Chapter 12
Assessment of Significant Learning Outcomes

Richard Daugherty, Paul Black, Kathryn Ecclestone, Mary James, and Paul Newton

12.1 Introduction

The nature and quality of the outcomes of learning are central to any discussion of the learner's experience, from whichever perspective that experience is considered. For those outcomes to be assessed it is also necessary to articulate in some way the constructs on which such judgments are based. The relationship between the intended outcomes of learning and the outcomes as evidenced through assessment is typically conceptualized in terms of the alignment of assessment to curriculum or of congruence between them (Baker, 2005; Porter, Smithson, Blank, & Zeidner, 2007; Beck, 2007; Biggs & Tang, 2007). In principle, for the assessment of outcomes to be valid the inferences drawn from the evidence of learning should be in line with the intended learning outcomes. In practice, the way in which learning outcomes are defined and assessed varies greatly within and across systems of education (European Centre for the Development of Vocational Training, 2008).

The project that is reported here suggests that the relationship between assessment and curriculum is more multi-dimensional and multi-level than the terms "alignment" or "congruence" would imply. That project, "Assessment of Significant Learning Outcomes" (ASLO), was a seminar series funded by the Teaching and Learning Research Programme (TLRP) of the UK Economic and Social Research Council (ESRC) – http://www.tlrp.org/themes/seminar/daugherty/index.html. Five case studies were chosen to illuminate the relationship of assessment to curriculum in different educational contexts:

- A school subject: mathematics education in England.
- Learning to learn: a European Commission project to develop indicators.
- Workplace learning in the UK.
- Higher education in the UK.
- Vocational education in England.

R. Daugherty (✉)
School of Social Sciences, Cardiff University, Wales, UK
e-mail: daughertyr@cardiff.ac.uk

In each of the context-specific seminars in the ASLO series the participants analyzed the terms in which the alignment of assessment procedures to learning outcomes was discussed in that context. This involved exploring how, and by whom, control over programmes of learning is exercised as well as how those who are engaged in the discussions perceive and express the issues involved. The overall aim was to identify insights that may have applications beyond the context from which they emerged rather than to develop an overarching conceptual framework that could be applicable to any context.

12.2 Background

The roots of the ASLO project can be found in the work of the Assessment Reform Group (ARG) and in TLRP's Learning Outcomes Thematic Group (LOTG).

Since its inception as a response to the policy changes in curriculum and assessment brought in by the Education Reform Act 1988, the ARG has reviewed the implications for policy and practice of research on assessment. It has taken a particular interest in the relationship between assessment and pedagogy (Gardner, 2006) and between assessment and curriculum, especially through its work on enhancing quality in assessment (Harlen, 1994). In recent years the assessment/pedagogy interaction has been a prominent focus of the Group's work (for example ARG, 2002).

The ARG has argued, for example in the Assessment Systems for the Future project (Harlen, 2007), that assessment regimes that rely only on test-based measures of attainment may be insufficiently valid to be educationally acceptable. Implicit in that critique are such questions as:

- What are the significant learning outcomes that are not being assessed in a system that relies wholly on test-based assessment procedures?
- What are the indicators of student performance which have been/could be developed in relation to such learning outcomes?
- What are the assessment procedures that do not rely on testing but do give/could give dependable measures of student performance in relation to those indicators?

Consideration of validity is the natural starting point for the assessment dimension of the project, drawing on the work of Crooks, Kane, and Cohen (1996), Stobart (2008) and others. There are recurring themes concerning the technical aspects of validity that can be traced across diverse contexts. It is also clear that a focus on "consequential validity" (Messick, 1989) or, alternatively, on the "consequential evidence of validity" (Messick, 1995), necessarily raises questions such as "what consequences?" and "consequences for whom?"

The project also drew on work done by the TLRP, the remit of which was to sponsor research "with the potential to improve outcomes for learners". In 2004, a grounded analysis by the Programme's LOTG of the outcomes mentioned in the first thirty TLRP projects to be funded, led it to propose seven categories of outcome:

- *Attainment* – often school curriculum based or measures of basic competence in the workplace.
- *Understanding* – of ideas, concepts, processes.
- *Cognitive and creative* – imaginative construction of meaning, arts or performance.
- *Using* – how to practise, manipulate, behave, engage in processes or systems.
- *Higher-order learning* – advanced thinking, reasoning, metacognition.
- *Dispositions* – attitudes, perceptions, motivations.
- *Membership, inclusion, self-worth* – affinity towards, readiness to contribute to the group where learning takes place.

(James & Brown, 2005, pp. 10–11)

However, this list was insufficient to capture the range of theoretical perspectives on learning underpinning these projects. Therefore another categorization was based on the metaphors of learning represented in project outputs. A matrix was devised with the classification of learning outcomes on one axis and the metaphors of learning (drawing on Sfard, 1998, distinction between acquisition and participation metaphors), underpinning the construction of those learning outcomes, on the other.

It was evident that the TLRP projects had had difficulty in conceptualizing learning outcomes to take full account of dimensions of learning such as: surface and deep; process and product; individual and social; intended and emergent. James and Brown (2005) pointed out that a reconceptualization of learning outcomes would present considerable challenges:

> The first challenge would be to convince stakeholders that the existing models no longer serve us well; the second would be to convince them that alternatives are available or feasible to develop. Alternatives would also need to be succinct, robust and communicable... (p. 20).

It is to these challenges that the ASLO project was a response.

12.3 Contexts

The educational environment within which current policies and practices have evolved has inevitably shaped the way in which learning outcomes, and the assessment of them, are conceptualized. But the influence of the wider social, economic and political context on the prioritization of learning outcomes and on the approach taken to assessment is also clearly evident in the project's five case studies. The evidence reviewed here relates to the case study contexts at the time the seminars took place, between January and October 2007.

