise on the part of assessors and demand thorough rationalisation. The inferential distance between a technical report and a statement of competence is extensive and points to the requirement for expert and well-trained assessors in the ec^{uk} scheme.

Validity is concerned with whether the assessment measures what it purports to measure [27]

In the ec^{uk} assessment process the experience to be recorded in a technical report needs to be examined to ensure that it provides the opportunity for demonstrating the competences relevant to the future activities of a professional engineer. An assessment that claims to assess competence would have to have “predictive validity” and be an indicator of future performance. It is not sufficient for candidates to describe what they have done in their engineering practice, but to demonstrate competence they would need to provide evidence that their actions were considered, related to the context and capable of adaptation to different circumstances. Examples of such evidence are explanations, rather than descriptions, of what was done and why — explanations that would be judged by competent practitioners to be satisfactory. Since the work situation is likely to be unique, the rote learning and reproduction of published explanations is inadequate and easily detected. Satisfactory explanations will be, at least, adaptations of common accounts tailored to the circumstances of a specific project.

The regulations also indicate that the content should be technical, should demonstrate

their knowledge and understanding of engineering principles [1]

and that

a pure management study is not acceptable [15]

Such considerations are often said to ensure “content validity”, that is the evidence that is to be collected is aimed at the goal of the assessment, which in this case is to assess engineering

knowledge and understanding that has not been certified in some other reliable way.

Another characteristic of an assessment instrument is its reliability.

Reliability is the degree to which the same results or outcome of assessment would be obtained, for example, on a different occasion, in a different context or by a different assessor [27]

Reliability is often an issue because practicalities and economics limit the amount of evidence that can be collected. In a formalised system providing both learning experiences and assessment, there are commonly deadlines for the completion of assessment exercises and this precludes the generation, collection and evaluation of additional evidence except under special circumstances, for example, in the form of examination resits, special interviews or third-party pleas for mitigation. Normative assessments, relying on comparisons between candidates, are difficult to adjust to differences in treatment of candidates. However, in the ec^{uk} process without strict deadlines and a requirement for summative assessment, reliability can be improved in questionable cases by evaluating further evidence created, for example, by the candidate, or by the involvement of more assessors.

Variation between assessors is overcome in the ec^{uk} scheme partly by involving at least two assessors in evaluating each candidate's technical report; the same two assessors also interview the candidate and prepare a single assessors' report. In compiling their report the assessors will have to discuss the criteria and reach an agreement on their interpretation. With a pool of assessors, the administrators can ensure that assessors are paired only occasionally with the same partner. In this way an assessor will partner a number of other assessors and the views on the interpretation of criteria will be distributed and moderated by the varied partnerships.

The credentials offered by one PEI are similar to credentials issued by another. Both agencies would want to be assured that the credentials that they issue are not being undermined by weaker standards of assessment by their partners. Reassurance might come from making records publicly available or by allowing mutual observations of the process. Public scrutiny may be desirable, but the exposure of reports on the work of candidates or even the work of candidates may breach implicit or explicit confidentiality agreements and may unduly affect the candidate's approach to his or her work. An alternative, which can provide a greater degree of confidentiality for the candidates, is to use an independent auditor.

The issue of licences to PEIs by the ec^{uk} requires the PEIs' processes to be open to ec^{uk} audit. The ec^{uk} therefore acts as the authority in reporting on standards set by individual institutions. To avoid problems at the auditing stage a PEI can add internal audits, invite observers from another PEI to observe the assessment processes and respond to feedback on any deficiencies that the internal auditors and observers note.

REPUTATION

The value of an assessment is related to the public perception of the credentials that the assessment supports; often the specific details of the assessment process are unimportant and the value of the credentials is linked to a specific tradition or institution. For new forms of credentials a reputation has to be constructed. The options for building, strengthening or maintaining the value of credentials include trading on an established good reputation (for example, by exploiting a well-established institutional name), advertising (discretely) the rigour of the assessment processes and paying attention to reputation of those providing the expertise involved in judging the candidates. Crucial components of an assessment system are, therefore, assurances that there are good grounds for respecting its reliability and validity. But to generate a reputation these assurances must be widely received and accepted and therefore not kept secret. One clear way of disseminating a reputation is to ensure candidates are stringently assessed and that when successful they remind challengers that they have been through a rigorous process.

However to make honest claims about the strength of an assessment process attention has to be paid to the fairness of the processes, the accuracy of the administration, the availability of resources and the honesty of the participants.

