Learning and Assessment in the Knowledge Society[1]

Bertil Roos
Department of Education
Umeå University
S-901 87 Umeå
Sweden

Bertil.Roos@pedag.umu.se

Abstract:

This paper raises questions about assessment and the knowledge or learning society. It focuses on the emergent field of *alternative* assessment and its links with practices usually described, by default, as *traditional* assessment.

[1] This is a revised version of a paper discussed at the EDEN 2004 Annual Conference, *New challenges and Partnerships in an enlarged European Union: Open, Distance and e-Learning in Support of Modernisation, Capacity Building and Regional Development*, 16 – 19 June, University of Technology and Economics, Budapest, Hungary.

Background

A new climate has arisen in higher education in Europe. Elite education has retreated. Since the Second World War, the spread of mass higher education has, in some countries, approached the level of near-universal post-compulsory education, creating mass higher education systems of a totally new kind. Mass higher education has become 'a very high-stakes affair', a carefully-managed gateway to socially prestigious occupations and, accordingly, the 'good life' (Bakker, 2001, p. 1). Young people and their parents are aware of the value of continued education. New pressures have been placed on the examinations and tests which mediate university entrance

> *In a world where university diplomas play an ever-increasing role in determining people's lives, demand for tertiary education and continuing pressure on the secondary examinations used by gatekeepers, will continue for decades to come (Bakker, 2001, p. 6).*

Yet, increased pressure on university places has also been accompanied by another political demand – that universities should be inclusive institutions where greater attention is given to gender and ethnic equity and to democratic forms of intellectual life.

What, however, does this mean for the future of university examinations and assessment? Should they continue to serve elitist assumptions about social selection? Or should they take on additional roles in the modern university? One new task, widely discussed as 'alternative' assessment, is that universities should give greater attention to assessment *for* learning, rather than assessment *of* learning. Put another way, universities should link assessment to the promotion as well as the measurement of learning. To this extent, they will contribute to the advancement of the ´learning society´.[2]

The remainder of this paper situates alternative assessment as a complement to traditional assessment. It argues that assessment theory has at least two histories and two sets of assumptions; that these assumptions are often confused; and that an important task for educationists is to clarify these assumptions. By reference to an international literature that includes the work of Sfard (Israel), Wells (Canada), Linn and Shepard (USA) and Wiliam, Torrance and Black (UK), it

[2] Detailed discussion of the "Knowledge Society" will be included in another paper.

distinguishes between behaviourist and constructivist modes of learning and assessment; it explores the relationship between learning and feedback; ***and*** it clarifies a series of tensions that currently hamper the work of educationists notably the differences between formative and summative assessment, high- and low- stakes testing, and divergent and convergent assessment.

A concrete – and international - example of such tensions differences in thinking about the assessment of performance is the so-called 'Bologna Process' – an initiative to harmonise a European 'qualifications framework'. As an official Swedish document has commented:

> *Most European countries have more grading levels than the three usually applied in Sweden. A majority of countries have between five and nine levels while many countries lack a common grading system. Certain countries use relative[norm-related] scales while other work with goal- [criterion-] related grades. Accordingly, European grading scales are thus different and even vary within institutions in the same country. (Ds 2004:2, p. 112)*

Learning And Assessment Theories

As the Bologna process suggests, the investigation of teaching and learning is an international issue. Yet, as the Bologna Process also indicates, there is no consensus. Many models and metaphors circulate around the academy. The question: is a learner a bottle to be filled or a fire to be lighted goes back to Ancient Greece. But it has also been the subject of Lorrie Shepard's presidential Address to the American Research Education Association (2000), and her parallel contribution to the fourth *Handbook of Research on Teaching* (2001). She started by identifying the learning paradigm that dominated the USA in the 20th century, one that included behaviorist learning theories, hereditarian theories of intelligence and individual difference, and the scientific measurement of ability and achievement. The key assumption behind this model was that learning is hierarchical, sequential and occurs `by accumulating atomized bits of knowledge' (2000, p. 6). Thus, in the early part of the twentieth century, close links were forged between social efficiency thinking, scientific management, `hereditarian theories of individual differences' and `associationist and behaviorist learning theories.' Learning was regarded as the accumulation of stimulus response associations. In turn, these various respective psychological theories were served by assessment theories based on `scientific measurement of ability and achievement'. (2001, p. 1068).

This paradigm took shape early in the 1900s. It was part of a wish to increase the efficiency of factories. Social efficiency theorists believed it was possible to

use the same principles in schools as in factories. Students' abilities had to be measured since they predict their future role in life. It was claimed that students with low ability should not be taught things that they would never use in the factory or on the farm. A leader in the social efficiency movement John Franklin Bobbitt, voiced the principle that each student should be educated 'according to his capabilities.' His thought influenced US school curricula during the 1920s, so that they became highly differentiated and largely utilitarian (Shepard, 2001, p. 1069). Although the French psychologist who developed the first IQ tests, Alfred Binet, believed in 'the educability of the intelligence', most US psychologists in the early 1900s saw "IQ test results as a scientifically exact measure of a fixed trait' (Shepard, 2001, p. 1069). They saw intelligence as innate; and the way to handle it in education was by using a differentiated curriculum (see also Kliebard, 1995).

