



Pergamon

Studies in Educational Evaluation, Vol. 23, No. 4, pp. 279-298, 1997

© 1997 Elsevier Science Ltd

Printed in Great Britain. All rights reserved

0191-491X/97 \$17.00 + 0.00

S0191-491X(7)00018-7

INTRODUCTION

ASSESSMENT AS A TOOL FOR LEARNING

Filip J.R.C. Dochy* and Liz McDowell**

**Centre of Teacher Training and Research, University of Leuven, Belgium*

***Centre for Advances in Higher Education, University of Northumbria at Newcastle, UK*

The role of assessment and evaluation in education has been crucial, probably since the earliest approaches to formal education. However, change in this role has been dramatic in the last few decades, largely due to wider developments in society. The most dramatic change in our views of assessment is represented by the notion of assessment as a tool for learning, and this special issue will focus on several pertinent examples. Whereas in the past, we have seen assessment only as a means to determine measures and thus certification, there is now a realisation that the potential benefits of assessing are much wider and impinge on in all stages of the learning process. Portfolio assessment, self-assessment and peer-review are some forms of assessment which encourage students to engage continuously and foster a deep approach to learning. Key elements of these approaches are reflection, feedback, and integration of learning and assessment. In this contribution, we will outline some of the major developments in the learning society and their implications for educational practice, and we will reflect on the future of education within powerful learning environments, where learning, instruction and assessment are fully integrated.

Recent Developments in the Learning Society

The Information Age

As far back as in classical Greece, scientists were considered experts in all sciences and areas of knowledge. Until a decade ago, some scientists could claim a full grasp of all

knowledge within a certain discipline. However, this era has today definitely passed. Information is now exchanged very rapidly and knowledge is growing at an exponential rate. There is no possibility of individual scientists possessing all knowledge within their discipline. Scientists currently and in the future, need to master the basic knowledge of their field and then the skills necessary to navigate around their discipline. This shift is also true for teachers, who have traditionally been accepted as the sources of all knowledge and experience within the educational process. The teacher now is better represented as a key to open the door to domains of knowledge and experience. The information age, leading into the 21st century, is characterized by an infinite, dynamic and changing mass of information. Successful functioning in this era will require specific skills:

- a) cognitive competencies such as problem solving, critical thinking, formulating questions, searching for relevant information, making informed judgements, efficient use of information, conducting observations, investigations, inventing and creating new things, analysing data, presenting data communicatively, oral and written expression;
- b) meta-cognitive competencies such as self-reflection, or self-evaluation;
- c) social competencies such as leading discussions, persuading, cooperating, working in groups, etc. and
- d) affective dispositions such as for instance perseverance, internal motivation, self-efficacy, independence, flexibility, or coping with frustrating situations. (Birenbaum, 1996, p.4)

Internet, Multimedia and Educational Technology

As we enter a new era of technological possibilities, education will use the electronic highway in all its forms. Electronic forms of communication are already intensively used in academic contexts. The use of multimedia, local area networks, shared communication systems, the Internet, shared electronic databases, video conferencing facilities, electronic self-study materials, study support and guidance through networks, progress assessment systems, intake and monitoring systems, and so on, will lead to the development of new teaching and learning strategies. Audio, for example, is indispensable for teaching and learning when music or spoken language is a major focus. Video can be used when surrogate experiences must be presented (e.g. exotic rituals, dangerous physical phenomena, animations of invisible phenomena), or when dynamic processes or phenomena cannot be adequately presented in printed material (e.g. plasma currents). The computer can be seen as indispensable not only in testing and calculating, but also for simulating complex processes and systems (e.g. economic or physical models) and for practice in skills (e.g. solving equations). Software that supports interactive assessments based on different MC formats is already available (e.g. CAT system, Examiner). The use of other electronic devices, such as CD-ROM movies, audio or pictures within such systems is possible (e.g. in Question Mark). Very soon, Internet-based information is likely to be used in assessment software, and test formats will be expanded to include items such as cases, 3D events, problems and simulations. Course development software is likely to be integrated with high quality assessment software, so-called "integrated learning and assessment systems", in order to support assessment during learning. This will include prior knowledge assessments and assessment of progress as in the example of software

called Integrative Testing and Electronic Media (ITEM) and the Mercator course development software at our department (Dochy & Moerkerke, 1997).

Changing Labor Market

Recently, there has been rapid change in the labor market demanding a more flexible labour force with increased short-term, part-time and casual working. There is increased pressure from industry for higher education to deliver graduates who are immediately employable and effective in business and industry (Moerkerke, 1996). Employers organizations have pointed to an insufficient match between the outcomes of study programmes and labor market needs (Harvey & Green, 1994). If vocational and academic education systems are required to supply graduates who can be utilized immediately within the labor force, this implies a growing need for assessment procedures for relevant skills (De Rijke & Dochy, 1995). Higher education is being required to prove, through the outcomes of assessment procedures, that graduates are equipped for the labor market. Institutes for vocational training are developing assessment systems for vocational qualifications which encompass learning outside formal education and training (Jessup, 1991). Educational measurement specialists (Birenbaum, 1996; Shavelson, Xiaohong, & Baxter, 1996) have recognised that, in the near future, effective and efficient systems for performance assessment will have to be developed.

The Student as a Consumer

Free market ideology and strategies have recently entered the world of education. More and more initiatives are being taken at governmental levels to promote a more demand-driven approach education, aiming to balance the demands from industry and the preferences of students. Institutes of higher education are increasingly having to compete with each other to recruit students, and as a result the needs and wishes of students are receiving more attention, thus placing students in the role of consumers (Green, Brannigan, Mazelan, & Giles, 1994).

Lifelong Learning

In the last decade or so, there has been an increasing focus on training in industry and business. Economic pressures requiring major restructuring in the labor market were major factors leading Government and employers to emphasize the importance of adaptability within the labor force and the ability of employees to acquire new skills throughout their working lives. It is widely accepted that the need for lifelong learning will increase even more rapidly in the near future (Moerkerke, 1996). Research in the Netherlands has indicated that the highly technical society of tomorrow will require people who are flexible and able to continue to acquire new knowledge and learn new skills (De Rijke & Dochy, 1995).