12.3.1 Case Study 1: National Curriculum Mathematics in England

Consideration of school mathematics was particularly relevant to our enquiry, because it is subject to an unusual set of pressures. One critic can claim that all the mathematics the average citizen needs is covered in Key Stage 2 (for

students from age 7–11), another that the increased mathematization of our culture makes advanced understanding crucial, whilst an academic has asserted that real understanding of mathematics only begins at the level of an undergraduate course.

Ernest (2000) characterizes the many different stakeholders in terms of five categories:

- industrial trainers;
- technological pragmatists;
- old humanist mathematicians;
- public educators;
- progressive educators.

The views of each of these groups differ, over the aims of mathematics education, over the teaching needed to secure these aims, and over the means to assess their achievement. The operational meaning of their aims is often not clear, and the means are often ill thought-out and ill-informed. The ascendant tendency at present in the UK is to focus on "numeracy", or "application of number", or "quantitative literacy" or "functional mathematics" and on attempts to bring these into working practice (Wake, 2005).

Such groups exert pressures in contrary directions, so it is hardly surprising that many describe the school scene as fractured and unsatisfactory. Some align in approach with Ernest's "old humanist mathematicians". They will typically be well-qualified but have a limited approach to teaching and learning, giving priority to algorithmic capacity to solve well-defined mathematical problems. Others will have a similar vision, but, being less well-qualified and/or confident, will be more narrowly dedicated to teaching to the test; many see the latter as a particularly weak characteristic of mathematics education (Advisory Committee on Mathematics Education, 2005). Such teachers will find it hard to be clear about what counts as being good at mathematics, i.e. they will not have a clear concept of validity. Those practitioners who are "progressive educators" will have clearer views about validity, usually at odds with the aims reflected in the formal tests.

A consequence of this situation is that many pupils have an impoverished experience of the subject, in ways pointed out by Schoenfeld (2001), who wrote of his experience as:

- mainly consisting of the application of tools and techniques that he had just been shown;
- being mainly "pure" and lacking opportunity to be involved in mathematical modelling;
- not involving real data;
- not being required to communicate using mathematics.

The fault line which runs through much of this is between mathematics, seen as the performance of routine algorithms, and mathematics seen as a tool to tackle "everyday" or "real world" problems. The former leads to assessment of achievement with

well-defined exercises, which have a single right answer, with learners inclined to think of achievement as arriving at that answer. The latter looks for evidence of a capacity to tackle the rather messy contexts which are characteristic of every-day problems, problems for which there is no right answer, and where explanation of the way the problem has been defined and of the approach adopted, including justification for the methods used, are as important as the "answer" itself. Such work is much more demanding to guide, and harder to mark. Yet pupils taught in this way achieve as well in the General Certificate of Secondary Education (GCSE) as those taught in more traditional methods, will take more interest in the subject, will be better able to see mathematics as useful in everyday life and will be better able to tackle unusual problems (Boaler, 1997).

The National Curriculum in mathematics in England gives prominence, in Attainment Target 1 (AT1), to "using and applying mathematics". There are clear statements about different levels of competence in tackling problems, but no mention of the nature or context of such problems, so no guidance on "textbook" versus "everyday" choices. The other three ATs are about the formal content of mathematics. Teachers see this curriculum as overcrowded; this in part is due to the atomistic approach to the formulation. The ACME (2005) report recommended that "The Government should reduce the overall volume and frequency of external assessment in mathematics", and reported the general belief in the mathematical community that "many of the negative effects of external assessment are serious". The 2007 revision has reduced the content to focus on a few "big ideas", but teachers seem to be misinterpreting the text as broad statements which still imply that all the content has to be "covered".

The testing system is of course of crucial importance here. With time-limited tests to cover a very full curriculum, any activity which involves much more time than that in which a single examination answer can be given is ruled out, thus ruling out realistic problems. There was teacher based/coursework assessment for AT1, but teachers saw this as stereotyped and providing little opportunity for interesting activities or for ways to assess them. For such activities, the right-answer approach does not work, and it is difficult for teachers to work with the inevitable ambiguities (Morgan & Watson, 2002).

There is thus an invalidity block, which could in principle be cleared by strengthening the use of teachers' own assessments in national tests and public examinations. That these can achieve validity with an acceptable level of reliability has been argued in general terms by the ARG (ARG, 2006). Nevertheless, the current coursework assessment at GCSE is unpopular: a consultation by the Qualifications and Curriculum Authority (2006) showed that mathematics teachers "thought that existing coursework did not provide a reliable and valid assessment for the subject" and it has been abandoned. At the same time, the experience of the King's Oxfordshire Summative Assessment Project project (Black, Harrison, Hodgen, & Serret, 2006a, 2007) is that mathematics teachers can develop their own summative assessment in ways that they find rewarding and which can produce dependable results, but that such development will be hard to achieve.

In summary, whilst the National Curriculum could be interpreted to reflect a valid representation of mathematics, the testing system does not realize this potential. However, to repair this mis-alignment would require changes which would demand extensive professional development for teachers, and a consensus about the aims of mathematics education which does not at present exist.

12.3.2 Case Study 2: Learning to Learn

The seminar on the assessment of "learning to learn" (L2L) drew on evidence from three UK projects and from the European Union (EU) Learning to Learning Indicators (Fredriksson & Hoskins, 2007). The papers revealed, more clearly than any of the other project case studies, the significance for the way assessment and learning are conceptualized of the contexts in which the constructs involved are developed. As McCormick argued in his commentary on the EU project (McCormick, 2007), it is essential to understand the *purposes* of measuring L2L as well as the *views of learning* underpinning its conceptualization.