FAILURE

Any failure in an assessment system is a cause for concern. It may be a failure of the candidate to meet the assessment criteria or a failure of an assessor to consider all the available evidence. To avoid failures checks are required throughout the assessment system together with the possibility of taking remedial action. Failures might be considered fewer than two headings: failure to communicate accurately and failure to perform a task as required. The usual strategies for dealing with the possibility of failure are quality controls on the components before they become operational, testing, cross-checking, providing feedback, inspection and replication of activities.

At any point in the process a candidate may fail to meet the criteria for entering a following stage. While a purely summative assessment should generate a result, consideration for the candidate may lead the assessing institution to elaborate on the result and provide a detailed explanation. Such an explanation gives the candidate guidance on how they might revise their evidence, perhaps after gaining more experience or a better understanding of the criteria, and re-enter the assessment process with an improved chance of success. Extreme caution is required since explanations of why people fail are complementary to expressions of criteria for success. If there are no published criteria, the author of the explanation is effectively defining the criteria for the assessment process. If there are published criteria then the explanation must be that specific criteria were not met; any attempt to go beyond the published criteria is a refinement or modification. Thus explanations to candidates need the full approval of those responsible for setting criteria and the implications for criteria would need to be formally recorded and disseminated to all the assessors.

The intricacy of the processes involved in assessment will give rise to occasional slips and inconsistencies. There will always be room for improvement in the processes and in the actions of individuals. Where there is scope for improvement there is the potential for criticism and any criticism will have repercussions in the perceived value of the award resulting from the assessment. The effects of criticism can be ameliorated if the assessment process includes mechanisms for responding by carrying out improvements. One requirement for such a mechanism is intelligence on perceived defects. Information can come from independent auditors who are employed to examine the processes, the participants in the process can provide their comments on their experience and data can be collected. It is, of course, wasteful if the information is collected and there is no process for reviewing the collected data and acting on the review.

HONESTY

The initial design of an assessment scheme is likely to be based on the assumption that most people — assessors, decision makers and candidates — behave honestly. Since the outcomes of the assessment process have value, some candidates may be tempted to cheat or persuade administrators or assessors to subvert the process. Cheating assessment processes has a long and dishonourable history and one consequence is that established assessment systems offer examples of all kinds of mechanisms for detecting and countering attempts at subversion.

Even with everyone behaving honestly, the possibility for undetected fraud can rouse suspicions and undermine the confidence that third parties have in the assessments. Processes that would detect misdemeanours before credentials are issued are needed to protect the value of the process. Deterrence, for instance, is a preventative measure that requires severe penalties, robust disciplinary procedures and a high probability of detection.

Examples of dishonest actions and illustrations of countermeasures include

- **Plagiarism:** Candidates may be tempted to copy substantial passages of existing material. Detection relies on comparisons with known sources held in databases or reliance on assessors who are knowledgeable about the domain covered by the report. Interviewing the candidate can test the candidate's familiarity with the material where plagiarism is suspected.
- **Collusion:** A potential assessor and a candidate may have a mutual interest in the candidate obtaining credentials. To avoid the temptation for an assessor to collude, more than one assessor is commonly employed and care must be taken over the recruitment of and vetting of assessors.
- **Falsification of evidence** The falsification of evidence might involve constructing a fictitious report. This is checked, first by getting third parties to endorse the work history of the candidate to check that the candidate had the opportunity described in the report. The second form of check is to question the candidate closely on the report to detect any inconsistencies between the oral and the written accounts of the project.

- **Falsification of records.** Administrators may also collude with candidates to alter the assessment records or candidates may in some other way get records changed. There are a variety of ways of providing checks on the authenticity of records including the publication of results, the duplication of records, and the endorsement of records by third parties.
- **Impersonation:** Candidates may be tempted to allow someone else to impersonate them, when the evidence is in written form, assessors who are acquainted with the candidates work may be able to detect occasional changes in style. Observing a candidate while he or she creates a whole report is impractical. An interview based on a lengthy report is perhaps more secure when the interview questions are not known beforehand. The interviewee's familiarity with the material in the report gives some assurance about his or her involvement in creating the report.

Fairness

The assessment processes can be subverted by accident if the participants are uninformed about the regulations. Some unfairness can occur through slips, errors and misunderstandings of participants who genuinely believe they are acting honestly: examples include differences in the understanding of the criteria for assessment, prejudice on the part of assessors, mistakes in requests for evidence and unrecognised personal difficulties (illness, lack of access to facilities, conflicting demands for the candidate's time and so on). The thorough briefing of administrators, assessors and candidates on the regulations is a crucial component of the assessment system.