Innovations in Assessment: Consequences of the Developments in Society

From Testing Towards Assessment

Birenbaum (1996) has made a useful distinction between two cultures in the measurement of achievement and relates them to developments in the learning society. In the traditional so-called testing culture, instruction and testing are considered to be separate activities. Instruction is the responsibility of the teacher, whereas testing is the responsibility of the psychometric expert, who can use elaborate procedures for test development and sophisticated psychometric models for the analysis of test responses. The work of such experts within testing agencies was stimulated by the demand for objectivity and fairness in testing and a high level of standardization where important decisions were to be based on test scores. Stemming from such demands, common tools for assessment in education were easy-to-score items such as multiple-choice items and true/false items. The testing culture fits well with the traditional approach to education where teaching is seen as an act of depositing the content which students receive, memorize and reproduce. The most efficient way of testing this process is by standardized tests delivered by external testing agencies.

The changing learning society has generated the so-called assessment culture as an alternative to the testing culture. The assessment culture strongly emphasizes the integration of instruction and assessment. Students play far more active roles in the evaluation of their achievement. The construction of tasks, the development of criteria for the evaluation of performance, and the scoring of the performance may be shared or negotiated among teachers and students. The assessment takes all kinds of forms such as: observations, text- and curriculum-embedded questions and tests, interviews, performance assessments, writing samples, exhibitions, portfolio assessment, and project and product assessments. Several labels have been used to describe subsets of these alternatives, with the most common being *direct assessment*, *authentic assessment*, *performance assessment* and *alternative assessment*. From a psychometric point of view, they can be characterized as badly standardized assessments that depend heavily on instruction. In these new contexts, assessment will be used more frequently, and will have different functions, different goals, and different administration techniques in addition to new assessment forms.

New Roles of Teachers in New Learning Environments

In a testing culture, professors were seen as carriers of knowledge which they had to transfer into the students' heads. According to Birenbaum (1996), the assessment culture is in accord with the constructivist approach to education. In this approach, learning is viewed as a process through which the learner creates meaning. The teacher is not a person who transfers knowledge, but a mentor who provides opportunities for learners to use the knowledge and skills they already possess in order to understand new topics. The teacher is expected to provide interesting and challenging tasks. When recent developments such as the use of new educational technological tools are taken into account, it seems likely that changes will even go further. The traditional professor had to

fulfil a range of different roles such as those of course material developer, student tutor, provider of feedback, guide through the learning process, tests administrator, test corrector, and so on. In the future some of these tasks will be carried out by new highly specialized professionals such as the item-bank manager, electronic course-materials developer, multi-media development expert or study counselor.

These changes are taking place with increasing moves towards what we call *Powerful Learning Environments* (PLEs). A number of characteristics of learning which are relevant to the design of PLEs, have emerged from recent research on learning and teaching (see De Corte, 1990, 1991). Two major aspects are: the constructive nature of learning, and the necessity to anchor learning in real life situations and contexts. A PLE is characterized by a good balance between discovery learning and personal exploration on the one hand, and systematic instruction and guidance, on the other. It also takes into account the students' individual differences in abilities, needs, and motivation. Students' constructive learning activities should preferably be embedded in contexts that are rich in resources and learning materials, that offer ample opportunities for social interaction, and that are representative of the kinds of tasks and problems to which the learners will have to apply their knowledge and skills in the future.

Towards Assessing Skills and Competencies

From all over the world there has been criticism of using standardized tests as measures of either learning or competence (Broadfoot, 1984; Darling-Hammond, 1994; Frederiksen, 1984; Kellaghan, Madaus, & Airasian, 1982; Linn, 1981; Rowe & Hill, 1996; Wigdor & Garner, 1982; Wolf, 1995). Newmann and Archbald (1990) argue that "most data fail to measure meaningful forms of human competence and that significantly new forms of assessment need to be developed" (p. 164). As a reaction to large-scale objective testing in the US, alternative approaches were investigated (e.g. Birenbaum & Dochy, 1996; Floden, 1994; Lesh & Lamon, 1992; Nitko, 1995; O'Connor, 1992; Resnick & Resnick, 1992; Shavelson, Xiaohong, & Baxter, 1996). All these attempts illustrate the development of more "in context" and "authentic" approaches to assessment (Archbald & Newmann, 1992; Hill, 1993). Rowe and Hill (1996) identified the key features of these more recent systems as:

- a concern for better methods of assessing and reporting on the actual and developing achievements and competencies of individual students;
- the development of national or system-wide standards frameworks for assessment, recording and reporting student achievement;
- the use of a broad range of assessment methods, including "common" or "standard" tasks undertaken by students; and
- a significantly enhanced role for teacher observation and judgement in the assessment of competency in relation to these tasks. (p.312)

Nisbet (1993) defines the term *authentic assessment* as

methods of assessment which influence teaching and learning positively in ways which contribute to realising educational objectives, requiring realistic (or

"authentic") tasks to be performed and focusing on relevant content and skills, essentially similar to the tasks involved in the regular learning processes in the classroom. (p. 35)

There is general agreement that skills such as communication, information technology, and personal or interpersonal skills such as working with others, improving own learning and performance, and problem-solving are important. However, there are differences in terminology and the conceptualisation of such skills. In higher education, the word "skill" is emotive and tends to be used as a way of drawing distinctions between education and training, and between academic and vocational aims.

Skills are thought to imply activities low in cognitive content that are typically learned through rote practice and therefore inappropriate in higher education. It is not surprising, therefore, that many... chose to avoid the connotations of skills and to employ the term "competence" to describe the attributes they sought to promote. (Otter, 1995, p.46)

The term *capability* has also been used (Clarke & Engel, 1986; Stephenson, 1994) to emphasize the holistic notion of a capable professional rather than a person with a collection of acquired skills.