The work of James and her colleagues (James et al., 2007) in England on "learning how to learn" (LHTL), has primarily focused on the development of pupils' learning practices. An early attempt to devise instruments to assess learning to learn "competence" encountered two obstacles. One was the dependence of the outcomes on the nature and context of the task. The second was that the project team could not agree on what the tasks were measuring. A deeper consideration of the concept of "learning to learn" (Black, McCormick, James, & Pedder, 2006b) led to the conclusion that "learning to learn" is not an entity, such as a unitary disposition or mental trait, but a family of *practices* that promote autonomy in learning. Thus the "how" in the project's preferred terminology was considered important, as was the close relationship between "learning how to learn" and learning per se. The implications are that LHTL practices can only be developed and assessed in the context of learning "something" in substantive domains; they are not easily, validly or comprehensively assessed by instruments similar to IQ tests or by "self report" inventories.

Thus, assessments of LHTL are likely to require sustained observation of how learners develop learning strategies for learning within domains – an argument for most emphasis to be placed on assessment by teachers in authentic learning contexts. The conceptualization of "learning to learn" and "learning how to learn" that emerged here (Black et al., 2006b) was not shaped by policy considerations and, if taken seriously, would call into question the appeal of these popular ideas as expressions of assessable learning outcomes.

Claxton and his colleagues at the University of Bristol were also interested in "learning to learn" for "lifelong learning" and how this might be assessed. They state the aims of their work as:

> . . . firstly, to seek to *identify* the elements that define a good learner. Secondly. . .. to devise an instrument that could be used to assess where an individual [is] located in relation to those elements at any given time and in any particular context. (Deakin Crick, Broadfoot, & Claxton, 2004, p. 248)

Their intentions, however, were not to develop a measure of "learning to learn" attainment that could be used in the policy arena, but to develop instruments for formative and diagnostic use by learners and their teachers. To this end they developed a self-report instrument, the Effective Lifelong Learning Inventory – ELLI, which focuses on "learning power", argued as being concerned with quality of learning (rather than with learning competences) and defined as:

> A complex mix of dispositions, lived experiences, social relations, values, attitudes and beliefs that coalesce to shape the nature of an individual's engagement with any particular learning opportunity. (http://www.ellionline.co.uk/research.php – accessed 26 July 2010).

Seven dimensions of "learning power" were identified, and scales for each were developed. These were described as: changing and learning, meaning making, curiosity, creativity, learning relationships, strategic awareness and resilience. Although these constructs are much more broadly defined than those to which conventional assessments of attainment are related, the "self-report" nature of the tools meant that they were relatively easy to construct. The instrument developers saw no need to devise tasks and contexts in which these dispositions and behaviours could be demonstrated. There are, of course, questions about whether respondents' answers to the questions are realistic, even if they strive to be honest, and whether the statements apply in all contexts, but the problems encountered by James and her colleagues (Black et al., 2006b), concerning the operationalization of constructs, were avoided.

The important point to be made here is that the origins and purposes of an instrument are crucial for understanding and judging its value. The ELLI project team wanted to develop measures of their constructs for diagnostic and formative purposes. Self-report instruments may be valid for at least some of these purposes though their validity, in relation to the constructs and to the particular uses of evidence from the instruments, is potentially problematic. If, however, the intention is to find measures of learning to learn for evaluation and decisions on matters of public policy, then their validity and reliability for those purposes may come more into question.

In contrast to these projects the work of Hautamäki and his colleagues in the University of Helsinki has been overtly linked to a declared purpose associated with national policy. Although the original purpose was to develop tools for school self-evaluation, it has been used to evaluate the outcomes of education in Finland and judge the "effectiveness" of the national system of education. Since 1995 the National Board of Education in Finland has sponsored work in the University of Helsinki to develop tools to assess learning to learn, one of five aspects of system effectiveness. School development is claimed to be the "first and foremost" purpose of the "learning to learn" assessment instruments although the assessment places the school on a national scale thereby directly comparing individual schools with national norms.

According to the researchers in Helsinki, "learning to learn" is defined as:

> the competence and the willingness – here referred to as beliefs – to adapt to novel tasks. Competence refers to the generalized knowledge and skills that develop by studying

different subjects at school and which is needed for learning new things. Beliefs and attitudes direct the use of these competencies in new situations. (http://www.helsinki.fi/cea/english/opiopi/eng_opiopi.htm – accessed 26 July 2010).

Learning competencies are assessed as generic skills demonstrated in specific contexts, for example, the ability to identify salient points in an argument developed in the context of a literature task, or the ability to use evidence in a science task. The assessment of beliefs and attitudes is based on self-report questionnaires similar to the ELLI instruments. The resulting 40 scales are described as an "easy to execute and cost effective measure", although the learning competences scales are vulnerable to the challenges that James and her team encountered, and the self-report scales have some of the limitations of the ELLI instruments.

These might not matter much if the instruments were primarily intended for internal diagnostic and formative use by schools though whether the evidence derived from the instruments is valid for such purposes would still need to be demonstrated. However, the discourse of policy is evident here in the wording of the question to which policy-applicable answers are being sought: "What kind of learning-to-learn skills does the education system produce?"

In terms of purpose, the current EU project to devise "indicators" of learning to learn is from the same mould. Its origins lie in the aspirations of the leaders of EU states meeting in Lisbon in 2000 which led in time to the European Education Council's support for a programme of work on eight such key competencies, one of which is learning to learn. In the absence of accepted Europe-wide measures of this as yet loosely defined construct, a new working group was set up "to develop new indicators for the monitoring of the development of education and training systems" (Fredriksson & Hoskins, 2007, p. 4). Thus, assessment as a source of performance indicator data has been the explicit driver of this EU project from the outset.

