

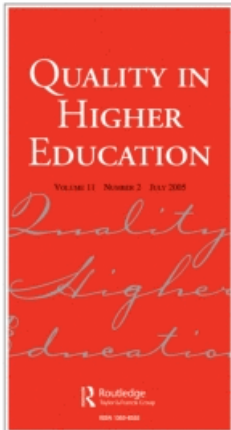
This article was downloaded by:

On: 21 April 2011

Access details: *Access Details: Free Access*

Publisher *Routledge*

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Quality in Higher Education

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713443244>

A Transatlantic View of Assessment and Quality in Higher Education

T. Dary Erwin^a; Peter Knight^b

^a James Madison University, Virginia, USA ^b Department of Educational Research, Lancaster University, Lancaster LA1 4YL, UK

To cite this Article Erwin, T. Dary and Knight, Peter(1995) 'A Transatlantic View of Assessment and Quality in Higher Education', *Quality in Higher Education*, 1: 2, 179 – 188

To link to this Article: DOI: 10.1080/1353832950010208

URL: <http://dx.doi.org/10.1080/1353832950010208>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

A Transatlantic View of Assessment and Quality in Higher Education

T. DARY ERWIN

James Madison University, Virginia, USA

PETER KNIGHT

Department of Educational Research, Lancaster University, Lancaster LA1 4YL, UK

ABSTRACT *Drawing on practice in the US and in the UK, the authors argue that the ways in which student learning is assessed constitute a sensitive set of indicators of the quality of the undergraduate experience. Indifferent assessment arrangements can poison otherwise well-conceived curricula. Distinguishing between the formative and the summative functions of assessment, they identify characteristics of assessment arrangements that support good quality learning, while observing that common practice in the US and in the UK falls short of these ideals. However, socio-political developments in western democracies make it risky for universities to continue to give tacit support for this state of affairs. A set of characteristics of assessment systems that support quality in higher education is proposed.*

Assessment—Poison or Power Source?

'Assessment' means different things to different groups. In the US, the word may refer to the assessment of student learning or it may refer to the evaluation of programmes. It will often be understood as testing, that is as a reliable procedure for collecting summative data, but it can also refer to the making of inferences based upon student performance on 'authentic' learning activities, whether the inferences are for summative or formative purposes. In this article, we argue that continuous quality improvement in higher education depends upon well-conceived approaches to assessment that have both summative and formative functions. Not only does the absence of such approaches betoken deficiencies in a university's quality improvement procedures, it also invites governments to impose their own views of quality upon them.

If we understand curriculum better, then we can better understand the formative and summative purposes of assessment and the significance of the arrangements for the assessment of student learning for continuous quality improvement in higher education. Instead of talking about curriculum in higher education, it is useful to distinguish between the planned, the 'delivered' and the 'received' curriculum, as well as to think about the used curriculum. The scare quotes indicate that both delivered and received are misleading metaphors. When we have planned a curriculum, we do not deliver it to learners like the postal service delivering a parcel: it is widely recognised that implementing plans involves changing them in many ways. It would be better to speak of the *created* rather than of the

delivered curriculum. That should not disguise the main point, that if we are interested in quality, then we need to look not just at paperwork (a tempting option to some quality appraisers on both sides of the Atlantic) but also at the ways in which curriculum is created through the actions of the tutors (and quick classroom observation does not do justice to that). Likewise, what is created by the tutor is not what the learner 'receives', for it is well known that human cognition is an active process whereby new information is processed, stored and understood in relation to learners' own mental structures. Learners transform what it is they are expected to learn, often in some rather unexpected ways (Biggs, 1993). So, rather than the received curriculum, we might speak of the *understood* curriculum. Those interested in the quality of learning are naturally interested in the quality of this understood curriculum. The implication is that where the quality of learning is being judged, then evidence needs to be provided showing what has been learned (the understood curriculum) and only then evidence of what was created by the tutor, and what had been planned.

The fourth sense of curriculum concerns the 'used' curriculum, that is the extent to which the understood curriculum contributes to life beyond the undergraduate experience. These contributions may be in the workplace or as citizens, parents, care-takers, or neighbours. Do graduates make use of their education? For example, is education in tune with societal needs, broadly understood?