Often regulations evolve in response to criticisms and events but they embody broad principles that get lost in the detailed wording. For the candidates the broad principle is to present the results of their own work and where it is admissible present the work of others and acknowledge the source. For the assessment team the goal is to give the candidate the opportunity to present his or her work with the minimum of hindrance, and to judge the work according to the agreed criteria and without prejudice.

Candidates can be intimidated by the setting in which evidence is collected. It would be difficult to generalise but it is clear that timed examinations in regimented examination rooms can induce severe stress but may affect different people in different ways. Apparently minor issues such as what to wear for a face to face interview can create severe anxiety that affects the performance of a candidate. Any unfairness can be alleviated by choosing assessment options that are known to cause less stress and by treating candidates with sensitivity. For example, aside from the important questions of racial and gender equality there will, in the ec^{uk} scheme, candidates who have worked with large budgets in well-equipped and supportive organisations while others work in declining industries which are under resourced. To alleviate such apparent unfairness a criterion can be added that credits candidates for working within the constraints placed upon them and for exploiting the opportunities presented to them.

The assessors may be unaware of these restrictions and opportunities and with this added criterion it is incumbent on the candidate to describe the constraints and resources available in the project they are reporting.

Challenges from candidates, the public or whistle-blowers about the fairness of the processes demand investigation. Fairness is about the differences in treatment of candidates that are unrelated to the goals of the assessment. Any investigation into unfairness would require evidence that exposed the rationale for judgements made in the assessment process. The evidence is more robust if it is based on records made at the time of the judgement rather than recollections or worse inventions long after the events. And in a fair assessment any rationale should be based on the criteria used for assessment of all candidates. This reinforces the need for explicit statements about the criteria used in the assessments and for the securely maintained and witnessed records.

Equality of opportunity in assessment involves ignoring features of a candidate's identity that might cause the assessors to bring to bear irrelevant or additional criteria. The treatment of candidates can also be unequal when they are not credited with satisfying a criterion that is relevant. To indicate that they have given a candidate due consideration, assessors can record their assessment of the candidate's evidence against all the criteria, including any evidence that they disregarded.

The use of more than one assessor provides some insurance against prejudice. Assessors can be screened for more pernicious forms of bias and they can be trained to be aware of criteria that are not to be applied. It is common to collect statistics in selection schemes and this can be extended to assessment schemes. The statistics provide hints on where a possible bias may be occurring and can be a trigger for further investigation or preventative action.

Economies are often made in assessment systems by asking for a limited amount of evidence. The assumption is then made that the student's performance on these samples is representative of the student's overall capability. A second form of sampling takes place when candidates are asked random questions as a check on the breadth of their knowledge and understanding this form of sampling requires effective security measures since the selections must be genuinely unpredictable. Usually a compromise has to be reached between the coverage of a test and the costs, mainly in time, to the stakeholders. A judgement has to be made as to what is an appropriate coverage. An inadequate coverage might be deemed to be unfair; excessive coverage might be seen as being oppressive.

CREDENTIALS

A candidate's self esteem is likely to be affected by the assessment, but he or she may be more interested in the effects that the assessment will have on others, perhaps employers and colleagues. To convince future contacts that they have completed the assessment, the candidate will require credentials and may need to demonstrate that they are authentic. Often results of assessment are summarised in a transcript or on a certificate. If such documents are used to judge an individual by a third party some means must be available to

validate the claims made on the certificate. This would require comprehensive and secured records maintained by the awarding agency.

Credentials have commonly been issued as elaborate documents that are difficult to reproduce or fake. Developments in document reproduction technologies have reduced the effort required to forge certificates. Assurance about the authenticity of credentials has also come from the publication of the results of assessment, award ceremonies and the archiving of the published results. Trust in credentials is then reinforced by the records kept by the assessment institution and the witnesses to the award. The *ec^{uk}* regulations therefore insist that the license holders to maintain records of the individual assessments.