McCormick has argued that defining and developing measures of learning to learn as a way of supplying governments with performance data could distort and damage the construct which the LHTL team have been trying to nurture in the pedagogy of schools in England:

> . . . in a field where we have trouble defining the concept of L2L, where there are probably few well tried classroom practices for various aspects of L2L, and where we have to struggle to find the instrument that represents whatever we can agree L2L means, we start to improve [education] by measuring. This is the proverbial assessment tail wagging the curriculum dog! (McCormick, 2007, p. 1)

Thus, regardless of its uncertain foundations, the construct of "learning to learn" is being shaped by the need for it to be measurable in ways that will supposedly illuminate the performance of the diverse education systems to be found in the nation states of the EU. Or, put another way, the measures currently being devised by this EU indicators project seem to aim at emphasizing validity for monitoring system performance, and at de-emphasizing validity for identifying individual student learning needs.

12.3.3 Case Study 3: Workplace Learning in the UK

The seminar on workplace learning considered evidence about the nature, scope and ethos of assessment in workplaces, drawing on case studies by Fuller and Unwin (2003) of the Modern Apprenticeship programme in three companies associated with the steel industry, and discussion in two papers by Eraut (2007a, 2007b). One paper focused on the ways in which feedback in different workplace contexts hinders or enhances professional learning and competence and the other on progression in forms of expertise and knowledge over a period of time in different professions.

Fuller and Unwin highlight (p. 408) "the relevance of the institutional arrangements, including the nature of the employment relationship and the formal qualifications required by the programme". The nature of these relationships and the ways in which a workplace deals with the formal requirements for apprentices to develop particular knowledge and competences through a framework of minimum qualification requirements offers some apprentices very "restrictive" environments to "get them through" the formal competences demanded in the qualification, and "expansive" environments that enable apprentices to develop more extensive knowledge and competence.

Understanding the alignment between assessment and learning outcomes in work-place learning is made more complex by the extent to which formal summative requirements are specified tightly or loosely. This takes different forms at different levels of work-based qualifications. For example, the Modern Apprenticeship scheme requires workplaces to enable trainees or workers to achieve tightly specified competence-based qualifications as part of National Vocational Qualifications (NVQs) while an accountant might complete several competence-based qualifications followed by a degree. At different qualification levels, and across different professions and occupations, workplaces vary in having loose frameworks of codified knowledge, skills and notions of progression in expertise, or no codified frameworks at all.

This complexity makes it necessary to understand more about the interplay between informal and formal assessment and the ways in which these are sometimes linked to forms of appraisal and performance review. There is also an interplay between the use of formal, codified knowledge in such systems and the tacit, intuitive forms of knowledge that professionals use often without realizing, but which are crucial to effective performance as part of "capability". These include knowledge embedded in task performance, personal development, team work, the performance of different roles, the application of formal academic knowledge and skills, decision-making and problem-solving.

The work of Eraut and colleagues illuminates some of the subtle and complex ways in which different types of knowledge inter-relate through their studies of five occupational groups – doctors, health scientists, nurses, accountants and engineers. That work shows numerous variables shaping the learning of individuals in workplaces that are very diverse, where the learning and informal and formal assessment cultures that nurses, for example, experience can vary between wards, even in the same hospital (Eraut, 2007a, p. 10). The specification of learning outcomes

and forms of assessment, formal and informal, summative and formative, therefore varies enormously across professions and workplaces. Eraut and colleagues' detailed longitudinal analysis of the factors that lead to effective support and mentoring, particularly through good feedback, has implications for assessor training and development in workplaces, both for those designated with formal assessment roles and for those who support colleagues more informally but are, nevertheless, carrying out assessments.

This analysis has several implications for how knowledge is defined, taught and assessed and for how workplaces can foster the intuitive and uncodified knowledge essential to effective practice. First, attempts to capture, codify and then assess informal and tacit uses of knowledge will not necessarily lead to more effective learning in the workplace. The more restricted, formalized and reified the assessment artefacts and forms of knowledge become, and the more they are tied to formal assessments, such as appraisal and performance review, the more likely they are to hamper the sort of conversations and feedback that lead to effective professional learning. On the other hand, if they are just left to chance, essential activities that develop capability, such as induction into the particular learning climates of groups being joined, the mentoring and management of different roles, and day-to-day formative assessment, will not be developed to best effect. Summative assessments are also crucial but perhaps more as snapshots of a professional's learning trajectory rather than as a dominant feature of workplace assessments.

This implies that workplace mentors, assessors and colleagues need to help novices become inducted into the practices of their new occupations so that they can apply tacit and formal knowledge to complex situations as and when they arise. Notions of progression, from novice to expert, and the types of knowledge they use are illuminated through the work of Eraut and colleagues over many years of study. Recent work shows the ways in which feedback can be used more effectively to develop what Eraut refers to as "capability" (rather than competence) as integral to expertise (Eraut, 2007b, p. 4). Developing the skills and processes of effective feedback in different workplaces is crucial for developing capability since the ability to deal effectively with an unfamiliar situation in medicine or engineering, for example, could be vital.

The very obviously situated nature of learning in the workplace, and the complexities of how feedback is used in specific contexts, has implications both for the codification of relevant knowledge and for how the learner's performance is assessed. Eraut and colleagues' work suggests that finding effective ways to align learning outcomes, formal and informal assessment and to codify the right sorts of knowledge without over-specifying them, must be done in the context of each profession or occupation and its relevant stakeholders and interest groups.