Frequently, in the appraisal of quality a paper audit of planning takes precedence over enquiry into the understood curriculum. In other cases, notably in some North American colleges, off-the-peg tests of student learning are administered as if they could offer reliable and valid data about the understood curriculum and hence of student learning and the quality of education. The blind adoption of a colleague's examination, or in the US, buying a test from a testing organisation without determining whether this measure fits their intended curriculum are often quick fixes to provide numbers without anticipating the value of the information. In such cases, the assessment method is unlikely to reflect the comprehensiveness of the programme and is likely to produce surprising or unsatisfactory results. A consequence can be that a 'teaching to the test' phenomenon follows as teaching staff, consciously or unconsciously, shift their delivered curriculum to match the test content. Where this is done unconsciously, then the teaching staff have unwittingly allowed the test to dictate the curriculum, sacrificing their own goals and academic freedom in the process. In the US, governmental officials implemented statewide tests when they were dissatisfied with higher education's response to calls for better measures of quality, suggesting that where academic staff do not have in place assessments of student learning that are fit for the purposes of sustaining learning and of supporting public accountability, then governments will intervene and impose their own agenda. National assessment requirements do not presently exist within the UK but lack of attention to public accountability needs could cause this to happen.

A second problem, which is especially acute where courses are taught by more than one tutor, is lack of reliability about the measures of quality. If teaching staff differentially rate an essay, which rating is correct? If no consistency or inter-rater reliability about student performance is reached, then any conclusions about the delivered or received curriculum are suspect. When wide disagreement occurs in marking students' performance, product, or process, then a lack of collaboration on the planned curriculum is evident. In addition, low inter-rater reliability may also be due to a poor rating scale. For example, one rating scale asked teaching staff to evaluate the ability of theatre majors to critique plays 'in a critical, specific, and well-ordered fashion'. Low inter-rater reliability of +0.3 was found among the four teaching staff rating the students' critique, suggesting low agreement. Why

might this have happened? The rating scale asked the tutors, or raters, to respond at least to three criteria simultaneously, namely 'is the student's paper: (a) critical, (b) specific, and (c) well-ordered?'. This approach seems to have caused two problems. One was that raters were unsure about the weighting to be given to each of the three criteria, some preferring one criterion, others another. In addition, there was evidence of confusion about how to interpret and apply vague terms such as 'critical' and 'well-ordered'. Needless to say, quite apart from the problems this causes in rating a performance, any confusion on the tutor's part leads to confusing feedback to the students, thereby jeopardising the quality of learning.

The information collected is, of course, limited by the worth or validity of the assessment method. Some general education measures in the US measure general verbal ability rather than knowledge in the arts or natural science, as the tests claim. Reliance on information from a faulty test leads to faulty conclusions about the *created* and *understood* curriculum. In another example, a variety of critical thinking tests are available on the market. Although critical thinking is a concept widely used and valued by academic staff, universal consensus about the meaning of critical thinking is largely lacking. Some tests of critical thinking define critical thinking as logic; others, such as Perry's scheme (Erwin, 1986; Perry, 1970), as intellectual development; others as inductive reasoning; others about esthetic judgment or about ability to integrate from several sources; and so on. It is important that the assessment method match the planned and delivered curriculum.

The implications of ill-conceived assessment methods for the quality of learning may be further explored. Suppose that we had planned a course that would teach certain material and that would encourage learners to adopt deep approaches to that material, approaches where the emphasis was upon creating an understanding of the material, rather than striving to be able to reproduce it. Suppose too that this course was to be assessed on the basis of a conventional examination and conventional coursework, for example laboratory reports or essays. Given student expectations, and given that there is a body of students who believe that essays, for example, should be collections of information, rather than arguments (Hounsell, 1984), then it can be said that some, perhaps most students, are likely to interpret those assessment arrangements as prompts to adopt surface learning approaches.

Learning takes place in a context and within a system (Biggs, 1993). Phenomenographic (or phenomenological, or ethnographic) research into learning has strongly suggested that students' learning approaches are affected by their intellectual environment (Trigwell & Prosser, 1993), of which assessment is a part. Furthermore, this environment might be seen as a system. The point is that if one part of a system (the goal of deep learning) is out of kilter with the other parts (student conceptions of the academic activity, teaching styles, learning tasks and assessment processes), then it will be lacking in impact.