Once credentials have been issued it is almost impossible to withdraw them. The decision to award credentials therefore requires considerable caution. It is common to have an interval between the final encounter with the candidate and the award, to have a panel of decision makers some of whom will not have been directly involved in the expert assessments and to bring to bear summaries of all the evidence relating to a candidate in borderline cases. The candidates may in some systems also have the opportunity to submit further endorsed evidence on any special circumstances surrounding the period during which assessment took place. Meetings that make decisions on the basis of a candidate's assessment are performative, that is they transform the identity of the candidate. There can be no room for ambiguity in the record of what took place. The careful servicing of the decision making and accurate record keeping of decisions are additional administrative tasks.

COMMUNICATION

With a range of types of tasks to be carried out on different occasions is likely to require a range of people in different roles. From the list of activities it is possible to distil six distinctive roles for people.

- The **administrators** follow well-documented procedures and securely maintain records. They set up arrangements and resources for conducting the assessment, compile statistics to measure the performance of the assessment system and issue the credentials once the decision to make an award has been decided. They do not have to be knowledgeable in the subject area of the assessment.
- **Mentors** advise the candidates on the interpretation of the assessment criteria and possibly provide feedback to the candidates on how they might improve their performance in the assessment exercises.
- **Expert assessors** examine the candidate's work, compile reports on how well the evidence provided by the candidate matches the criteria for assessment.
- **Witnesses** observe different stages of the process, endorse the proper execution of procedures and confirm the identity of authors of the different documents. Some witnesses are likely to be subject experts.
- **Decision makers** review evidence and reports on evidence at key stages. Their decisions authorise the candidate to progress through the early stages and authorise the issuing

of credentials in the final step. The decision makers must have sufficient expertise to interpret the reports of the experts.

- Finally **auditors** examine the records for consistency and completeness.

Inevitably an assessment system employs a variety of agents and to ensure the robustness of the overall system, the interactions within the system must be reliable and secure. Since the assessment task is likely to be broken down into stages and it is likely that different people will be involved in the different stages, there is a need for accurate communication free of interference. The intervals between the stages of the assessment may mean that there are delays in passing on messages between the participants. The fallibility of human memory and the abiding possibility for misunderstanding means that care needs to be taken over the protocols. First, adequate time needs to be allocated for conducting the communication, the communications will need to be recorded close to or at the time when evidence, judgements or interpretations are ready to be conveyed. The communication should be restricted to what is pertinent and this can be aided by providing forms that remind authors what the communication is about. Explicit assessment criteria can also steer communication towards what is significant in the assessment process.

The participants separated in their involvement by time may never meet. Questions may be raised about the authenticity of the communication. Authenticity checks are provided by signatures (written or digital), by all parties to the communication keeping copies of the communication and by involving witnesses who may also keep records.

APPOINTMENTS

During the operation of an assessment scheme, appointments will be made and some thought must be put into the characteristics required of the appointees. Assessors, for example, will be involved in interviewing candidates. The ec^{uk}'s procedures ask that a candidate is interviewed on his or her project. The regulations state:

The report shall be assessed, through a process of informed peer assessment, by two assessors ... who will explore the report in an interview. [15]

An interview based on a written report is similar in format to viva voce examinations that are commonly a part of the assessment for the award of a Doctorate in the UK. Research has been carried out on the goals of this oral examination and found that there are “*significant inconsistencies*” and a “*lack of transparency*” in the assessment practice and this is partly a consequence of the limited number of people involved in these interviews and hence the limited opportunity to acquire shared experience. This is being exacerbated by the wider involvement of examiners, who are not familiar with the processes adopted in a system that relies on “*the tacit knowledge about PhD examining that academics are assumed to acquire*” and raises the question, in the context of this paper, of how interviewers are to acquire skill in assessing, and how suitable assessors are to be found. Especially with a lack of agreement on exactly what the purpose of an oral examination

is. But perhaps this is an indication that the purpose of the viva will vary and this, the researchers discovered was the opinion of academic assessors. Academic assessors saw the purposes of the oral as an authentication of the candidate's involvement in the reported work, examination of the candidate, monitoring standards, providing guidance and advice, or simply part of a ritual. Many candidates felt the viva was a painful experience particularly where the interview did not redeem the thesis. [28]

Overall the research sounds warnings about

The inexplicit ways in which examiners are selected, and the ‘confidentiality’ of the viva, [which] mean that doctoral assessment can be unacceptably uncertain. [29]

Since there are likely to be uncertainties, in the ec^{uk} scheme the assessors produce a report on the interview, but they do not make a decision. A larger panel of experienced assessors can make the decision about a satisfactory level of underpinning knowledge and understanding based on all the reports of the interviewers and assessors and any expressed concerns of the candidates. If there are any doubts the reports and the candidates submission can be assessed by a second assessment team.