By the end of 1900s, Shepard notes, the tide of informed opinion had turned. Evidence, for instance, on the negative impact of labeling children had led to a re-assessment of learning. By the millennium the intellectual climate meant that: `most scientists and educated citizens´ assigned a `much more limited role to heredity´. They recognized the multidimensional nature of ability´, and they were `aware of the large effect of past learning opportunities on both test performance and future learning´. (Shepard, 2001, p. 1069)

Nevertheless, Shepard also recognised that old theories were still supported by public opinion and, as result, continued to exert their influence on school practices – through, for example, grading and assessment practices (p. 1069-1070).

Figure 1, taken from Shepard's address, summaries this argument. It shows the 20th century dominant paradigm on the left, present-day teaching practices with assessment separated from instruction in the middle, and an emergent, paradigm of learning and assessment, on the right. In the extended version of her presidential address, Shepard (2001) characterised U.S. practice as a mismatch of traditional testing, on the one hand and instruction influenced by an emergent paradigm, on the other. Theories from the past, that learning is like building a brick wall, layer by layer, serve as the default framework for the assessment of learning. She highlighted the absence, therefore, of a valid connection between assessment and instruction. (p. 1067)

Constructivism – an emergent paradigm

In another review of learning theories, the American researchers, Duffy and Cunningham (1996), describe the brick wall theory as the `objectivist' model. Knowledge is transferred from the teacher to the student, the reception of knowledge is passive, and such learning is unreflective. Learning, therefore, is about acquiring knowledge; and instruction is about delivering knowledge.

Duffy and Cunningham, like Shepard, go on to describe an alternative model which they characterized (1996, p. 175) as the 'constructivist' view. Here, too, they make a crucial distinguish between the activities of learning and the activities of instruction or teaching:

> *1) learning is an active process of constructing rather than acquiring knowledge, and (2) instruction is a process of supporting that construction rather than communicating knowledge (Duffy and Cunningham 1996, p.171)*

Despite using `communication' rather than `delivering' Duffy and Cunningham, have nothing to say about a constructivist view of assessment. This separation of learning from assessment has been a recurrent problem in educational theory: What then, is the relationship between teaching and assessment? One answer to this question has emerged alongside constructivists views of learning that build on the work of Piaget, Vygotsky, and other cognitive scientists. This later post-behaviorist view holds that new knowledge is developed - or constructed - from old knowledge. Anna Sfard, an Israeli researcher, highlights this constructivist dimension of learning by refering to the 'learning paradox', also known from the writings of Plato as Menos´ paradox : "*How can we want to acquire a knowledge of something that is not yet known to us?*[3] Sfard's assumption, is that human beings reach new states of knowledge through building upon their current knowledge (see for an earlier analysis Bruner, 1968).

To summarize: Shepard (2001) declared that the old paradigm, with hereditarian theories of intelligence, social efficiency curriculum and scientific efficiency has led to an atomistic view of both learning and the measurement of learning. The new paradigm, constructivism, stresses that all human knowledge is constructed (or reconstructed). Individuals organize information in their own way, transcending their own – and pre-existing - knowledge structures. Moreover, as in Sfard's case, the new paradigm entails cognitive development through social experience. Individuals are introduced to new ways of thinking, with the help of social experiences that serve as 'scaffolding' (Shepard, 2001, p. 1075). Such scaffolding may be erected by themselves, or by others who are held responsible for their learning (e.g. parents, teachers).

[3] Another version of Meno's paradox is: If you don't know what you are searching for, then what is it you are searching for; and if you do know what you are searching for, why are you searching? (see also Marton & Booth, 1997, ch. 1).

Learning and Feedback

One form of instructional scaffolding, provided by teachers, takes the form of dialogue. Learners, that is, are helped into new ways of thinking by resources - words, ideas, values, symbol systems - mediated in dialogue. Shepard (2001) asserts that teaching, in the 21st century, should pay closer attention to this view of mediated learning. By using reciprocal teaching – with mutual agreement concerning outcomes, question generating, summarizing, and clarifying - teachers not only offer learners new knowledge but also new understanding (p. 1076). Through such understanding, according to Shepard (2001), the learner can transfer knowledge from one context to another. These, understanding therefore, become "robust', enabling the transfer of knowledge and understandings. (p. 1076).

Shepard links this view of learning to a specific learning environment. Students see their teacher as a resource, a 'guide in the learning process, rather than as an evaluator'. Scaffolding emphasises mastery-based feedback rather than normative evaluation. Evaluating `errors and mistakes' is seen as a 'normal part of learning'. Without such learning environments, Shepard suggests, students will developed a performance orientation rather than a learning orientation. The brick wall paradigm is re-furbished, as 'extrinsic rewards' are given for 'success on easy tasks'. Dialogue through scaffolding, however, takes place in students' zones of proximal development, learning environments where, she suggest, learners can "participate in active inquiry and discussion of challenging problems' (p. 1077).