Consider this point a little further. Assessment, it is suggested, shapes the understood curriculum. Much of students' work comprises doing assessment tasks. The way they approach learning is often shaped by the assessment tasks, and the way they feel about their learning and themselves as learners is also shaped by the assessment tasks. So, we suggest that assessment has considerable power to shape student learning. If all other elements of the course point in one direction and the assessment arrangements in another, then the assessment arrangements are likely to have the greatest influence on the understood curriculum. Good curriculum cannot contain bad assessment but bad assessment can poison good, planned and created curricula. It follows that attempts to examine the quality of higher education need to be sensitive to the quality of the assessment arrangements within the curriculum. In England, for one, that was not the case in the early years of the

system of academic audit, which did not recognise that the quality of student assessment was one of the most sensitive indicators of the quality of learning provision as a whole. A recent report (HEQC, 1995) shows greater awareness of the significance of assessment arrangements for student learning. Unfortunately, in some areas of the US, statewide tests are mandated without consideration of teaching staff wishes or intent.

Purposes of Assessment

If it is accepted that assessment arrangements can have such power in the curriculum (and hence that they are sensitive indicators of the quality of the curriculum), it follows that unless the purposes of assessment are clear and appropriate, quality is compromised.

Formative Assessment Purposes

It is widely accepted that assessment can have a formative function, which is to say that it can help faculty (academic staff) to improve their teaching and learners to improve their learning. Valid and reliable assessment data tell tutors what has been understood (illuminating the size of the gaps between the planned, the created and the understood curricula), and in this way can make for better learning in the future by helping faculty to see where the course needs to be strengthened.

Formative assessment may be described as assessment for learning. It is diagnostic identifying what learners do not know, as well as that which they do well enough. Feedback will inform students how well they are progressing. This feedback serves as a reality test and as a motivator for further learning. On this view, if properly structured, such assessment information can and perhaps should also be part of the learning process itself, as it is in some colleges, Alverno College in Milwaukee being a well-known example. If this approach to assessment is to work, it is important that learners get good quality feedback about their performance, which implies that they are familiar with the criteria that are applied to the assessment of their work. This, it should be stressed, means understanding what the criteria mean in the context of the course or programme, not simply recognising the names of the criteria. An implication is that learners will spend some time becoming familiar with those criteria and, ideally, applying them, perhaps by marking each other's work, or by marking anonymised work done in previous years (see, for example, Stephani, 1994, on peer assessment). In some cases, these criteria will be a mix of faculty-selected and student-suggested criteria (see Knight, 1995a, for examples).

While clarity about criteria is necessary for effective feedback (and supports a deep approach to the task in the first place), time also needs to be found for feedback and it is assumed that faculty have the skill to engage in the feedback process. However, English evidence is that feedback is often cursory and unhelpful (Brown & Knight, 1994, ch. 3 and 8). This is worrying, since there is evidence that learning depends upon good feedback: Laurillard (1993) has the interplay between learner and tutor at the heart of her model of learning, and while we recognise that interaction *between* learners can have a valuable function and that learners may engage in a critical dialogue with themselves, we see no substitute for interaction between learner and tutor. This, as Laurillard makes clear, may take place in a variety of settings using a variety of media (and computer conferencing has a strong appeal). Unfortunately, promising developments in such academic guidance, such as Records of Achievement and portfolio assessment, make considerable demands on faculty time and demand considerable skill and preparation if they are to deliver the proclaimed benefits (Trowler & Hinnett, 1994; Knight, 1995b). Yet, good learning is

conceptually related to the provision of feedback from tutor to learner (and from learner to tutor too) and the evidence is that in many assessment systems it is not provided. This absence of a prerequisite for learning might be considered to be an ethical failing in higher education.