The Use of Co-, Self- and Peer-Assessment

A significant new form of assessment is the use of self-, peer- and co-assessment. The type of student self-assessment referred to most frequently in the literature is a process which involves teacher-set criteria and where students themselves carry out the assessment and marking. Another form of student self-assessment is the case where a student assesses herself or himself on the basis of criteria which she or he has selected, the assessment being either for the student's personal guidance or for communication to the teacher or others. According to Hall (1995) there are two critical factors for genuine self assessment: the student not only carries out the assessment, but also selects the criteria on which the assessment is based. Similarly, peer-assessment can indicate that fellow students both select the criteria and carry out the assessment. Any situation where the tutor and students share in the selection of criteria and/or the carrying-out of the assessment is more accurately termed co-assessment (Hall, 1995). However, it is still frequently the case that teachers control the assessment process, sometimes assisted by professional bodies or assessment experts, whereas students' assessments and criteria are taken seriously but considered to be additional to the assessment undertaken by the teacher or professor rather than replacing it (Rogers, 1995).

Implementing forms of self-, peer- and co-assessment may decrease the time-investment professors would otherwise need to make in more frequent assessment. In addition to that advantage, using these assessment forms assists the development of certain skills for the students, e.g. communication skills, self-evaluation skills, observation skills, self-criticism.

Different Functions of Assessment in the Learning Process

It has been demonstrated that assessment has a significant impact on instruction and learning. A number of authors have reported negative effects of assessment on teaching and learning (Frederiksen, 1984; Ridgway & Schoenfeld, 1994). Poor assessment practices can often be held responsible for low quality instruction and learning. According to Boud (1990), in many cases:

assessment tasks are set which encourage a narrow, instrumental approach to learning that emphasizes the reproduction of what is presented, at the expense of critical thinking, deep understanding and independent activity. (p.104)

Tests that are inadequately linked to instruction have led to undesirable consequences such as inappropriate information about learning progress and learning difficulties, reduction of students' motivation for learning, and incorrect evaluation of the effectiveness of instruction. Negative effects have been reported, even with tests well-linked to instruction, due to too frequent formal testing (Tan, 1992). In the United States, where student test performance was linked to inferences about schools and teacher competency, schools and teachers narrowed their curricula and courses with the aim of helping students pass tests from external agencies (Baker, 1994).

For effective assessment, there is a need to reconceptualize conventional assessment instruments and the theory underlying them (Glaser, 1990). Such instruments are not always congruent with current research-based conceptions of the acquisition of knowledge and skills. As Masters and Mislevy (1993) have observed:

Most items on standard achievement tests assess students' abilities to recall and apply facts and routines presented during instruction. Some require only the memorization of detail; they seek evidence that students have absorbed factual details presented in class and are able to reproduce these on command. Other achievement test items, although supposed to assess higher-level learning outcomes like "comprehension" and "application", often require little more than the ability to recall a formula and to make appropriate substitutions to arrive at a correct answer. (p. 219-220)

Many argue that sound assessment practices can be used to improve instruction. Most assessment specialists, like Birenbaum and Dochy (1996), Nitko (1989), and Keeves (1994), take the position that appropriately used educational assessments are potent tools that enhance the instructional process. An important condition for appropriate test use is that tests are linked or integrated both with instructional materials and procedures (Nitko, 1989). Nitko developed a number of prescriptions for sound assessment practice in relation to instruction. One of the main issues is clarity about the type of instructional decision that is supported by the assessment procedure. Nitko identifies four types of instructional decisions that are supported by tests:

Placement decisions: deciding in which instructional sequence or at what level in the instructional sequence a student should begin in order to avoid unnecessary repetition of what is already known and to allow more rapid attainment of new goals. Placement or prior knowledge state tests are given before a student begins a unit of instruction. They may focus on the prerequisite knowledge and skills for proposed learning or on the outcome knowledge and skills to determine whether students have already attained the desired outcome of the instruction. Thus, instructors use prior knowledge state tests to place students in instructional modes by (a) determining the degree to which prerequisite entry behaviors or skills are present or absent, (b) determining entering mastery of course objectives, and also for (c) matching students to alternative instructional modes, based on student characteristics.

Monitoring decisions: deciding (a) whether students are learning appropriately, and (b) whether the assigned learning activity is working effectively, or a different activity should be assigned. So-called progress assessments or assessments of growth are the instruments for the assessment of a student's progress toward the final learning objectives. Progress assessments provide feedback to students on how their learning is progressing and can lead to the (formative) decision that there is a need to remediate or relearn a certain instructional unit. Progress assessments are used by instructors to (a) choose or modify subsequent learning activities, (b) prescribe remediation of group or individual deficiencies, and (c) provide on-going feedback to the student for the purpose of directing advanced or remedial study. Progress tests are taken during instruction, but measure student behaviors and cognitive processes at the level of the desired outcomes.

Attainment decisions: deciding, at the end of a particular instructional segment, whether a student has attained the desired instructional goals. Final tests are used for certifying student learning, and evaluating teacher effectiveness.

Diagnostic decisions: deciding which learning outcomes a student has not acquired and the probable cause(s) of the failure to acquire them, in order to remediate or correct incomplete or erroneous prior learning. Diagnostic assessments are designed to provide specific information on individual learning deficiencies and misunderstandings. When prior knowledge state assessments, progress assessments, and final assessments are designed and interpreted properly they can also provide information of a diagnostic nature since students and teachers can use this information in order to regulate learning processes.

Masters and Mislevy (1993) also remark that the scoring of responses to assessment items as either "right" or "wrong", is in accordance with the additive and incremental view of the learning process. Although this conception of learning is applicable, to some degree, in some domains and under certain circumstances, it is not the appropriate view with respect to the most important objectives of present-day education since these objectives relate to understanding and problem solving. It is also inappropriate to forms of learning that involve the construction of meaning by the student, and the development of strategies for approaching new problems and learning tasks. Therefore, new types of instruments are required which allow the assessment of qualitatively distinct levels of

understanding (or misunderstanding), as well as strategic differences in learners' approaches to unfamiliar problems and challenging learning situations. The need for these new types of instruments is especially obvious when we aim for a better integration of instruction and assessment (Dochy, 1992). Indeed, due to their static and product-oriented nature, traditional achievement tests not only lack diagnostic power but, as a consequence, they also fail to provide relevant information to assist in adapting instruction appropriately to the needs of the learner (Campione & Brown, 1990; Dochy, 1994).