A more developed discussion of dialogue as scaffolding occurs in the work of someone who has worked in the UK, Canada and the USA, Gordon Wells. His *Dialogic Inquiry: Towards a socio-cultural practice and theory of education.* (2001) builds on Vygotsky's view of constructivism (see also Daniels, 2001). In turn, he emphasises three features of dialogic pedagogy: (1) 'the essentially dialogic nature of the discourse in which knowledge is co-constructed'; (2) the 'significance of the kind of activity in which the knowing is embedded'; and (3) the 'important role played by the artifacts that mediate the knowing' (p. 127). Wells believes, therefore, that an important role for the teacher is to create a

> *… classroom community in which the search for understanding, and the dialogue through which this is accomplished, pervades all areas of the curriculum and is inclusive of all students, whatever their social, ethnic, or linguistic background.* (Wells, 2001, p. 119-120).

He also stress that ‘an adequate curriculum theory must utilize an interactive model of teaching and learning’ and that the teacher’s role is based on ‘initiating and guiding this dialogue’. Equally, teacher must encourage students to ‘be agentive in directing their own learning’ and, accordingly, that teachers should ‘seek to equip them [learners] with socially valued ways of thinking and acting’ (p. 119-121).

To foster dialogue, Wells points to the value of collaboratively produced knowledge artifacts, ‘psychological tools’ of various kinds, which are ‘available for further knowledge building’. Further he delegates to each community of inquiry ‘how these principles can most effectively be realized in different settings’ (p. 129-130). One of these tools is feedback and, as Wells acknowledges, the organization of feedback is both contextual and social. Thus, it is possible to speak of constructivist feedback. Nevertheless, Wells’ analysis of dialogue misses an important feature of feedback, one that contributes to the terminological and practical confusions that hamper the parallel development of educationists. To be specific, he fails to recognize that there are different kinds of feedback, and that only some forms of feedback - here described as cybernetic - can be considered constructivist.

Feedback and Cybernetics

Feedback is a concept that came into everyday use from communication engineering and, in particular, from the field of cybernetics (see Roos & Hamilton, 2004). In the process, however, the nuances that were incorporated by the early pioneers of cybernetics (e.g. Norbert Wiener) were lost. They clearly distinguished different kinds of feedback: homeostatic feedback (self-regulation), servo-mechanical feedback (which keeps a system on a pre-defined course) and self-directing feedback (whereby feedback not only affects a system's behaviour but also its structure or organisation). Indeed, in this sense of new structures, a direct connection exists between cybernetics and constructivism (see for instance, von Glasersfeld, 1979). This last perspective has been described as “The property of a body, process, or machines (without closed loop control) of reaching a *new* steady state after a sustained disturbance” (Mayr, 1970, p. 134, emphasis added).

These ideas may have been influential the field of informatics, but their effect on educational practice seems to be limited. `Feedback´ took on a restricted meaning, merely the return flow of information. In an effort to monitor their

performance, for instance, institutions and agencies, including university course teams, seek `feedback' from clients and consumers. This information, ostensibly, is used to evaluate and revise existing practices. This consumer-oriented sense of feedback is problematic, however, because it says nothing about how the information is used. In Sadler's terms, the information is left 'dangling' (Sadler, 1989, p. 121). Return flow may take place but nothing changes. Such feedback, therefore, is may not used cybernetic ally – to create new structures. Thus, popular use of the word 'feedback' may not embrace the essence - and etymology - of the word *cybernetic*, that feedback serves as a governor.

This cybernetic, constructivist sense of feedback lies at the heart of the so-called alternative or 'authentic' assessment movements. Yet, at the same time, alternative assessment is sometimes troubled by the fact that feedback, like learning theory, also has a two histories. One set of origins may have been the growth of information engineering in the 1930s, but another history reaches back to reflexology and stimulus-response theories of brain activity that took shape in the nineteenth century. Thus, the establishment of an alternative sense of assessment, requires that it distinguishes itself not only from common sense uses of 'feedback' but also from behaviourist conceptions of 'response'.

This last problem has been discussed by Bangert-Drowns and his co-workers (1991). They avowed that in the early 1900s classroom equipment was developed that could give test takers immediately response to their answers. This sense of feedback remains in place, and was central, for instance, to notions of programmed learning in the middle of the twentieth century. Students who are provided with information about their performance improve their results while students denied such information show little improvement. This finding, according to Bangert-Drowns et al., was still true in the 1980s. Effective feedback from the teacher about students performance could, if matched with the students own expectations about performance, improve students learning (p 213-214). 'Without feedback', they suggest, 'mutual influence is impossible' (p. 214).