Unfortunately, even if the time were found for better feedback, it is not clear that academic staff have, in general, the skill to give feedback that helps. In part this is because it is quite clear that different learners on different tasks at different times want different sorts of feedback: sometimes detailed, sometimes pointing to a few general points for development. Yet evidence from English schools (Bennett *et al.*, 1984; Desforges & Cockburn, 1987; Bennett & Kell, 1989) has repeatedly shown that schoolteachers find it hard to diagnose learning difficulties and to help learners with difficulties to surmount the block to understanding. A cautious extrapolation would suggest that university teachers may not often be well placed to identify, let alone to remedy learning difficulties: after all, what training have they had in so doing? Where work is marked holistically, that is one mark is awarded on the basis of a single judgement of the quality of the whole piece of work, it is harder to give feedback that relates to key learning criteria: more analytical or structured feedback is desirable. This implies that analytical marking should take place, where a variety of criteria, relating to levels of performance, is used to identify the pattern of strengths and weaknesses in a piece. This allows academic staff to provide students with a profile of their performance, rather than with a general, overall comment. For example, evaluating writing ability may be accomplished through holistic or analytical type rating scales. A holistic rating scale evaluates *total* writing ability separating students into, say, five qualitative levels. A holistic rating scale certifies overall writing ability but offers little feedback to the student or programme director about what could be improved in the delivery of the curriculum. An analytical rating scale may have several diagnostic categories defining writing ability such as usage, sentence construction, organisation/content, mechanics, support, and focus. Ratings are given in each of the six areas (or whatever writing sub-areas are selected). An overall holistic rating would not describe which aspect of the writing curriculum is working well or should be improved. Conversely, analytical scales may, in some forms, tend to reductionism, giving learners information about sub-areas but failing to convey an appreciation of the piece of writing as a whole. This is not an inevitable outcome of analytical rating scales, but it is a danger of them.

Where such analytical criteria are in place, it is feasible to certify student competence on entry to a programme. In the past, the mode of instruction was fairly fixed and assessment methods variable. In the future, instructional strategies are likely to become more diverse, not least because of the increased application of technology to learning (Chodorow, 1995), and the set of assessment methods may become the constant. This diversity of ways of learning may be seen in a greater use of workplace-based learning, of independent projects, and of technology-based instruction. However, this will be diversity in pursuit of a set of learning goals, exemplified in assessment systems, that defines what learners have to demonstrate, irrespective of their mode of learning, in order to be certified as competent in a programme. Assessment will give students who have gained learning in non-traditional ways opportunities to demonstrate that learning has been accomplished. For example, the Internet offers wide possibilities for access to information. Whether the student has independently learned through accessing the Internet, this learning experience may be validated through an institution's assessment arrangements and accredited as prior learning. Examining assessment arrangements for evidence of quality in higher education entails looking to see that they support this formative function, and paying particular attention to clarity of criteria and to arrangements for feedback.

Summative Assessment Purposes

Assessment also has summative purposes, which is to say that it provides end-of-course information about learners' achievements. It not only certifies these achievements but it also provides performance indicators that have a myriad of policy and management uses.

First, we need to ask about the focus of summative assessment: are summative data really related to the quality of learning or are they (often rather inadequate) proxies for that quality? It is useful to distinguish between summative data relating to outputs and those relating to outcomes. In the US, and to some extent in the UK as well, output data have been collected and used in the judgement of quality. Such output data include retention rates, graduation rates and graduate employment records. While they are not without interest, they appeal to a 'bean counting' mentality and say little about the quality of learning. Retention rates, for example, may be strongly related to the characteristics of the student intake and to the socio-economic structure of the area in which the university is situated.

Our claim is that summative assessment needs to be based on evidence of *learning outcomes*, directly providing evidence about the university's core business, which is student learning. Summative assessment data will show quite clearly what those outcomes are and they can be subjected to critical appraisal as a part of the evaluation of the quality of learning. On both sides of the Atlantic it is likely that quality assessors would expect to see evidence not only of a view of the domain-specific learning that would be supported by a programme, but also of cross-curricular or generic learning that the programme promotes. Where inspection of summative data showed that a course team neither had a view of the desired disciplinary outcomes, nor of the generic outcomes, adverse judgements about the quality of learning might follow. Where neither a disciplinary, nor a generic account of learning outcomes existed, the conclusion could be that faculty had not thought clearly about teaching and learning; that there would therefore be a lack of consistency and progression in the programme; and that learners would have to re-tune their expectations each time they began a class with a new tutor. Where summative data related only to disciplinary goals, it might be concluded that, desirable though this is, the programme's goals were too narrow in the present socio-political climate on both sides of the Atlantic (and of the Pacific too). In developed countries, governments increasingly expect to hear that expensive undergraduate programmes develop generic competences, such as communication, group working, self-management skills, and the application of information technology. Where summative assessment data do not provide evidence of achievement in these areas, the conclusion could be that a programme's quality was deficient either because these competences are not fostered, or because they are not valued enough to be assessed or to the educators. Following the model set out at the beginning of this paper, if some things are not valued enough to be assessed, the understood curriculum is less likely to embrace them than if they were assessed: the lack of assessment signals that these competences are not really important.