Standard Setting and the Use of Profiles

Hambleton and Sireci (1997) argue that criterion-referenced standard setting procedures should be preferred to norm-referenced standards. However, we believe that, in new assessment and instructional contexts, a range of developments in standard-setting methods will be needed rather than a shift to criterion-referenced standards, although they may provide one starting point. There will also be developments in norm-referenced standards, and there are promising developments towards knowledge and skill profiles, alternatively termed subject or score profiles.

Firstly, we may need to move away from criterion-referenced standards with the increasing use of some forms of student-centered learning. For example, where students negotiate to determine their own learning goals, it becomes rather difficult to formulate standards which cannot be adapted to the individual students or the group. Nor will norm-referenced standards be appropriate in every case. For example, in new learning systems such as problem-based learning, there is always a reasonable probability that one of the following situations will arise: students themselves decide not to put very much effort into a certain part of a course; assessment results are so low that simply guessing would lead to much the same score; results are so high that they do not discriminate students' levels of achievement; students do not study parts of the course because they know that there is a relative standard (Wijnen, 1990). This may result in using criterion-referenced standards together with a norm to be used if too many or too few students pass. Such compromise-oriented methods have been elaborated by Berk (1986) in situations where the course content or skills are very difficult to assess and the assessment seems to be too easy or too difficult.

Secondly, starting from a student-oriented learning system, a clear way forward is in the use of knowledge and skill profiles. Baumgart (1986) notes that the term profile has long been used in the literature of educational measurement and evaluation "to describe the presentation of an individual's achievements or characteristics on multiple dimensions" (p. 42). In recent years, according to Rowe and Hill (1996), the concept has acquired a particular meaning related to the reporting of students' academic and other achievements including personal qualities, in the form of specific, disaggregated statements or checklists of students' achievements and attributes. Different forms have appeared over the last ten years and terms such as knowledge profiles, skill profiles, subject profiles, student profiles or records of achievement, are all used in this sense (Assiter & Shaw, 1993; Dochy, 1996; Mansell, 1986). Experiences with other profiles in education have shown that profiles are useful instruments for communicating assessment results (Birenbaum, 1994). Subject profiles, as used by Rowe and Hill (1996), refer to a framework of empirically-calibrated,

descriptive performance indicators, located on a developmental growth continuum, designed to assist teachers and schools with assessing, recording, reporting and monitoring students' progress. Rowe and Hill (1996) propose three core elements in their framework for a given subject: "areas of knowledge and skills; aspects within them to be assessed, recorded and reported on; and levels of achievement and competency which describe and illustrate both attainment and developing competence" (p.318).

To provide more detailed analysis of the complex components of knowledge states, Dochy (1992) introduced so-called "knowledge profiles". A profile is a plot, in the form of a graph, of the raw or standardized scores of a group or individual on certain parameters (Keeves, 1988). For each person or group, a knowledge profile can be obtained by combining the test results on a set of parameters. The comparison between individual profiles is known by the generic term *profile analysis*. The term *knowledge profile* is used when achievement scores (e.g. on prior knowledge state assessments, or progress assessments) are analyzed following the parameters along a dimension and when results of this analysis are plotted on a graph. Figure 1 shows an example of a knowledge profile. In Figure 1, the relationships between some key concepts are illustrated. A dimension is used to construct a knowledge profile. Each dimension, consisting of several parameters, represents an approach to the structure of knowledge (Dochy, 1996).

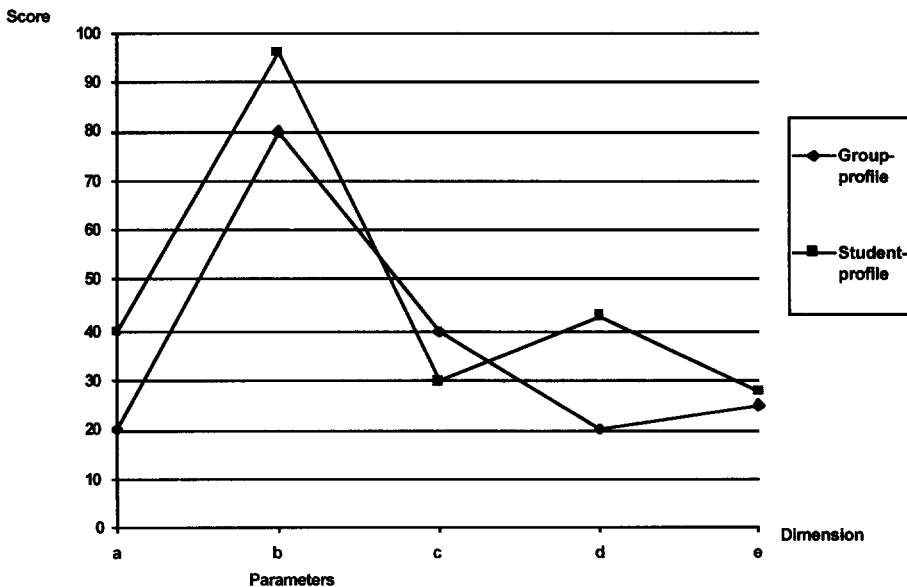


Figure 1: Example of a Knowledge Profile on a Single Dimension (adapted from Dochy, 1996, p. 236)

Knowledge profiles can give practical indications of student achievement and learning, thereby making it possible to support the learning process (Dochy, 1992; 1993). Wolf, Bixby, Glenn and Gardner (1991) also advocated this approach and, in an overview of

student assessment, stressed the need for a new brand of educational psychometrics, to match more developmentally oriented assessments. Dochy (1996) implemented this idea by trying to identify multiple components of the prior knowledge state, by using prior knowledge state tests, and then using the same kind of tests as progress tests administered several times during the year. In this approach, the analysis of prior knowledge suggests "that in situations where there are significant differences between the prior knowledge state of specific subpopulations, the profile dimensions will help to detect and dissect in more detail the strengths and weaknesses of the students involved" (Dochy, 1992, p. 185). This could be a promising starting point for differentiated diagnostic and guidance approaches both for students who have to select study modules or programs (Ager, 1993) and for instructors who have to provide students with the most optimal study paths (Lee, Pliskin, & Kahn, 1994).