The problem, however, with the argument of Bangert-Drowns and his colleagues is that it embraces an undifferentiated sense of feedback. What, for instance, is meant by 'mutual influence'.

By 2001, Shepard was well aware of the feedback problem – the fact that it is central to *both* behaviourist and constructivist theories of teaching and learning.

> *The idea of feedback comes from electronics where the output of a system is reintroduced as input to moderate the strength of a signal. Correspondingly, both behaviorist and constructivist learning theories take for granted that providing information to the learner about performance will lead to self-correction and improvement. (Shepard, 2002, p. 1091-1092)*

Further, she recognised - unlike Bangert-Drowns - that meta-analysis of the literature on feedback is of limited value.

> *For the most part, however, meta-analyses of the feedback literature are of limited value in re-conceptualizing assessment from a constructivist perspective, because the great majority of existing studies are based on behaviorist assumptions. (1092) [And] Relatively few studies have been undertaken in which explicit feedback interventions have been tried in the context of constructivist instructional settings. (Shepard, 2002, p. 1092)*

The recovery of a cybernetic sense of feedback seems to have come from the fields of systems and management science, themselves a fusion of 'operations research', (military) targeting, and information engineering (cf. McCorduck, 1979, passim). A key mediator seems to have been Arkalgud Ramaprasad, then working at Southern Illinois University. His ideas have been progressively refined by Elshout-Mohr and currently (2004) professor in Information and Decisive Sciences (IDS) at the University of Illinois (Chicago) and, above all, D. Royce Sadler.

Ramaprasad (1983) defined feedback as 'information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way' (p.4). The information is itself not feedback. 'For feedback to exist, the information about the gap must be used to alter the gap. If the information on the gap is merely stored without being utilized to alter the gap, it is not feedback' (Ramaprasad, 1983, p. 5).

Elshout-Mohr (1994) also noted problems with using the term 'feedback' :

> *What educational psychologists mean precisely by feedback varies. One can simply understand it to mean 'the obtaining of information of good (corresponding to a norm) or bad (diverging from a norm) quality of what the pupils produce (or reproduce). Or the term can be*

> *defined more broadly. The term feedback is used more to indicate a educational function that should be fulfilled in all learning processes, even the higher ones (Elshout-Mohr, 1994, p. 2).*

Royce Sadler, working in Australia, also focused his attention on feedback, but made a significant intervention. He started with the idea that feedback 'is usually defined in terms of information about how successfully something has been or is being done' (1989, p. 120); yet went on to redefine feedback: 'in terms of its *effects* rather than its informational content …' (p. 120, emphasis added). 'If', as Sadler remarks, 'the information is simply recorded … or is too deeply coded (for example, as a summary grade given by the teacher) to lead to appropriate action, the control loop cannot be closed' (Sadler, 1989, p. 121). The assessment might be formative in purpose but it would not be formative in function. Indeed, Sadler's discussion of the control loop extended to defining the 'indispensable conditions for improvement' as being that 'the *student.*..is able to monitor continuously the quality of what is being produced *during the act of production itself*' (Sadler, 1989, p. 121).

Sadler's insights have been seminal in discussions of assessment. Feedback, therefore, can be seen either as a behaviourist response or a constructivist response. In the latter case, as Sadler suggests, the constructivist dimension relates to feedback that fosters 'new' or 'robust' understandings.

Black and Wiliam (1998), working at London University, have developed this position. Like Weiner, Elshout-Mohr and Sadler, they recognised that constructivist feedback requires a cognitive theory which links learners understanding, their interactions with assessment tasks and the design of such tasks.

> *In other words, the quality of the feedback provided becomes a feature of any procedure for formative or constructivist assessment (Black & Wiliam, 1998, p. 28).*

Yet, in regard to discussions about the quality of feedback, various commentators, including Shepard (2000), Torrance (1993) and McClellan, (2004) have identified problems that surround the concept and practice of alternative assessment. Earlier sections of this paper suggest that 'traditional' and 'alternative' assessment have had different histories, represent different views of teaching and learning and, not least, different views of feedback. But what does alternative assessment mean, and how can it be differentiated from earlier forms of assessment?

Alternative assessment for learning

The literature on alternative assessment is not a coherent body of knowledge. It has emerged piecemeal. It has been established more though bricolage than through the work of a single theorist. Nevertheless, the concept has spread throughout the world of educational research. For example the google search-engine generated over 50 pages of alternative assessment sites (2004-04-26) and the US data-base, ERIC (Educational Resources Information Center) yielded 7096 references for the boolean search 'alternative *and* assessment'. To understand the different elements of alternative assessment, it is useful to turn to recent reviews. One example is provided by Effie Maclellan from the University of Strathclyde. She suggests that, 'broadly speaking' alternative assessment is:

> *characterized as an alternative to standardized, norm-referenced, multiple-choice testing and typically claims the following features:*
>
> - *Student involvement in setting goals and criteria for assessment*
> - *Performing a task, creating an artifact/product*
> - *Use of higher level thinking and/or problem solving skills*
> - *Measuring metacognitive, collaborative and intrapersonal skills as well as intellectual products*
> - *Measuring meaningful instructional activities*
> - *Contextualisation in real world applications*
> - *Use of specified criteria, known in advance, which define standards for good performance. (McClellan, 2004, p. 312)*

The diverse literature of alternative assessment is both a response to the past (e.g. the behaviourism described by Shepard) and a response to the future – to the claim that the human species is moving towards a 'knowledge society', a notion that Torsten Husén sketched out in the 1970s:

> *Among all the explosions that have come into use as labels to describe rapidly changing Western society, the term 'knowledge explosion' is one of the most appropriate. Reference is often made to the 'knowledge industry', meaning both the produces of knowledge, such as research institutes, and its distributors, e.g. schools, mass media, book publishers, libraries and so on. What we have been witnessing since the mid-1960s in the field of distribution technology may well have begun to revolutionize the communication of knowledge within another ten years. (Husén, 1974, p. 239)*

This ideas has gain considerably official attention since the 1970s. It is argued that the new economic conditions of human life require learners who can operate effectively and can "go on learning throughout life, who are capable of coping with uncertainty, diversity and the need for collaboration with others' (Broadfoot, 2002, p. 1).

Shepard made the same argument, that new times require new assessment practices:

> *classroom assessment must change in two fundamentally important ways. First, its form and content must be changed to better represent important thinking and problem solving skills in each of the disciplines. Second, the way that assessment is used in classrooms and how it is regarded by teachers and students must change. The content of assessments should match challenging subject matter standards and be connected to contexts of application (Shepard, 2000, p. 11).*

Such a constructivist assessment paradigm will contain, according to Shepard, student self-assessment and feedback from peers, as a `central part of the social processes that mediate the development of intellectual abilities, construction of knowledge, and formation of students' identities' (p. 2). A constructivist assessment paradigm will look `to assessment as a source of insight and help instead of it being the occasion for meting out rewards and punishments' (p. 15). Assessment has to be more useful and helpful in the learning process and therefore must be moved into the middle of the teaching and learning process becoming `integral to Vygotsky's idea of a zone of proximal development' (p. 16).

In the foregoing analysis, there are at least two paradigms of teaching and learning and, with them, two models of assessment. Equally, as noted, the differences between these models of assessment are neither widely appreciated or understood. One way, however, to clarify the essential differences between behaviourist and constructivist assessment is by reference to three dualisms: summative/formative, high stake/low stake and divergent/convergent assessment.

Summative and formative assessment

Black and Wiliam (1998) provide a concise differentiation of formative from summative:

> *summative assessments are best thought of as retrospective. The vast majority of summative assessments in education are assessments of what the individual has learnt, knows, understands or can do…. In contrast formative assessments can be thought of as being prospective' (Wiliam, 2000, p. 14).*

Nevertheless, the distinction between formativ and summative assessment is not clear. Terms like performance assessment, authentic assessment and portfolio assessment are often used to describe activities which may, in terms of their social functions, be formative or summative. Other adjectives, however, may be more focused. For instance according to Black and Wiliam (1998) the terms 'classroom evaluation', 'curriculum-based assessment', 'feedback' and 'formative evaluation' are used as synonyms for formative assessment (p. 44-45). Likewise, Sadler comments that the 'etymology and common usage associate the adjective *formative* with forming or modeling something, usually to achieve a desired end.' (1989, p. 120). Further, Sadler calls attention to the cybernetic dimension of formative assessment. It is 'concerned with how judgments about the quality of student responses … can be used to shape and improve the student's competence by short-circuiting the randomness and inefficiency of trial-and-error learning.' (p. 120).

As already noted, the distinction between formative and summative relates not to the assessment information itself but to how that information is used (Wiliam, 2000, p. 11). If the teacher merely communicates such information to the learner, it is not legitimate to call the assessment `formative'. Yet if the learner understands and acts on the information then the assessment will be formative (p. 12). Further, Black and Wiliam claim that a teacher has a choice of two complementary options in formative assessment. The first is to develop the capacity of students to recognise and appraise gaps, leaving the student with the responsibility of planning and carrying out any remedial action that may be needed. The second option is for teachers to take responsibility themselves for directing the remedial activity which follows.

As noted, the quality of the feedback is central to the differentiation of formative from summative. Yet, in practice, these differences may be confused. Torrance (1993) asserts, in line with the argument of this paper, that formative assessment and summative assessment are so different that the same assessment system cannot fulfill both functions. Indeed, combining formative and summative assessment may adversely affect the quality of the feedback that learners receive**.** The multiple objectives of the assessment may prevent the provision of

useful feedback. Students may find it hard to focus on the developmental aspects of feedback if they are sidetracked by personal feelings concerning the mark they have achieved.