Some teaching staff will not be easy with the somewhat vocational-looking set of competences that we have just mentioned. There are alternative accounts of the general learning that can come from higher education, such as that offered by Barnett (1994), who argues that the undergraduate experience should encourage wisdom and the ability to step outside disciplinary and occupational frameworks, making meaning for oneself as a part of a process of life-long learning or self actualisation. We too have already alluded to the value orientations that might be encountered within the undergraduate experience, not in the spirit of arguing that higher education should be indoctrinating (which would be an

oxymoron) but in the spirit that higher education cannot proceed as though it were a value-neutral enterprise. Generic skills do not have to be vocational and determinist. Howsoever they are conceived, summative data should give evidence of critical engagement with questions about the nature of generic learning that might take place within the university. If they do not, questions might be asked about the quality of the programme.

An implication of programmes that claim to promote a range of domain-specific and generic outcomes is that multi-strategy assessment is necessary: learning will be assessed through a variety of methods (IT-based, performances, presentations, memoranda) and media and for a variety of purposes (formative or summative) and by a variety of people (self-, peer- and tutor-assessment, assessment by business people and by others who are not education professionals). Heywood (1994) has provided a full case for this multi-strategy approach. We suggest that where there is not evidence of a *planned*, multi-strategy approach for a programme, questions ought to be raised about the quality of learning outcomes. How can a range of outcomes be validly documented, let alone supported, without a range of assessment strategies Erwin (1995)?

The final point about summative assessment data relates to the issue of interpretation. If graduates show mastery of a number of learning outcomes that is something to be welcomed but it might not be a sign that the undergraduate experience has been successful. Some of the generic competences may well have been substantially mastered on entry to the programme, with the result that the main impact of the undergraduate experience has been in students' major subject area. In other cases, the achievement of mastery represents considerable learning at university and a noteworthy achievement. Borrowing from research into effective schools, we suggest that summative assessment data should not just be considered at face value but that a before-and-after analysis be conducted to see what has changed during the undergraduate years. This implies that students might be assessed on entry to provide a point of comparison with their graduating achievements, hence provide an estimate of the impact of the undergraduate experience on their learning. There are considerable technical problems with this 'added value' treatment of summative data (PCFC/CNAA, 1990) but there is, nevertheless, enormous value in asking what students appear to have learned since entering the university (Astin, 1991), as innovators at James Madison University and at Alverno College have found (Alverno College Faculty, 1994; Erwin, 1991, 1995).

We have said that summative data should relate to learning outcomes. In part, this is an ethical requirement, in that we have a professional responsibility to do what works best. But it is also socio-political, based on the belief that macro-changes will rob universities of their freedom unless they anticipate those changes and act in ways that forestall government action. This point is further developed in the next section.

Assessment and Higher Education Policy

In discussion of the goals and achievements of higher education, on both sides of the Atlantic, there has been a tendency to give increasing weight to the views of employers and, through the agency of governments, those of the tax-payers. Continuing economic problems and the rise, internationally, of wariness of state welfare provision mean that we do not expect this tendency to be attenuated. From a macro perspective, Krueger (1992) has claimed that what is valued in higher education is funded, and what is funded is assessed. Academic staff generally believe that quality should be rewarded, and future public policy may shift in this direction as governments find the pressures of financing higher education on the basis of student numbers less attractive than the idea of financing them on the basis

of learning quality. In that case, the credibility and worth of our assessment methods will become paramount. If academic staff wish to see judgements about quality based upon the appraisal of learning outcomes, it is necessary for them to collect evidence about those learning outcomes as has been suggested. Where that does not happen, states that are trying to balance the budget (as with the US Federal government) or to restrain burgeoning expenditure (as in the UK and Australia) will have little choice but to make quality judgements on the inappropriate basis of learning outputs, including student numbers. Universities that wish to enter the 21st century on more or less their own terms have, it has been argued, little choice but to develop forms of accountability that show the value of their work in easily-understood terms. The experience of schools in England might be a warning of an alternative scenario.