The Future: Integrating Learning, Instruction and Assessment Through Flexible and Transformative Learning in *Powerful Learning Environments*

Formerly, and perhaps still today, professors were sometimes viewed as alligators swimming in the deep water waiting to catch the unwary student who slips whilst undergoing testing. Nowadays there is a movement to reduce this threatening aspect of testing and to view assessment as a means to assist students and teachers to reach their goals. Will the future bring us as far as an "overall assessment prophecy" would predict ?

It seems certain that teaching will develop into what we might call activities that support student transformative learning. Within another decade, with the use of more formative assessment in classrooms acting as a lever, we may see very much more integration of assessment and learning.

The Overall Assessment Prophecy?

For many years, Glaser (1981, 1990) has been an advocate of what we call the overall assessment prophecy. This prophecy holds that it is no longer possible to consider assessment only as a means of determining which individuals are already adapted to or have the potential to adapt to mainstream educational practice. An alternative goal is to reverse this sequence of adaptation; rather than requiring individuals to adapt to means of instruction, we may adapt the conditions of instruction to individuals in order to maximize their potential for success.

This objective can be realized if learning can be designed to take account of the individual's profile of knowledge and skills (Pellegrino & Glaser, 1979), but we first need to know how to achieve that end. Should we aim at assessment-driven instruction or at instruction-driven assessment? It is certainly the case that much educational practice is assessment-driven; teachers do teach to the test. However, this has proved to be problematic. It has been shown that teaching to the test leads, in many cases, to a sacrifice in the depth of the learning of both knowledge and skills. Specifically, as Hambleton (1997) argues, when multiple-choice items are used in the context of high-stakes assessments, teachers tend to emphasize the memorization of isolated factual

information suitable for this type of test format at the expense of, for example, high-order problem-solving skills.

Our view is that assessing high-order skills by means of authentic assessments will lead to the teaching of such high-order knowledge and skills. We agree with Knight (1996) that instructional reform towards the overall assessment prophecy can be attained by a careful reform in assessment. Several authors have argued (Birenbaum & Dochy, 1996) that the issue of test-driven instruction takes on a different flavour when associated with the new ideas of alternative, performance, or authentic assessment. Firstly, the assumption is made that alternative assessment will be instruction-driven by using authentic test items, directly related to instruction. Secondly, there is an assumption that alternative assessment will have a positive backwash effect (Biggs, 1996) on instruction, making it better related to real life experiences and resulting in more active learning.

Support for Student Learning by Means of Formative Assessment

Knight (1996) identified the key advantages of formative assessment.

- It involves giving learners feedback that can be used to improve the next performance.
- It is directly connected to the transformational notion of quality, since it is about improving student learning.
- It does not have to be conducted solely by tutors; co-, peer- and self-assessment are possible and, arguably, necessary.
- It is not high-stakes assessment.
- Where it is associated with a dialogue between learner and teacher, the initial reliability of the assessment need not be high; what matters is that a basis is laid for a conversation. However, if there is to be communication between assessor and assessed, there needs to be an understanding of the operational meaning of the relevant criteria.

Formative assessment might be most useful where it provides high-quality information on individuals' performances but it is difficult for professors to find more time for one-to-one assessment. How, then, might formative assessment be promoted? Three ways are suggested by Knight:

- Make more use of teaching and research assistants. However, they need training; they cost money, and they have pressing research and career commitments of their own.
- Make more use of computer-based learning and assessment packages. Learning and assessment systems based upon artificial intelligence principles have great promise in some areas, especially where performance and likely errors can be quite precisely defined. However, they demand a considerable initial investment which only repays itself over many years.
- Make more use of self-, peer- and co-assessment.

The option of self-, peer and co-assessment seems appealing. Not only does it ensure that learners have to understand the meaning of criteria and apply them seriously, it

also promotes important skills to do with judgement and self-appraisal (Boud, 1992). Knight adds that if learners do act to undermine the power of the formative assessment, they alone are the losers. Where learning processes are to be assessed, then self-, peer- and co-assessment may be the only valid ways of doing so and the value and importance of assessing learning processes were shown in the findings of research into the skills that industry values in new graduates (Harvey & Knight, 1996).

A problem of self-, peer- and co-assessment is that they may shift the time burden from the teacher to the student. Since students are not practised assessors, it takes them far longer than an experienced teacher to make a formative judgement. This implies that very good assessment frameworks or tools are needed to support students as assessors, or perhaps that less material will have to be covered in order to make the time for students to use these new assessment strategies. However, Knight (1996) argues that whilst some tutors accept the enhanced learning in depth which self-, peer- and co-assessment offers, others value breadth of learning and become concerned if content coverage is restricted. There seems to be a need for:

- Consistency between program goals, teaching, learning and assessment procedures;
- the use of a wide range of assessment modes and methods;
- provision of opportunities for the development of shared meanings about assessment criteria;
- plentiful formative assessment, both at the classroom level and at the individual level;
- use of self-, peer- and co-assessment;
- a concern for depth of understanding in preference to breadth of knowledge, and
- provision of opportunities for all faculty to learn about ways of teaching that are consistent with an emphasis on student learning and on the transformative potential of higher education.