Wiliam (2000) points out that since schoolteachers are involved in both summative and formative assessment, it is relevant for teacher education to focus on such tensions between summative and formative assessment (p. 3-4). Practical problems that has arise from the fact that formative assessment is not well understood by teachers. Like Shepard, they suggest its implementation calls for deep changes both in teachers' perceptions of their own role and their classroom practice. Wiliam (2000) assert that years of constructivist learning and teaching may be completely distorted by summative assessments carried out at their conclusion (p. 3).

Black and Wiliam (1998) suggests that assessment must be seen as central to learning; students have to be active in their own assessment; and that students must be able to revise their own learning in the light of an understanding of what it means to get better (p. 22). If, therefore, assessment is an occasion for learners to learn, it is important to induct learners, through dialogue, into a discourse where they can recontextualize their understanding. Indeed Wells (1999) links this recontextualisation to the provision of educational scaffolding:

> *… one of the chief functions of the use of language in the classroom is to induct students into modes of discourse that provide them with frames of reference with which to 'recontextualize' their experience, and that it is this task that gives educational scaffolding its particular character (Wells, 1999, p. 127).*

High stakes and low stakes testing

Assessment can be used to intervene in someone's life. It may, that is, alter their life chances – in positive or negative directions. Summative assessment, then, is risky. It is therefore a high stakes activity. Its task is to identify winners, separate the sheep from goats, the bulls from bears (on the stock market). Everything depends on the calculation of a single score or narrow range of scores. These scores are reductionist. They are the condensation of evidence about performance. Accordingly they are 'deeply coded' (in Sadler's phrase), or 'hyper-real' in (Hanson's terms) (1993, p. 298). Their significance is symbolic;

and their impact can have social consequences that may nor be merit by the initial data (cf. Messick's discussion of consequential validity, 1989).

In 'high-stakes' assessments, as Wiliam (2000) points out, 'there is an incentive for teachers and students to concentrate on only those aspects of competence that are likely to be assessed' (p. 1). He adds: 'we start out with the intention of making the important measurable, and end up making the measurable important' (p. 1). The higher the stakes, the greater the pressure placed on teachers to teach to the test and to devote more and more time to prepare students to do well on the tests. As Shepard put it, "WYTIWYG" or "What You Test Is What You Get." (2001, p. 1082).

Robert Linn, President of AERA (2003-4) has made a comparable argument: 'the unintended negative effects of the high-stakes accountability uses often outweigh their intended positive effects' (2000, p. 14). Indeed, Linn goes on to make seven 'suggestions' that might enhance the 'validity, credibility and positive impact of assessment and accountability systems'. These include 'don't put all of the weight on a single test', 'recognize, evaluate, and report the degree of uncertainty in the reported results' and 'put in place a system for evaluating both the intended and unintended positive effects and the more likely unintended negative effects of the system (p. 15).

Low stakes assessment, on the other hand, is significant because, ideally, its side-effects are zero; and, instead, only the desired effects are promoted. That is, it is designed to have a positive effect on the life chances of those who are assessed. It is low stakes because while it provides a characterization of the learner, it is not designed to discriminate among those who are tested in ways that are socially significant. It is about learning continua, not learning categories.

Divergent and convergent assessment

This distinction is known from the work of Harry Torrance and John Pryor, as reported in *Investigating Formative Assessment* (1998). Their claim is that two conceptual models of assessment can be identified on the basis of teacher's views of learning and the process of intervening to support learning. To clarify these differences they used the labels *convergent* and *divergent* assessment.

The task in convergent assessment is to find out whether the student has a predetermined knowledge, understanding or skill. It employes tick-lists and can-do statements. It prefers pseudo-open questioning and it focuses on contrasting errors with correct responses. Assessment with these characteristics can also be described as behavioristic. Using suitable probes, the teacher elicits the knowledge, understanding or skills of the learner.

By contrast, divergent assessment has students' understanding in focus. It is goal-free. It is not constrained by predetermined knowledge, understanding or skill. It aims to find out what a student knows or can do. The assessment is performed by the teacher and the student working together (cf. dialogue). Divergent assessment is characterize by flexible planning, open forms of recording, emphasizing the learners understanding, open tasks, open questioning and descriptive, qualitative feedback. In effect, it takes place in the student's zone of proximal development, and is closely integrated into the school curriculum.

It is not claimed, however, that teachers should always use divergent assessment. Both forms of assessment serve different purposes. The argument of assessment theorists, like Torrance and Wiliam, is that school teachers and other educationists should be aware of the differences.

Conclusion

This paper has explored conceptions of teaching, learning and assessment. It has discussed issues surrounding the notion of alternative assessment; it has noted problems associated with clarifying the differences between 'alternative' and 'traditional' assessment; and, it has sought to resolve these problems by focusing on the integrity of 'constructivist assessment' – something that is quite distinct from 'traditional' assessment.