For some 20 years government has been concerned about the quality of schooling in England. It saw reform of the curriculum as a powerful way of raising standards in education. In the 1980s it supported school-based curriculum development and offered schools a series of guidelines to help them to reform their own curricula. By the mid-1980s it had become clear that reform was patchy, both in terms of uptake and of quality. Fundamental questions were not being addressed and structural weaknesses persisted. Government response was to impose a complex national curriculum on schools in the face of considerable teacher hostility. Some seven years on, this curriculum has been made more teacher-friendly and more manageable but there is no doubt that schools' failure to reform themselves led to government action that took the initiative away from them and curbed teacher autonomy. It is not only the British government that has shown a commitment to consumer power and to the regulation of state provision and it could be unwise to assume that higher education, both in the UK and elsewhere, will escape such regulatory attention if it fails to provide readily-understandable evidence of what it provides for the taxpayers' money.

In the US, a similar pattern has followed. In the minds of many US government officials, improvements in the educational system also lead to improvements in the economy. This belief gives rise to the view that there should be greater monitoring of educational systems, including the performance of institutions of higher education. Accrediting agencies, which represent a peer-review process within US higher education, have been attacked since the 1980s for focusing on resources and institutional reputation and not student learning (Astin, 1991). National Education Goals of critical thinking, problem solving, and communication skills, have been federally formulated for all college students (Corallo, 1991).

The line that has been taken in this article implies a substantial change in culture and practices in many universities. This takes time—it could be sensible to think in terms of a 10-year plan to improve learning quality through assessment reform and through other innovations in learning (perhaps through a World Wide Web-based 'virtual university') and in teaching (perhaps through re-conceptualising teaching to shift emphasis away from the quality of face-to-face contact and to the quality of course design). We are also attracted by the idea of incremental change within such a strategic vision. Not only has some work on innovation in school teaching shown that bottom-up 'tinkering' can be highly effective as an approach to curriculum renewal (Huberman, 1992) but we are also mindful of a study of the impact of 'total quality' thinking on US firms. This survey of 584 companies in three nations concluded that quality improvement needed to take different forms according to the current performance of a company: low performers need to start from a different point to high performers. Moreover, 'the most popularly used practices are least effective, and the most effective practices are least used' (Fuchsberg, 1992, p. B2). The suggestion is that assessment reform, as a part of continuous quality improvement, should

take the form of an incremental series of local initiatives within a medium- to long-term institutional strategy. Moreover, there are substantial funding implications and many universities will need to consider whether they invest sufficient monies in their equivalent to commercial research and development: where a business might commit between 4% and 10% of income to research and development, it appears that many universities assume that much smaller amounts are appropriate.

A two-fold argument has been outlined. In the first place, it has been contended that the quality of learning and the assessment systems used in higher education are conceptually related. Those interested in the quality of learning and in the quality of teaching need to pay close attention to the ways in which student learning is assessed, for this provides powerful evidence about teaching and learning quality. The second line of argument has been that where higher education institutions do not establish sound assessment systems, then they are vulnerable to governmental regulation, which may not be sympathetic to academic goals and which may jeopardise academic freedom. This leads to a suggested list of the characteristics of assessment systems that are consonant with quality in higher education.

Assessment of Learning and Quality in Higher Education Summary

A quality programme in higher education should be characterised by an assessment system that has the following features:

- the intended curriculum aims should be clearly defined;
- the expectations attached to each learning aim need to be clearly expressed, which is to say that both staff and students need to understand the assessment criteria that will be applied;
- a range of learning outcomes, both subject-specific and generic, should be assessed;
- assessment methods should be fit for the purpose, that is they should be valid measures of the intended learning outcomes;
- multiple programme aims demand multiple assessment methods;
- there should be evidence that students get useful feedback on their work through interaction with teaching staff and, perhaps, their peers;
- consideration should be given to making learning programmes, aims, criteria and outcomes public through the World Wide Web and through documentation made available to the public;
- assessment data should feature in university decision-making processes, reflecting academic staff's professional and ethical responsibility to identify what works. Likewise, assessment data should be used in programme review and resourcing decisions;
- it follows from the foregoing points that the collection and use issues require that summative assessment data be centrally stored in a form that is readily accessible to authorised staff and which can be readily analysed using standard statistical packages;
- similarly, universities should consider establishing Offices for Student Assessment, whose officers should be professionals trained in psychometrics, in educational enquiry, and in interpersonal and management skills;

- if government intervention is to be forestalled, universities need to develop a central steer within which diversity and academic freedom are possible.