Assessment and Student Behaviour

Research into learning and instruction shows that in learning "the tail wags the dog", meaning that what students do is very much guided by the ways in which they are assessed (Gibbs, 1992). Traditional exams and tests have tended to encourage memorization and rote learning even when lecturers intended something rather different. Alternative assessment methods such as varied writing tasks, science journals, Overall Tests, portfolios, group projects, oral presentations, use of case studies and simulations, offer opportunities to make assessment a valuable learning experience in addition to allowing grades to be allocated (Birenbaum & Dochy, 1996, Brown & Knight, 1994). McDowell (1995) suggests that students find newer forms of assessment intrinsically interesting and motivating. They remain aware of the need to achieve high grades, but tend not to focus exclusively upon achieving them. There is evidence that they learn and behave differently than they do in courses where there is a traditional test or exam. Thus, alternative assessment may contribute to changing the culture amongst students from a testing culture to an assessment culture (Birenbaum & Dochy, 1996). More use of formative assessment is one way to convince students that assessment has the purposes

of indicating their strengths, weaknesses and progress, and guiding them towards their learning goals in addition to verifying their levels of achievement.

Research shows here that alternative assessment methods are less threatening to most students than traditional testing. They are perceived as fairer tests (Sambell, McDowell & Brown, 1997, this issue). Moreover, students seem to find meaning in assignments such as projects, group exercises and portfolios perhaps because they are seen as more like real life activities, more appropriate to Powerful Learning Environments, than examinations. Such assessment enhances students' intrinsic motivation, although the extrinsic motivation of grades remains present. Changing assessment methods is an effective way of encouraging students to change their learning methods. When assessment stays the same, students often will not accept the need to change their approaches to learning; for example, students often prepare for exams by rote learning even if this is not appropriate. Assessment is one of the most effective tools for innovating in both instruction and learning.

Conclusion: Future Research and Developments

We believe that the assessment culture can be used to change instruction from a system that transfers knowledge into students' heads to one that tries to develop students who are capable of learning how to learn (Ridgway & Schoenfeld, 1994). The current societal and technological context requires education to make such a change. The explicit objective is to interweave assessment and instruction in order to improve education. For example, Clarke and Stephens (1996) used assessment as a means for systematic reform of mathematical education in Australia. Their study demonstrated the effectiveness of curricular change as a result of the implementation of the Victorian Certificate of Education, which is a mix of traditional assessment using multiple-choice tests and new assessment forms such as investigative projects, challenging problems and extended-answer analytical tests. Although enthusiastic and promising results have been reported, the first attempts to introduce new assessment procedures have not been entirely positive. Madaus and Kellaghan (1993) reported problems with the organization, time and costs of assessment programs.

A number of lessons can be learned from the early applications of alternative assessment programs. Firstly, one should not throw the baby out with the bath water. Objective tests are very useful for certain purposes, such as high-stake summative assessment of an individual's achievement, although they should not dominate an assessment program. Increasingly, measurement specialists recommend so-called balanced or pluralistic assessment programs, where multiple assessment formats are used (Birenbaum, 1996; Clarke & Stephens, 1996; Ridgway & Schoenfeld, 1994). There are several motives for these pluralistic assessment programs (Birenbaum, 1996; Messick, 1984): a single assessment format cannot serve several different purposes and decision-makers, and each assessment format has its own method variance, which interacts with persons.

There is a need to establish a system of assessing the quality of alternative assessment and implement quality control (Birenbaum, 1996). Various authors have recently proposed ways to extend the criteria, techniques and methods used in traditional psychometrics (Cronbach, 1988; Kane, 1992; Linn, Baker, & Dunbar, 1991). Others, like

Messick (1984; 1995), oppose the idea that there should be specific criteria, and claim that the concept of construct validity applies to all educational and psychological measurements, including performance assessment.

There is a need for well-designed and well-evaluated heuristics which would help teachers to implement and design high-quality alternative assessment procedures. Teachers need clear criteria for the development and use of assessment procedures, and easy-to-use techniques for quality improvement and quality control.

There is no one ideal assessment format for all cases. All assessment formats can have negative effects on teaching and learning. From this perspective, Frederiksen's (1984) criticism can be interpreted more broadly than simply as an exposure of the negative effects of multiple-choice items; it also shows that some aspects of any assessment may have negative as well as positive educational consequences. However, alternative assessment with its potential to develop and assess learning achievements which are congruent with contemporary views of the nature of knowledge and the needs of society for educated persons, is an avenue which we must continue to explore.

References

- Ager, T. (1993). Online placement testing in mathematics and chemistry. *Journal of Computer Based Instruction*, 20 (2), 52-57.
- Alexander, P.A., & Dochy, F.J. (1995). Conceptions of knowledge and beliefs: A comparison across varying cultural and educational communities. *American Educational Research Journal*, 32 (2), 413-442.
- Archbald, D.A., & Newmann, F.M. (1992). Approaches to assessing academic achievement. In H. Berlak, F.M. Newmann, E. Adams, D.A. Archbald, T. Burgess, J. Raven & T.A. Romberg (Eds.) *Toward a new science of educational testing and assessment* (pp. 139-180). Albany: State University of New York Press.
- Assiter, A., & Shaw, E. (Eds.) (1993). *Using records of achievement in higher education*. London: Kogan Page.
- Baker, E. (1994). Learning based assessments of history understanding. *Educational Psychologist*, 29 (2), 97-106.
- Baumgart, N. (1986). An outside observers' view of profiles in Britain. In P.M. Broadfoot (Ed.), *Profiles and records of achievement: A review of issues and practice* (pp. 42-61). London: Holt, Rinehart, & Winston.
- Berk, R.A. (1986). A consumer's guide to setting performance standards on criterion-referenced tests. *Review of Educational Research*, 56, 137-172.
- Biggs, J. (1996). Assessing learning quality: Reconciling institutional, staff and educational demands. *Assessment & Evaluation in Higher Education*, 21 (1), 5-15.