The confusion surrounding the field of assessment has several origins. In part, it arises because much of the discussion is driven by educational ideals – especially the notion of the learning society. Thus, assessment is bothered by what counts as the learning society. For instance, is the learning society the same as the knowledge society? In addition, confusion arises because assessment cannot be separated from its behaviourist past. Common sense assumptions about the selective role of examinations and human differentiation still dominate discussions about the reform of schooling, not only in Sweden but also throughout Europe (cf. the Bologna process). And finally, confusion arises because much discussion relates to assessment theory rather than assessment practice. As Ulf P. Lundgren noted, there is always a gap between 'hope' and happening', between 'text' and 'context' (1983).

The idea that assessment should be moulded to the designs of the learning society is explicit in the work of three theorists working in the United Kingdom: Lee Harvey (1997), Caroline Gipps (1994) and Patricia Broadfoot (2002). Harvey argues for a transformation of assessment in higher education, moving the emphasis from teaching to learning, and developing assessment procedures that encourage transformative learning (p. 66). Gipps (1994) has suggested a

paradigm shift is needed to transform the current assessment paradigm (summative assessment) to one that prioritises the formative functions of assessment. This, she suggests, will lead to the replacement of a testing culture by an assessment culture and a shift from psychometrics to the assessment of learning (chapter 9). Finally, Broadfoot (2002) identifies a key challenge for assessment in the 21st Century is to turn it into 'the gearbox in the engine of personal reflection and target setting that is at the heart of effective education'. She also points to clarifying and harnessing technology's power in support of effective learning. In the 'learning society', she concludes, teachers become key agents of change - as 'empowered, reflective and active learners, rather than passive respondents to external dictats' (p. 2).

The recognition that assessment cannot be entirely liberated from its past underpins the arguments of Black and Wiliam (1998). They argue that, for a variety of reasons, formative and summative assessment will continue to operate in parallel.

> *… there is no such thing as a formative assessment. The formative-summative distinction applies not to the assessment itself, but to the use to which the information arising from the assessment is put. The same assessment can serve both formative and summative functions, although in general, the assessment will have been designed so as to emphasise one of the functions (Wiliam, 2000, p. 11).*

They also note, however, that summative and formative assessment emphasis different communities of practice:

> *This notion of 'understanding the standard' is the theme that unifies summative and formative functions of assessment. Summative assessment requires that teachers become members of a community of practice, while formative assessment requires that the learners themselves become members of the same community of practice (Wiliam, 2000, p. 12).*

Yet, as this suggests, the rise of formative assessment will require a paradigm shift in teachers views of learning and in their own classroom practice (p. 13). In effect, Black and Wiliam recognize that, necessarily, teachers are engaged with both formative and summative assessment. Teachers and learners need to

develop the capacity to interpret and respond to assessment results in a formative - or cybernetic - manner (p. 29), just as students must acquire the ability to assess and develop knowledge for themselves. As Lee Harvey put it students must be considered as 'principal stakeholders' in the learning society (1997, p. 66).

The third position described above - the gap between 'hope' and happening' - has been discussed by Torrance (1993) and Black and Wiliam (2003). Torrance discussed the problems associated with changes in assessment policy in England and Wales and, ten years later, Black and Wiliam report their own efforts at putting these changes into practice. In both cases, these authors recognize that efforts to reform assessment in the direction of formative assessment run the risk of reinforcing the hegemony of summative assessment.

Torrance's early work developed into efforts to explore teachers assessment practices (Torrance & Pryor, 1998), just as Black and Wiliam invested energy in 'putting it [Assessment for learning] into practice' (Black et al., 2002). The latter group, for instance, have emphasise the importance of 'supported development' for teachers; and they suggest how formative assessment could be the integrated into the every-day practices of teachers – through attention, for instance, to 'rich questions, comment-only marking, sharing criteria with learners, and student peer- and self-assessment' (2003, p. 630; see also Black *et al.*, 2002, chapter 4).

To conclude, this paper has been written at a time when educational assessment is in a state of flux –lacking in 'articulated' formulations (Torrance, 1993, p. 339) or 'consensus' (Maclellan, 2004, p. 312). Moreover, this unease has arisen because formative and summative assessment are linked together as subsets of a more general activity – assessment. Besides giving a sense of the historical origins of this reclassification of assessment, this paper raises *other* taxonomic issue – that formative assessment, itself, implies a reassessment of the boundaries that, hitherto, have separated teaching from assessment (cf. Bernstein's, three message systems: curriculum, pedagogy and evaluation, 1971). Perhaps this will turn out to be the most significant feature of the learning society.

References

Bakker, S. (2001). Editorial: Examinations and entry to university: pressure and change in a mass system. *Assessment in Education: Principles, Policy & Practice, 8* (3), 6-7.

Bangert-Drowns, R.L., Kulick, J.A. & Morgan, M.T. (1991) The instructional effect of feedback in test-like events, *Review of Educational Research*, 61, pp. 213-238.

Bernstein, B. (1971). On the classification and framing of educational knowledge. In M. F. D. Young (Ed.), *Knowledge and Control: New directions for the sociology of knowledge* (pp. 47-115). London: Collier-Macmillan.