There can be no illusions that it is easy to change traditional patterns of assessment and learning in higher education on the lines suggested. Yet, whether because the aim is better quality education, or whether the aim is to avoid state regulation of higher education, action is necessary. Of course, there is a do-nothing alternative. Those choosing it ought carefully to consider the consequences at a time when continuous quality improvement is seen as a necessary feature of effective organisations and when governments are interested in diverting funds from the ineffective to the effective.

References

- ALVERNO COLLEGE FACULTY, 1994, *Student Assessment-as-Learning at Alverno College* (Milwaukee, Alverno College Institute).
- ASTIN, A., 1991, *Assessment for Excellence: the philosophy and practice of assessment and evaluation in higher education* (New York, Macmillan).
- BARNETT, R., 1994, *The Limits of Competence* (Buckingham, Open University Press/Society for Research into Higher Education).
- BENNETT, N. *et al.*, 1984, *The Quality of Pupil Learning Experiences* (London, Erlbaum).
- BROWN, S. & KNIGHT, P.T., 1994, *Assessing Learners in Higher Education* (London, Kogan Page).
- BENNETT, N. & KELL, J., 1989, *A Good Start?* (Oxford, Blackwell).
- CHODOROW, S., 1995, 'The virtual university', *Penn Printout*, 11(6), pp. 1, 21–3.
- CORRALLO, S., 1991, 'Critical concerns in assessing selected higher-order thinking and communication skills of college graduates', *Assessment Update*, 3(6), pp. 5–6.
- DESFORGES, C. & COCKBURN, A., 1987, *Understanding the Mathematics Teacher* (Lewes, Falmer Press).
- ERWIN, T.D., 1983, 'The scale of intellectual development: measuring Perry's scheme', *Journal of College Student Personnel*, 24, pp. 6–12.
- ERWIN, T.D., 1986, 'Students' contribution to their college costs and intellectual development', *Research in Higher Education*, 25, pp. 194–203.
- ERWIN, T.D., 1991, *Assessing Student Learning and Development: a guide to the principles, goals, and methods of determining college outcomes* (San Francisco, Jossey-Bass).
- ERWIN, T.D., 1995, 'Attending to assessment: a process for faculty, in: P. KNIGHT (Ed.) *Assessment for Learning in Higher Education* (London, Kogan Page).
- FUCHSBERG, G., 1992, 'Total quality is termed only partial success', *Wall Street Journal*, 10 January, pp. B1–B2.
- HEYWOOD, J., 1994, *Enterprise Learning and its Assessment in Higher Education* (Sheffield, Employment Department).
- HUBERMAN, M., 1992, 'Teacher development and instructional mastery', in: A. HARGREAVES & M. FULLAN (Eds) *Understanding Teacher Development* (London, Cassell).
- KNIGHT, P.T. (Ed.), 1995a, *Assessment for Learning in Higher Education* (London, Kogan Page).
- KNIGHT, P., 1995b, *Records of Achievement in Higher and Further Education* (Lancaster, Framework Press).
- KRUEGER, D.W., 1992, 'A return to excellence', *Assessment Update*, 4, pp. 1–N4.
- LAURILLARD D., 1993, *Rethinking University Teaching*, (London, Routledge).
- PERRY, W., Jr., 1970, *Forms of Intellectual and Ethical Development in the College Years: a scheme* (New York, Holt, Rinehart & Winston).
- POLYTECHNICS & COLLEGES FUNDING COUNCIL & COUNCIL FOR NATIONAL ACADEMIC AWARDS (PCFC/CNAA), 1990, *The Measurement of Value Added in Higher Education* (London, CNAA).
- STEFANI, L.A.J., 1994, 'Peer, self and tutor assessment: relative reliabilities', *Studies in Higher Education*, 19, pp. 69–75.
- TROWLER, P. & HINNETT, K., 1994, 'Implementing the recording of achievement in higher education', *Capability*, 1(1).