ECONOMICS

It is important to realise that no system for assessing the capabilities of individuals is perfect and indeed most are generously laced with compromises. One of the severe restrictions on an assessment scheme is the available time of the participants. The following lists give an illustration of the scale of effort involved.

The candidate will invest effort in

- Finding out what the specific assignment is
- Planning its completion in conjunction with the concurrent tasks he or she has
- Conducting any research or practical activity associated with the assignment
- Addressing the criteria
- Collating and formulating evidence of his or her effort
- Critically comparing the collected evidence against the criteria
- Revising the evidence
- Handing over the evidence to the assessment administration

Producing a report though also involves learning from experience and this will mean that the first construction of the report will suggest to the author a better formulation. It is unlikely therefore that a single draft will be sufficient to create a satisfactory result. Drafting and redrafting will carry substantial costs in time beyond the gaining of experience that provides the material for the report.

The assessors' efforts include

- Disseminating of the criteria for assessment
- Evaluating the candidates' work
- Setting assignments and checking them to ensure they will provide an opportunity for candidates to demonstrate their ability to satisfy the assessment criteria without any unwarranted overheads.

- Participating in any appeals
- Revising criteria for future assignments
- Considering any feedback. Administration involves
- Scheduling the assessment and ensuring the schedules are maintained
- Managing the logistics and storage associated with assessment materials
- Policing where security is an essential part of the assessment system.
- Normalising of assessment standards which requires experienced assessors and a level of formalisation that adds to the administrative overhead.
- Conducting appeals processes for those who feel they have been the victims of overzealous assessors, unfair assessment or those accused of fraud.

Often the costs of educational programmes and the shaping of educational programmes concentrate on teaching and learning. But formal educational programmes also provide assessment of individuals, the ec^{uk} technical report process assumes that teaching and learning have been completed and the task is to assess that learning. Assessment even in this cursory analysis proves to be expensive in people's time and much of the expenditure of time has financial implications. Attempts to economise are likely to reduce the value of the product of the assessment. Candidates may wish to have quick, cheap and easy assessment, but the long-term effects are likely to be detrimental and hard to predict.

AUTHORITY

With a great deal at stake, challenges to the accuracy of the processes are bound to occur. While such challenges are to be welcomed in helping to detect and correct errors and failures, continued challenges can postpone decisions. Ultimately there has to be a way of terminating the assessment process and this implies the assignment of authority to a decision making body with an assured termination procedure such as a majority vote, or handing the decision to a single senior judge. Clearly the authority has to be well-informed, well-qualified to make the judgement and willing to defend decisions.

CONCLUSIONS

The ec^{uk} technical report assessment process is compactly described in a few regulations. The implications of the regulations, as with any rigorous assessment practice, are far reaching, but can be informed by existing practices and research results.

Current research suggests that to ensure fairness, the dissemination of criteria for assessment is desirable but gaining a common understanding of criteria amongst candidates and assessors requires an interaction which may be limited by the available time of the participants.

Work situations, the literature notes, can provide excellent opportunities for learning and it would appear that the costs are low. However, the cost of rigorous assessment with its checks on the honesty of candidates, effectiveness of assessors and

security of credentials necessarily involve substantial effort and time.

Candidates, assessors and the designers of the assessment system all need to judge finely how much effort they need to expend. Inadequate effort on the part of the candidates can result in failure, insufficient endeavour on the part of the assessors can result in unfairness and ineffective work on the part of the scheme designers can result in a damaged reputation for the organisation. There is, unfortunately, no obvious way of estimating how much effort is needed to create a worthy scheme or identifying where economies can be safely made.