12.3.4 Case Study 4: Higher Education in the UK

The seminar on higher education discussed a report on "innovative assessment" across the disciplines (Hounsell et al., 2007) together with two further papers from

Hounsell and colleagues (Hounsell & Anderson, 2008; Hounsell, 2007). A defining feature of the relationship between curriculum and assessment in this sector is that "a distinctive and much-prized characteristic of higher education is that the choice not only of curriculum content and teaching-learning strategies but also of methods of assessment is to a considerable extent devolved" (Hounsell et al., 2007, p. 12). Even the "academic infrastructure" put in place by the UK regulatory body, the Quality Assurance Agency (QAA), emphasizes the fact that its codes of practice, qualification frameworks and subject benchmarks "allow for diversity and innovation within academic programmes". In higher education the regulatory texts have a relatively low profile within the discussion of curriculum and assessment. However, it is crucial to note that this profile varies considerably across disciplines and across institutions, shaped by the learning cultures of disciplinary communities and of institutions. For example, the QAA regulatory texts appear to exert more influence on programme planning and on the assessment of students' work in the post-1992 universities sector than in the pre-1992 sector.

Higher education is one of only two of the case study contexts (vocational education being the other) in which the term "learning outcomes", as used generically by the LOTG, has established currency. Except in a minority of institutions that are content to rely on long-established practices, usually involving responsibility for curriculum design and for assessment resting with the course tutor(s), the specification of "intended learning outcomes" has become integral to the practices of teaching and learning in UK higher education. Among the problems discussed at the ASLO seminar were the difficulty of capturing high quality learning in the language of learning outcomes, with the pitfalls of vagueness/vapidity on the one hand and undue particularity and prescriptiveness on the other.

In this respect the discussion echoed the project's concern about neglect of "significant" outcomes without suggesting ways of resolving dilemmas about both defining and assessing such outcomes. But what was also evident were the pressures on the specification of learning outcomes, typically articulated at institutional level, that were generated by governments' expectations of higher education, for example to demonstrate student employability. Such instrumentalism, communicated by government through its agencies, has similar roots to equivalent influences on the school mathematics curriculum and work-based training and assessment in qualifications such as NVQs.

In spite of the impact of the regulatory framework there is also ample evidence of the staff responsible for course programmes evolving their own interpretations of the "what", "how" and "why" of the learning involved (see, for example, the exploration of "ways of thinking and practising" in two subject areas, biology and history, discussed in Hounsell and Anderson (2008)). However, the goal of many course designers in higher education of "introducing students to the culture of thinking in a specific discipline" (Middendorf & Pace, 2004) may not be compatible either with the aspirations of the diverse student population on first degree courses or with the procedures that universities often adopt for assessing student attainment. While the enculturation approach to course design may move discussion beyond reductive lists of measurable learning outcomes it presents the challenge of valid assessment in a

different form – how to judge a student's progress in terms of "connoisseurship" of the subject area.

For most if not all first degree courses in UK universities, the end-of-programme requirement to summarize student performance in a degree classification is a powerful influence on curriculum and pedagogy as well as, more directly, on assessment practices. A picture emerged of assessment in higher education constrained by "delivery" models of teaching and learning. The potential for formative feedback to enhance the quality of learning is undermined by the variability and often poor quality of such feedback as lecturers and their students are typically preoccupied with "what counts" in the reckoning that awards students an end-of-programme degree. In those circumstances, the issue of validity does not appear as an explicit item on the agenda of course designers, disciplinary groups or the institutional committees that have oversight of programme specifications. Instead questions of alignment are buried deep in the interface between course content and assessment, with assumptions about learning and learning theory that are implicit in the formal curriculum and in the associated pedagogy seldom being made explicit.

In contrast to the context in which school mathematics is evolving in England, with the policy texts dominating the discourses, the issue of alignment of curriculum and assessment in UK higher education is being worked through at the local level as the tutors responsible for course units/modules plan their teaching. In the traditional subject-based first degree programme questions about how to assess student learning are more likely to be influenced by departmental colleagues or within-discipline assumptions than by a thorough consideration of the extent to which intended learning outcomes and the evidence elicited by assessment of student performance are aligned. Amid such diversity as is allowed for by responsibility for curriculum and assessment being devolved there are, of course, many exceptions to that generalization. Such exceptions can be found not only among the instances of "innovative" assessment reported by Hounsell et al. (2007) but also in degree programmes where specific content knowledge and skills are required for the programme to be accredited.

12.3.5 Case Study 5: Vocational Education in England

Questions about definitions of outcomes, standards and curriculum content, and their effects on assessment practices in vocational education, arise in a context of numerous failed attempts since the late 1970s to create "parity of esteem" between vocational and academic education, and to encourage young people to see vocational education as a genuine high status alternative to programmes based on traditional subjects.

Assessment based on prescriptive and detailed specifications of learning outcomes, portfolios of achievement, unit-based assessment, locally-devised, teacher-assessed projects and grading based on "learning to learn" skills, has been used partly as a motivating device to encourage young people to gain a credible qualification, partly as an attempt to foster independence as part of "lifelong

learning" skills and attitudes and partly as way of reflecting in the curriculum the concerns of employers.

Although these developments have influenced broader education debates about what comprises fair and useful assessment, there is little political, professional or public agreement about curriculum design and content in vocational education, nor about its purpose in relation to the content and outcomes of general education. The combined effect of lack of consensus and ad hoc reforms has been programmes comprising a range of functional, generic and personal skills, attitudes and dispositions and a very uncertain subject base, where diverse bodies compete to have their learning outcomes included (see Ecclestone, 2002; Stanton, 1998).