Black, P. (2001). Dreams, strategies and systems: Portraits of assessment past, present and future. *Assessment in Education*, 8(1), 65-85.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for Learning: Putting it into practice*. Maidenhead: Open University Press.

Black, P., & Wiliam, D. (2003). 'In praise of educational research': Formative assessment. *British Educational Research Journal, 29*(5), 623-637.

Broadfoot, P. (2001). Editorial. Assessment for Lifelong Learning: challenges and choices. *Assessment in Education: Principles, Policy & Practice, 8* (3), 5-7.

Bruner, J. S. (1968). *Toward a Theory of Instruction*. New York: Norton.

Daniels, H. (2001). *Vygotsky and Pedagogy*. London: RoutledgeFalmer.

Duffy, T. M., and Cunningham, D. J. (1996). *Constructivism: Implications for the design and delivery of instruction. In D. H. Jonassen (Ed.),* Handbook of Research for Educational Communications and Technology (pp. 170-198). New York: Simon Schuster Macmillan.

Elshout-Mohr, M. (1994) *Feedback in self-instruction.* European Education; Summer94, 26 (2), 58-73.

Gipps, C. (1994). *Beyond Testing: Towards a theory of assessment.* London: RoutledgeFalmer.

Glasersfeld, E. v. (1979). *Cybernetics, Experience, and the Concept of Self in Mark N. Ozer (Ed.),* A Cybernetic Approach to the Assessment of Children: Toward a More Humane Use of Humane Beings. Westview Press / Boulder, Colorado.

Hanson, F. A. (1993). *Testing, Testing. Social consequences of the examined life.* University of California Press, Berkley , Los Angeles Press.

Harvey, L (1997). Transforming higher education: Students as key stakeholders. Quality assurance as support for processes of innovation. Högskoleverkets skriftserie 1997:1 S. http://helios.hsv.se/publikationer/skrifter/skrift97.html

Husén, T. (1974). *The Learning Society*. London: Methuen.

Kliebard, H. M. (1995). *The Struggle for the American Curriculum (2nd ed.).* New York: Routledge.

Linn, R. L. (2000). Assessment and accountability. *Educational Researcher*, 29(2), 4-16.

Lundgren, U. P. (1983). *Between Hope and Happening: Text and context in curriculum*. Geelong: Deakin University Press.

Maclellan, E. (2004). How convincing is alternative assessment for uses in higher education. *Assessment and Evaluation in Higher Education, 29*(3), 311-321.

Marton, F. & Booth, S. (1997). *Learning and Awareness.* Mahwah, N J: Lawrence Erlbaum.

Mayr, O. (1970). *The Origins of Feedback Control*. The M.I.T. Press. Cambridge, Massachusetts, and London, England.

McCorduck, P. (1979). *Machines Who Think: A personal inquiry into the history and prospects of artificial intelligence*. San Francisco: Freeman.

Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35(11), 1012-1027.

Messick, S. (1989). Validity. In R. LINN (ed) *Educational Measurement* (3rd edition, 13-103) (New York, Macmillan).

Ramaprasad, A. (1983) On the definition of feedback, *Behavioral Science*, 28, pp. 4-13

Ramsden, P. (1992) *Learning to Teach in Higher Education.* London: Routledge.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18,* 119-144.

Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education,* 5(1), 77-84.

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

Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.

Shepard, L. A. (2001). The role of classroom assessment in teaching and learning. In V. Richardson (Ed.), *Handbook of Research on Teaching* (4th ed.). Washington, DC: American Educational Research Association.

Torrance, H. (1993). Formative Assessment: some theoretical problems and empirical questions. *Cambridge Journal of Education, Vol. 23, No. 3, 1993*, 333-343

Torrance, H. (Ed.). (1995). *Evaluating Authentic Assessment*. Buckingham: Open University Press.

Torrance, H., & Pryor, J. (1998). *Investigating Formative Assessment: Teaching learning and assessment in the classroom*. Buckingham: Open University Press.

Torrance, H., & Pryor, J. (2001). Developing formative assessment in the classroom: using action research to explore and modify theory. *British Educational Research Journal*, 27(5), 615-631.

Utbildingsdepartement (2004) *Högre utbilding I utveckling: Bolognaprocessen I svensk untlysning* (Ds 2004:2). Stockholm: Utbildningsdepartementet.

Wells, G. (1999). Dialogic Inquiry: Towards a socio-cultural practice and theory of education. Cambridge: Cambridge University Press.

Wiliam, D. (2000). An overview of the relationship between assessment and the curriculum. In D Scott (Ed.) (2000) *Assessment and the Curriculum. Greenwich, CT: JAI Press*

Wiener, N. (1965) *Cybernetics: Or control and communication in the animal and the machine* (2nd ed.). Cambridge, MA: MIT Press.

Wiener, N. (1968). *The Human Use of Human Beings*. London: Sphere Books [reprint of 1954 ed.].