REFERENCES

- [1] "UK Standard for Professional Engineering Competence, Chartered Engineer and Incorporated Engineer Standard", Engineering Council UK: London, 2003, p.12
- [2] "Applicant's Handbook for Chartered Professional Engineer (CPEng) Chartered Engineering Technologist (CEngT) and Chartered Engineering Associate Officer (CEngO)", Engineers Australia: Canberra, 2004, p.5
- [3] "Students - So, you want to be an engineer?", http://www.engineeringk12.org/students/so_you_want_to_be_an_engineer/default.htm, American Society for Engineering Education, October 2004
- [4] Malpas, R., *The Universe of Engineering — A UK Perspective*, The Royal Academy of Engineering: London, June, 2000, p.20
- [5] Maclellan, E., "Assessment of Learning: the differing perceptions of tutors and students", *Assessment and Evaluation in Higher Education*, 26, No4, 2001, pp.307–318
- [6] Astin, A.W., *Four Years that Matter: the college experience twenty years on*, Jossey-Bass: San Francisco, 1997
- [7] Knight, P.T., "Summative Assessment in Higher Education: practices in disarray", *Studies in Higher Education*, 27, No.3, 2002, pp.276–286
- [8] McAlpine, L., Designing learning as well as teaching. *Active learning in higher education*, 5(2) 119–134 (2004)
- [9] Kilpatrick, W.H., "The project method—The use of the purposeful act in the educative process", *Teachers College Record*, 19, No.4, 1918, pp.319–335
- [10] Alberty, H.B., *A study of the project method in education*, Ohio State University Press: Columbus, Ohio, 1927
- [11] Earl, L.M., *Assessment as Learning*, Corwin Press: Thousand Oaks, 2003, p.34
- [12] Smith, K. & Tillema, H., "Clarifying Different Types of Portfolio Use", *Assessment and Evaluation in Higher Education*, 28, No6, 2003, pp.625–648
- [13] Maclellan, E., "How convincing is alternative assessment for use in higher education?", *Assessment and Evaluation in Higher Education*, 29, No3, 2004, pp. 311–320
- [14] Brown, G., *Assessment: A guide for lecturers*, LTSN Generic Centre, 2001
- [15] *Regulations for Registration*, Engineering Council, London, 2004
- [16] Hornby, W., "Assessing Using Grade-related criteria: a single currency for universities?", *Assessment and Evaluation in Higher Education*, 28, No4, 2003, pp.435–454
- [17] Tariq, V.N., Stefani, L.A.J., Butcher, A.C. & Heylings, D.J.A., "Developing a new Approach to Project Work", *Assessment and Evaluation in Higher Education*, 23, No3, 1998, pp.221–240
- [18] Orsmand, P., Merry, S. & Reiling, K., "A Study in Self-assessment: tutor and students' perceptions of performance criteria", *Assessment and Evaluation in Higher Education*, 22, No4, 1997, pp.357–369
- [19] Rust, C., Price, M. & O'Donovan, B., "Improving Students' Learning by Developing their Understanding of Assessment Criteria and Process", *Assessment & Evaluation in Higher Education*, 28, No2, 2003, pp.147–164
- [20] Wharton, S., "Defining Appropriate Criteria for the Assessment of Master's Level TESOL Assignments", *Assessment and Evaluation in Higher Education*, 2No6, 2003, pp.649–663
- [21] Ritter, L., "The Quest for an Effective Form of Assessment: the evolution and evaluation of a controlled assessment procedure (CAP)", *Assessment and Evaluation in Higher Education*, 25, No4, 2000, pp. 307–320

- [22] Norton, L., "Using assessment criteria as learning criteria: a case study in psychology", *Assessment & Evaluation in Higher Education*, 29, No6, 2004, pp.687-702
- [23] Sambell, K. & McDowell, L., "The construction of the hidden curriculum: messages and meanings in the assessment of student learning", *Assessment and Evaluation in Higher Education*, 23, No4, 1998, pp.391-402
- [24] Webster, F., Pepper, D. & Jenkins, A., "Assessing the Undergraduate Dissertation", *Assessment and Evaluation in Higher Education*, 25, No1, 2000, pp.72-80
- [25] Wittgenstein, L., translated by Anscombe, G.E.M., *Philosophical Investigations*, Blackwell: Oxford, 1958
- [26] Baartman, L.K.J., Bastiaens, T.J. & Kirschner, P.A., "Requirements for Competency Assessment Programmes", *Onderwijs Research Dagen 2004*, Universiteit Utrecht, 2004
- [27] Martin, S., "Two models of educational assessment: a response from initial teacher education: if the cap fits", *Assessment & Evaluation in Higher Education*, 22, No3, 1997 pp.337-344
- [28] Jackson, C. & Tinkler, P., "Back to Basics: a consideration of the purposes of the PhD viva", *Assessment and Evaluation in Higher Education*, 26, No4, 2001, pp.355-366
- [29] Morley, L., Leonard, D. & David, M., "Variations in Vivas: quality and equality in British PhD assessments", *Studies in Higher Education*, 27, No3, 2002, pp.264-273