Learning outcomes in vocational education also reflect competing aims:

- motivating learners who would otherwise not stay on in post-16 education or who are disaffected in Key Stage 4 by responding to and rewarding their expressed interests and notions of relevance
- expanding routes into higher education whilst also making sure that expansion does not lead to over-subscription for limited places
- preparing students for progression into work and job-related NVQs
- encouraging learners to carry on gaining qualifications
- keeping students labelled by defenders of A-levels and GCSEs as "less-able" from "undermining" standards in these qualifications
- convincing learners, teachers, admissions tutors that vocational education has parity of esteem with long-running, higher status academic qualifications
- ameliorating poor levels of achievement in numeracy and literacy through "key skills"
- unifying disparate and confusing post-16 qualification pathways
- satisfying demands from different constituencies, such as employers' representatives or subject associations, to include "essential" content and skills
- having credibility in the school sector which has less experience of mainstream vocational education

A number of studies show these factors affect teaching and assessment practices, ideas about "types" of young people suitable for vocational education, beliefs about their motivation and attitudes to learning.

First, despite political targets to raise levels of participation and achievement, there are large gaps between notions of "choice" and "opportunity" and actual progression. Vocational students often choose progression routes that reflect their images of themselves as "types" of learners suited for different "types" of assessment and while they see themselves as "vocational", many students' vocational aspirations are erratic and vague (see Biesta & Davies, 2006; Davies & Biesta, 2007; Bathmaker, 2003; Torrance et al., 2005; Ecclestone, 2002).

Second, choice is affected by the ways in which learning outcomes and assessment both reflect and reinforce certain "learning identities" and "learning careers", and the creation of self-fulfilling images of learning, progression and appropriate assessment activities. The concept of "learning cultures" illuminates the subtle ways

in which students and teachers develop implicit and explicit expectations about teaching, learning and assessment, and how, in turn, these interact with peer norms and relationships, official requirements, institutional ethos and structures and the nature of the relationship between teachers and students (Ecclestone & Pryor, 2003; Ecclestone, 2004).

Third, dispositions and attitudes cannot be isolated from employment prospects, the effects of educational selection and differentiation in a local area, students' social class and cultural background and the educational institutions they choose or are sent to. Images of achievement and failure, and a learning career associated with those images, affect students' and teachers' perceptions about the suitability of a vocational or academic qualification and are rooted in teachers' and students' perceptions about employment and education prospects in local labour markets.

Fourth, ideas about "achievement" and "learning" are influenced by targets to raise attainment of grades, overall pass rates, retention on courses and progression to qualifications at the next level. "Learning" and "achievement" are often synonymous with learning outcomes and criteria prescribed by the awarding body, so that "assessment" is frequently the "delivery of achievement".

Finally, assessment is affected by teachers' images of what students like, need and want. Vocational tutors regard "good assessment" as practical, authentic and relevant activities, work-experience and field trips: there is a widespread view that "these students" do not want or like written assessment, that they are less secure, need more group affinity and should be in a more protected, safe environment. Many vocational teachers see assessment as integral to a strong ethos of personal development that minimizes stress or pressure. Assessment to develop subject knowledge is not prominent in their espoused goals for students, an attitude reinforced by learning outcomes that emphasize generic skills and attitudes rather than subject content. Vocational teachers and students like to work in a lively and relaxed atmosphere that combines group work, teacher input and time to work on assignments individually or in small friendship-based groups. Goals for relevance and real-life application are reinforced by concerns that assessment should engage and retain young people in formal education who are deemed to be demotivated and disengaged.

One effect is a growing tendency to avoid "burdening" vocational students with "too much written work" or with methods that alienate them from formal education. It is now commonplace to elide vocational education with practical activities loosely related to work, so that learning outcomes and assessment are associated with the need to motivate and engage young people. A recent phenomenon is to associate disaffection with "fragile learning identities" and "low self-esteem".

The ad hoc evolution of learning outcomes and assessment methods in vocational education in England over the past 30 years has been a central factor in creating and maintaining certain images and attitudes to learning in vocational education. Difficulty in creating an enduring, high status vocational counterpart to general education, and a stable system of organizations and bodies to implement it, might be countered by a better understanding of:

- how learning outcomes, pedagogy and assessment are inextricably linked
- how they are affected by political imperatives for achieving targets and
- how they are shaped by the learning cultures of different vocational education settings.

12.4 Discussion

Several themes, discussed more fully in Daugherty, Black, Ecclestone, James, & Newton, 2008, recur across the five case studies.

Construct definition – how, and by whom, the constructs involved are defined, interpreted and made real – has emerged as a major issue in each of the contexts. Construct validity has long been a central concern in the field of assessment without the constructs themselves necessarily being critically explored or closely defined. Even if the constructs have been considered at the levels of assessment theory and qualification design, they may not be applied in the day-to-day practice of assessors. At the other end of the curriculum/assessment relationship the constructs informing the design of programmes of learning have in some contexts been strongly contested. What this suggests is a need to clarify the constructs within a domain that inform the development both of the programmes of learning, in principle and in practice, and of the related assessments.

A second theme, *progression,* is crucial to the design and implementation of learning programmes, and in particular for the implementation of assessment for learning. Its relevance to summative assessment depends on the structure of the assessment system. If the only high-stakes summative test is a terminal one, then the desired final outcomes are laid down, the test constructors have to reflect these in as valid a way as they can, and the teachers discern, from study of a syllabus and of examples of the test instruments and procedures, how best to focus their work. Enabling progression is absolutely central to formative assessment but there is evidence in these case studies that summative assessment requirements, driven by pressure for uniformity and for accountability, can constrain teachers and trainers in using their own judgment to nurture progression.