- Birenbaum, M. (1994). Toward adaptive assessment - The student's angle. *Studies in Educational Evaluation*, 20, 239-255.
- Birenbaum, M. (1996). Assessment 2000: Towards a pluralistic approach to assessment. In M. Birenbaum & F.J.R.C. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes and prior knowledge* (pp. 3-30). Boston: Kluwer.
- Birenbaum, M., & Dochy, F., (Eds.) (1996). *Alternatives in assessment of achievement, learning processes and prior knowledge*. Boston: Kluwer.
- Birenbaum, M., & Dochy, F.J.R.C. (1996). Introduction. In M. Birenbaum & F.J.R.C. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes and prior knowledge* (pp. xiii-xv). Boston: Kluwer.
- Boud, D. (1990). Assessment and the promotion of academic values. *Studies in Higher Education*, 15 (1), 101-111.
- Boud, D. (1992). The use of self-assessment schedules in negotiated learning. *Studies in Higher Education*, 18 (5), 529-549.
- Broadfoot, P.M. (1984). From public examinations to profile assessment: The French experience. In P.M. Broadfoot (Ed.), *Selection, certification and control*. Sussex: Falmer.
- Brown, S., & Knight, P. (1994). *Assessing learners in higher education*. London: Kogan Page.
- Campione, J.C., & Brown, A.L. (1990). Guided learning and transfer: Implications for approaches to assessment. In N. Frederiksen, R. Glaser, A. Lesgold, & M.G. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 141-172). Hillsdale, NJ: Erlbaum.
- Clarke, D., & Stephens, M. (1996). The ripple effect: The instructional impact of the systematic introduction of performance assessments in mathematics. In M. Birenbaum & F.J.R.C. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes and prior knowledge* (pp. 63-93). Boston: Kluwer.
- Clarke, R.M. (1986) Professional education for capability and change. *Higher Education Review*, 18 (3), 27-35
- Cronbach, L.J. (1988). Five perspectives on validity argument. In H. Wainer & H.I. Braun (Eds.), *Test validity* (pp. 3-17). Hillsdale, NJ: Erlbaum.
- Darling-Hammond, L. (1994). Performanced-based assessment and educational equity. *Harvard Educational Review*, 64, 5-29.
- De Corte, E. (1990). Toward powerful learning environments for the acquisition of problem solving skills. *European Journal of Psychology of Education*, 5, 519-541.
- De Corte, E. (1991). Recent onderzoek over leren en instructie: Een selectief overzicht [Recent research on learning and instruction: A selection]. *Tijdschrift voor Hoger Onderwijs*, 9 (1), 2-17.

- De Rijke, R., & Dochy, F. (1995). *Assessment centres in onderwijs, opleiding en HRM* [Assessment centres in education, training and Human Resources Management]. Utrecht: Lemma.
- Dochy, F. (1992). *Assessment of prior knowledge as a determinant for future learning: The use of prior knowledge state tests and knowledge profiles*. Utrecht/London: Lemma/Jessica Kingsley.
- Dochy, F. (1994). Prior knowledge and learning. In T. Husen & T.N. Postlethwaite (Eds.), *International encyclopedia of education*, second edition (pp. 4698-4702). Oxford/New York: Pergamon.
- Dochy, F.J.R.C. (1996). Assessment of domain-specific and domain-transcending prior knowledge: Entry assessment and the use of profile analysis. In M. Birenbaum & F.J.R.C. Dochy (Eds.), *Alternatives of assessment in achievements, learning process and prior learning* (pp. 227-264). Boston: Kluwer.
- Dochy, F., & Moerkerke, G. (in press). Assessment as a major influence on the present and future of learning and instruction. *International Journal of Educational Research*.
- Dochy, F., Moerkerke, G. & Martens, R. (1996). Integrating assessment, learning and instruction: assessment of domain-specific and domain-transcending prior knowledge and progress. *Studies in Educational Evaluation*, 22 (4), 309-339.
- Floden, R.E. (1994). Reshaping assessment concepts. *Educational Researcher*, 23, 4.
- Frederiksen, N. (1984). The real test bias, influences of testing on teaching and learning. *American Psychologist*, 39 (3), 193-202.
- Gibbs, G. (1992). Improving the quality of student learning through course design. In R. Barnett (Ed.), *Learning to effect*. Milton Keynes, UK: SRHE & Open University Press.
- Glaser, R. (1981). The future of testing. A research agenda for cognitive psychology and psychometrics. *American Psychologist*, 36 (9), 923-936
- Glaser, R. (1990). Toward new models for assessment. *International Journal of Educational Research*, 14, 375-483.
- Green, D., Brannigan, C., Mazelan, P., & Giles, L. (1994). Measuring student satisfaction: A method for improving the quality of the student experience. In S. Haselgrove (Ed.), *The student experience* (pp. 100-107). Buckingham, UK: SRHE & Open University Press
- Hall, K. (1995). *Co-assessment: Participation of students with staff in the assessment process*. Invited address at the 2nd EECAE Conference (European Electronic Conference on Assessment & Evaluation), EARLI-AE list, March 10-14 (listserv: listserv@nic.surfnet.nl).
- Hambleton, R., & Sireci, S. (in press). Future directions for norm-referenced and criterion-referenced achievement testing. *International Journal of Educational Research*.
- Harvey, L., & Green, D. (1994). Quality in Higher Education Project: Employer Satisfaction Summary Report. Birmingham, UK: University of Central England.

- Harvey, L., & Knight, P. (1996, forthcoming). *Transforming higher education*. Buckingham: Open University Press.
- Hill, P.W. (1993). *Profiles and the VCE: Authentic assessment in a high stakes environment*. Paper presented to the VCTA Comview Conference, Melbourne, 1 December.
- Jessup, G. (1991). *Outcomes, NVQs and the emerging model of education and training*. London: Falmer.
- Kane, M.T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112 (3), 527-535.
- Keeves, J.P. (1994). Methods of assessment in schools. In T. Husen & T.N. Postlethwaite (Eds.), *International encyclopedia of education* (pp. 362-370). Oxford/New York: Pergamon.
- Keeves, J.P. (Ed.) (1988). *Educational research, methodology and measurement: An international handbook*. Oxford/New York: Pergamon.
- Kellaghan, T., Madaus, G.F., & Airasian, P.W. (1982) *The effects of standardized testing*. Boston: Kluwer-Nijhoff.
- Knight, P. (1996). *Quality in higher education and the assessment of student learning*. Invited paper at the Third European Electronic Conference on Assessment and Evaluation, March 4-8, EARLI-AE list European Academic & Research Network (EARN) (EARLI_AE on Listserv@nic.surfnet.nl).
- Lee, D.M.S., Pliskin, N., & Kahn, B. (1994). The relationship between performance in a computer literacy course and students' prior achievement and knowledge. *Journal for Educational Computing Research*, 10 (1), 63-77.
- Lesh, R., & Lamon, S.J. (Eds.) (1992). *Assessment of authentic performance in school mathematics*. Washington, DC: American Association for the Advancement of Science.
- Linn, R.L. (1981) Measuring pretest-posttest performance changes. In R. Berk (Ed.), *Educational evaluation methodology: The state of the art* (pp 84-109). Baltimore, MD: Johns Hopkins University Press.
- Linn, R.L., Baker, E.L., & Dunbar, S.B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 20 (8), 15-21.
- Madaus, G.F., & Kellaghan, K. (1993). British experience with "authentic" testing. *Phi Delta Kappan*, 74, 458-469.
- Masters, G., & Mislevy, R.J. (1993). New views of student learning: Implications for educational measurement. In N. Frederiksen, R.J. Mislevy & I. Bejar (Eds.), *Test theory for a new generation of tests* (pp.219-242). Hillsdale, NJ : Erlbaum.
- McDowell, L. (1995). The impact of innovative assessment on student learning. *Innovations in Education and Training International*, 32 (4), 302-313.