Another theme to emerge across the case study contexts was *the impact of assessment procedures* on the alignment between intended or desirable outcomes from learning and those outcomes which actually emerge. From a measurement perspective, alignment is often conceived quite narrowly – in terms of content validity – where misalignment between an assessment instrument and intended learning outcomes represents a threat to the integrity of inferences from assessment results. However, it can be conceived more broadly too, where misalignment represents a threat to the integrity of learning itself, resonating with the notion of "systemic validity" (Frederiksen & Collins, 1989). The five case study contexts highlighted numerous situations in which the nature of an assessment procedure threatened to disrupt the acquisition of desirable learning outcomes by students. This disruption occurred when assessment procedures led either to the failure to acquire

desirable outcomes from learning, or to the acquisition of undesirable outcomes from learning. For both types of disruption potential impacts were attributable either to the design of the assessment instrument or to the nature of the assessment event itself.

A fourth theme to emerge was *system-level accountability as a driver of alignment.* Accountability takes very different forms, has different purposes and stakeholders and has different effects on the interpretation of learning outcomes within each of the contexts reviewed. Two of the case studies in particular – the school mathematics curriculum and the learning to learn indicators – revealed just how influential the political imperatives for system level accountability can be. They can be seen to be determining not only the role of assessment in defining the relevant constructs but also, perhaps more crucially, in shaping how teachers and students then interpret and enact those constructs.

12.5 Conclusion

It became clear in the course of the ASLO seminar series that the language of intended outcomes, alignment and curriculum is embedded in different ways in the assumptions, histories and practices of the different sectors of formal education. It has also been increasingly evident that, in asking whether the inferences drawn from assessments are *aligned* to intended learning outcomes, the project was not using the most appropriate language to express the dynamics of the assessment/curriculum relationship. It is certainly true that "alignment of an assessment with the content standards that it is intended to measure is critical if the assessment is to buttress rather than undermine the standards" (Linn, 2005, p. 95). But "alignment" implies that there is something in place – content standards in the case of the US contexts to which Linn is referring – to which assessments could, at least in principle, be aligned. All the ASLO case studies have exposed a lack of clarity in defining the underlying constructs, whether in terms of content standards or of narrower/broader formulations.

The case study evidence reviewed here has taken the analysis of the relationship between curriculum and assessment beyond the simple notion of explicit outcomes of assessment being in some way aligned to, or congruent with, a pre-specified curriculum. Instead we see a multi-layered process of knowledge being constructed, with numerous influences at work at every level from the national system to the individual learner.

References

Advisory Committee on Mathematics Education (2005). *Assessment in 14-19 mathematics*. London: Royal Society.

Assessment Reform Group (2002). *Assessment for learning: 10 Principles*. Cambridge: University School of Education.

Assessment Reform Group (2006). *The role of teachers in the assessment of learning*. London: University Institute of Education.

Baker., E. (2005). Aligning curriculum, standards and assessments. In C. A. Dwyer (Ed.), *Measurement and research in the accountability era* (pp. 315–335). Mahwah, NJ: Lawrence Erlbaum Associates.

Bathmaker, A. M. (2003). *Learners and learning in GNVQs*, PhD thesis. Coventry: University of Warwick.

Beck, M. D. (2007). Review and other views: 'alignment' as a psychometric issue. *Applied Measurement in Education*, *20*(1), 127–135.

Biesta, G., & Davies, J. (2006). Going to college: aspirations and experiences of vocational education students. Paper from the *ESRC TLRP Project, Transforming Learning Cultures in FE*. Exeter: University of Exeter.

Biggs, J., & Tang, C. (2007). *Teaching for quality learning at university* (3rd ed.). Buckingham: SRHE and Open University Press.

Black, P., Harrison, C., Hodgen, J., & Serret, N. (2006a). *Strengthening teacher assessment practices, learning and evidence*. Paper presented at the 2006 BERA Conference, Cardiff: a report of findings from the first stage of the King's-Oxfordshire Summative Assessment Project (KOSAP).

Black, P., Harrison, C., Hodgen, J., & Serret, N. (2007). *Riding the interface: an exploration of the issues that beset teachers when they strive for assessment systems that are rigorous and formative*. Paper presented at the 2007 BERA conference, London: 2nd report of findings from the King's-Oxfordshire Summative Assessment Project (KOSAP).

Black, P., McCormick, R., James, M., & Pedder, D. (2006b). Learning how to learn and assessment for learning: A theoretical inquiry. *Research Papers in Education*, *21*(2), 119–132.

Boaler, J. (1997). *Experiencing school mathematics: Teaching styles, sex and setting*. Buckingham: Open University Press.

Crooks, T., Kane, M., & Cohen, A. (1996). Threats to the valid use of assessment. *Assessment in Education*, *3*(3), 265–285.

Daugherty, R., Black, P., Ecclestone., K., James, M., & Newton, P. (2008). Alternative perspectives on learning outcomes: Challenges for assessment. *The Curriculum Journal*, *19*(4), 243–254.

Davies, J., & Biesta, G. (2007). Coming to college or getting out of school? The experience of vocational learning of 14- to 16-year-olds in a further education college. *Research Papers in Education*, *22*(1), 23–41.

Deakin Crick, R., Broadfoot, P., & Claxton, G. (2004). Developing an effective lifelong learning inventory: The ELLI project. *Assessment in Education*, *11*(3), 247–272.

Ecclestone, K. (2002). *Learning autonomy in post-compulsory education: The politics and practice of formative assessment*. London: RoutledgeFalmer.