- Messick, S. (1984). The psychology of educational measurement, *Journal of Educational Measurement*, 21, 215-238.
- Messick, S. (1995). Validity of psychological assessment. *American Psychologist*, 50 (9), 741-749.
- Moerkerke, G. (1996). *Assessment for flexible learning*. Utrecht: Lemma.
- Moerkerke, G., Dochy, F., & Martens, R. (1995). *Effects of self-assessment by means entry tests and progress tests on study results*. Paper presented at the annual EARLI conference, Nijmegen, August 26-31.
- Newmann, F.M., & Archbald, D.A. (1990). Organizational performance of schools. In P. Reyes (Ed.), *Teachers and their workplace: Commitment, performance and productivity*. Newbury Park, CA: Sage.
- Nisbet, J. (1993). Introduction. In *OECD-curriculum reform: Assessment in question* (pp. 25-38). Paris: Organisation for Economic Cooperation and Development.
- Nitko, A. (1995). Curriculum-based continuous assessment: A framework for concepts, procedures and policy. *Assessment in Education*, 2, 321-337.
- O'Connor, M.C. (1992). Rethinking aptitude, achievement and instruction: cognitive science research and the framing of assessment policy. In B.R. Gifford & M.C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement and instruction* (pp. 9-35). Boston, Kluwer.
- Otter, S. (1995). Assessing competence: The experience of the enterprise in higher education initiative. In A. Edwards & P. Knight (Eds.), *Assessing competence in higher education* (pp. 43-64). London: Kogan Page.
- Pellegrino, J.W., & Glaser, R. (1979). Cognitive correlates and components in the analysis of individual differences. *Intelligence*, 3, 187-214.
- Resnick, L.B., & Resnick, D.P. (1992). Assessing the thinking curriculum. In B.R. Gifford & M.C. O'Connor (Eds.), *Changing assessment: Alternative views of aptitude, achievement and instruction* (pp. 37-75). Boston: Kluwer.
- Ridgway, J., & Schoenfeld, A. (1994). *Balanced assessment: Designing assessment schemes to promote desirable change in mathematics education*. Invited paper at the First European Electronic Conference on Assessment and Evaluation, February 21-23, EARLI-AE list European Academic & Research Network (EARN) (EARLI_AE on Listserv@nic.surfnet.nl).
- Rogers, P. (1995). *Validity of assessments*. Contribution to the 2nd EECAE conference (European Electronic Conference on Assessment & Evaluation), EARLI-AE list, March 10-14 (listserv: listserv@nic.surfnet.nl).
- Rowe, K.J., & Hill, P.W. (1996). Assessing, recording and reporting students' educational progress: the case for 'subject profiles'. *Assessment in Education*, 3 (3), 309-352.

Sambell, K., McDowell, L., & Brown, S. (1997). "But is it fair?": An exploratory study of student perceptions of the consequential validity of assessment. *Studies in Educational Evaluation*, this issue.

Segers, M.S.R., & Dochy, F.J.R.C. (1996). The use of performance indicators for quality assurance in higher education. *Studies in Educational Evaluation*, 22 (2), 115-139.

Shavelson, R.J., Xiaohong, G., & Baxter, G. (1996). On the content validity of performance assessments: Centrality of domain-specifications. In M. Birenbaum, & F. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes and prior knowledge* (pp. 131-142). Boston: Kluwer.

Stephenson, J. (1994). Capability and competence: Are they the same and does it matter? *Capability*, 1 (1), 3-4.

Tan, C.M. (1992). An evaluation of the use of continuous assessment in the teaching of physiology. *Higher Education*, 23, 255-272.

Wigdor, A.K., & Garner, W.R. (Eds.) (1982). *Ability testing: Uses, consequences and controversies*. Washington, DC: Academy Press.

Wijnen, W.H.W.W. (1990). Waarden en normeren van beoordelingen [Valuing and setting standards for assessment]. In H.J.M. van Berkel & A.E. Bax (Eds.), *Beoordelen in het onderwijs* [Assessing in education] (pp.94-99). Maastricht: Versluys.

Wolf, A. (1995). *Competence-based assessment*. Buckingham: Open University Press.

Wolf, D., Bixby, J., Glenn, J., & Gardner, H. (1991). To use their minds well: Investigating new forms of student assessment. *Review of Research in Education*, 17, 31-74.

The Authors

FILIP DOCHY is Professor of Instructional Science and Technology in Teacher Training at the Centre of Teacher Training & Research, University of Leuven, Belgium and Research Manager of the Centre for Educational Technology and Expertise, Division of Applied Assessment Sciences, at the Open University of Heerlen, The Netherlands.

LIZ MC DOWELL is Senior Lecturer in Educational Development in the Centre for Advances in Higher Education, University of Northumbria at Newcastle, UK. She specialises in assessment and evaluation within Higher Education, both as a researcher and as a consultant to a range of academic departments and universities.