Ecclestone, K. (2004). Learning in a comfort zone: Cultural and social capital in outcome-based assessment regimes. *Assessment in Education*, *11*(1), 30–47.

Ecclestone, K., & Pryor, J. (2003). 'Learning careers' or 'assessment careers'?: The impact of assessment systems on learning. *British Educational Research Journal*, *29*(4), 471–488.

Eraut, M. (2007a). *Feedback and formative assessment in the workplace*, expanded version of a paper for an AERA symposium (2006) on formative assessment in professional education.

Eraut, M. (2007b). *How do we represent lifelong professional learning?*, expanded version of a paper given to an EARLI SIG meeting.

Ernest, P. (2000). Why teach mathematics? In S. Bramall & J. White (Eds.), *Why learn maths?* (pp. 1–14). London: University Institute of Education.

European Centre for the Development of Vocational Training (2008). The shift to learning out-
comes. Conceptual, political and practical developments in Europe. Luxembourg: Office for
Official Publications of the European Communities. Accessed July 27, 2010, from http://www.
cedefop.europa.eu/etv/Upload/Information_resources/Bookshop/494/4079_en.pdf

Frederiksen, J. R., & Collins, A. (1989). A systems approach to education testing. *Educational
Researcher*, *18*(9), 27–32.

Fredriksson, U., & Hoskins, B. (2007). Indicators of learning to learn, paper presented to ASLO
project seminar, London, March 2007. Accessed July 27, 2010, from http://www.tlrp.org/
themes/seminar/daugherty/index.html

Fuller, A., & Unwin, M. (2003). Learning as apprentices in the contemporary UK workplace:
Creating and managing expansive and restrictive participation. *Journal of Education and Work*,
16(4), 407–426.

Gardner, J. (Ed.) (2006). *Assessment and learning*. London: Sage.

Harlen, W. (Ed.) (1994). *Enhancing quality in assessment*. London: Paul Chapman.

Harlen, W. (2007). *Assessment of learning*. London: Sage.

Hounsell, D. (2007). Towards more sustainable feedback to students. In D. Boud & N. Falchikov
(Eds.), *Rethinking assessment in higher education, learning for the longer term* (pp. 101–113).
London: Routledge.

Hounsell, D., & Anderson, C. (2008). Ways of thinking and practising in biology and history:
Disciplinary aspects of teaching and learning environments. In C. Kreber (Ed.), *Teaching and
learning in higher education: Within and across the disciplines*. London: Routledge.

Hounsell, D., Blair, S., Falchikov, N., Hounsell, J., Huxham, M., Klampfleitner, M., et al. (2007).
Innovative assessment across the disciplines: An analytical review of the literature (Review
Report and Database). York: Higher Education Academy.

James, M., & Brown, S. (2005). Grasping the nettle: Preliminary analysis and some enduring issues
surrounding the improvement of learning outcomes. *The Curriculum Journal*, *16*(1), 7–30.

James, M., McCormick, R., Black, P., Carmichael, P., Drummond, M. -J., Fox, A., et al. (2007).
Improving learning how to learn – classrooms, schools and networks. London: Routledge.

Linn, R. (2005). Issues in the design of accountability systems. In J. Herman & E. Haertel (Eds.),
*Uses and misuses of data for educational accountability and improvement: 104th yearbook of
the national society for the study of education* (pp. 78–98). Chicago: NSSE.

McCormick, R. (2007). Learning to learn: reflections on assessment, curriculum and pedagogy,
paper presented to ASLO project seminar, London, March 2007. Accessed July 27, 2010, from
http://www.tlrp.org/themes/seminar/daugherty/index.html

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13–103). New
York; London: Macmillan & American Council for Education.

Messick, S. (1995). Validity in psychological assessments. *American Psychologist*, *50*(9),
741–749.

Middendorf, J., & Pace, D. (2004). Decoding the disciplines: A model for helping students learn
disciplinary ways of thinking. In D. Pace & J. Middendorf (Eds.), *Decoding the disciplines:
Helping students learn disciplinary ways of thinking (New directions for teaching and learning,
98)* (pp. 1–12). San Francisco: Jossey-Bass.

Morgan, C., & Watson, A. (2002). The interpretative nature of teachers' assessment of mathemat-
ics: Issues of equity. *Journal for Research in Mathematics Education*, *33*(2), 78–110.

Porter, A. C., Smithson, J., Blank, R., & Zeidner, T. (2007). Alignment as a teacher variable.
Applied Measurement in Education, *20*(1), 27–51.

Qualifications and Curriculum Authority (2006). News release. Qualifications and Curriculum
Authority. Accessed July 16, 2007, from http://qca.org.uk/qca_9678.aspx

Schoenfeld, A. (2001). Reflections on an impoverished education. In L. A. Steen (Ed.),
Mathematics and democracy: The case for quantitative literacy (pp. 49–54). Princeton, NJ:
National Council on Education and the Disciplines.

Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational
Researcher*, *27*(2), 4–13.

Stanton, G. (1998). Patterns of development. In S. Tomlinson (Ed.), *Education 14-19: Critical perspectives*. London: Athlone Press.

Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. London: Routledge.

Torrance, H., Colley, H., Garratt, D., Jarvis, J., Piper, H., Ecclestone, K., et al. (2005). *The impact of different modes of assessment on achievement and progress in the learning and skills sector*. Learning and Skills Development Agency. Accessed July 16, 2007, from https://www.lsda.org.uk/cims/order.aspx?code=052284&src=XOWEB

Wake, G. (2005). *Functional Mathematics: more than 'Back to Basics'*. Nuffield Review of 14-19 Education and Training: Aims, Learning and Curriculum Series, Discussion Paper 17. London: Nuffield Foundation.