

The defining qualities of good assessment



Michael F. Christie

Chalmers University of Technology, Göteborg, Sweden
michael.christie@pedu.chalmers.se

Tom Stehlik

University of South Australia, Adelaide, Australia
Thomas.Stehlik@unisa.edu.au>

***Abstract:** In this paper we challenge the way that quality is defined in higher education and ask what characterizes quality in the assessment of student learning. In order to do this we are obliged to untangle some of the aims of assessment within universities. There are a number of stakeholders involved with assessment of learning and we argue that they do not always agree on what constitutes good assessment practice. The word 'quality' is often used as a talisman to ward off competing views about assessment of learning. In this paper we stress the need to use the term in one of its original meanings – excellence. Although this definition is often criticised because of its apparent vagueness, we argue that by using it we are forced to seek consensus on what constitutes excellence. In the absence of consensus we are forced to rely on qualified judges. In the matter of assessment in higher education we have to ask ourselves 'who are the best judges of a student's learning'. We argue that teachers and students are the best judges of effective learning but that external quality audits can undermine their judgement. Sweden is moving closer to a British model of quality assurance and the culture of compliance that this engenders can impede the use of collaborative, continuous, formative and innovative ways of assessing learning.*

***Keywords:** learning, assessment, quality*

Introduction

In January 2001 the government agency that oversees universities in Sweden (Högskoleverket) introduced a new system of quality assessment. Whereas previously audits had been carried out at the institutional level the new system was aimed at assuring the quality of subjects and programmes. The change of focus was due, according to the agency, to increasing demands from society that universities deliver quality education to a more numerous and more diverse clientele (see Högskoleverkets rapportserie 2001:9). As Sweden's economy slowed down in the late nineteen nineties the government insisted that the universities give 'value for money'. At 1.7% of GNP the amount of money invested in higher education was by no means trivial. The hope was that 50% of all Swedes would receive a higher education before the age of 25 and that such an educated workforce would help maintain Sweden's good standard of living. To ensure this, higher education had to be more relevant and of a higher quality. Högskoleverket (hereafter HSV) was concerned by falling numbers of university applicants and by an apparent dissatisfaction among students with the quality of university teaching. A survey of students at a large technical university in Göteborg (Eklöf and Svensson, 2000)

showed that only 43% of the 4004 respondents thought that their teachers were good or very good. The issue of teacher competence was also raised in newspaper articles and letters to the editor.

In the past, quality control of teaching, learning and research had been left to the universities. HSV checked on the mechanisms universities had in place to assure quality and kept an eye on student progression rates, publication records and research grants within the different universities. By the turn of the century, however, a shrinking economy, growing problems in meeting social welfare payments and a crisis in the health system made the Swedish government particularly sensitive to complaints about the quality of higher education. A government bill (1999/2000:28) gave HSV a new role in quality assessment. HSV recruited staff to organise a regular cycle of course and program audits. Experts and students in a particular field were invited to form panels with the object of thoroughly reviewing each of the major university disciplines once every six years. To ensure such audits were taken seriously universities were warned that if a course or program had serious defects and those defects were not remedied within a year then the right to award a degree in those programmes could be revoked.

The new Swedish system was partly inspired by the British Quality Assurance Agency (hereafter QAA). There are similarities and differences between the two. The QAA system is now well established and exerts a great deal of influence on how universities deal with quality assurance. Many British universities endeavour to put in place an overall quality control system that is consistent across faculties and matches the QAA model. In Sweden the new audit system complements the existing self evaluation process that universities engage in but is also likely that the audits will have an impact on the way these processes evolve. The most common objection to the British system is that it promotes a 'culture of compliance' within the universities. Harvey and Knight (1996: 97) have pointed out that the demands on teaching staff to respond to external monitoring can adversely affect efforts to enhance the student learning experience. The danger is that audits can improve quality assurance procedures in universities but not necessarily improve the quality of student learning (Newton, 2000). Another objection is that QAA reviews can inadvertently emphasise quantitative measures. These reviews are often taken up by the media which assumes that its readers are not interested in complicated reports. Instead the press focuses on such details as pass rates, awards, and ultimately the score out of 24 which review teams use to judge a programme. The result is a leagues table showing the best and worst universities.

In this paper we are particularly interested in defining quality assessment because we feel that it holds the key to a quality degree. The questions we want to ask in this paper are:

- Do external audits encourage teachers and students to continuously develop better ways of encouraging and assessing meaningful learning, or
- Do they rather lead to a 'culture of compliance' whereby assessment methods endorsed by the audit team become those used by the teachers and learners.

There need not be a conflict here. Hopefully reviewers, most of whom are teachers and students themselves, will espouse the best possible forms of assessment. Unfortunately, the process is bigger than the individuals involved. Quality audits are born in a political arena and their purpose is political as well as pedagogical. Comparisons may be odious but HSV is committed to making them. If a programme wants to compare well there is a danger that it will select assessment methods that are tried and true rather than innovative, that are easy to compare rather than specifically suited to one group of students and their particular context.

When funding is attached to the results of such audits, the audit itself and how to give the auditors what they expect, can take over from creating a better course. The danger is pedagogical compliance rather than pedagogical creativity.

Defining quality

In order to sustain such an argument we need to define what we mean by quality. We also need to refer to the definitions that agencies, like QAA and HSV use. The new Oxford dictionary defines quality as excellence or a degree of excellence. In this meaning quality can also be thought of as the 'best of its kind', a standard against which similar things are measured. The problem with the dictionary definition is that one has to go on and debate what is meant by excellence. In the area of higher education, some people, Boyle and Bowden (1997) for instance, think that debate on the term itself is a waste of time. 'Most progressive thinkers', they say, 'and those motivated by positive practical outcomes, have moved on from the endless esoteric debates on conceptions of quality'. We acknowledge attempting to define quality can drive you mad (as it did Phaedrus in Pirsig's book, *Zen and the Art of Motorcycle Maintenance*) but think it is important to try. Boyle and Bowden's view is that we should simply accept 'fitness for purpose' (Ball, 1985) as the most workable definition. But this definition has its own problems because it ignores multiple or competing purposes. Even if the debate about quality can not be resolved it is essential to engage in it. Not to do so is to give the field to those who have the power to enforce their purpose.

Quality as 'excellence or goodness' was the definition preferred by philosophers. 'Fitness for purpose', is the definition preferred by business people. It is, incidentally, in the world of business that both the definition 'fitness for purpose' and the quality assurance movement has its origins. The definition works well because the aim of business is straightforward. Most businesses intend to make a profit, and the steadier and more long-term those profits are, the better. Quality is therefore assured when the process of production, distribution and sales fits the stated aims and purpose of the company - the largest possible long-term profit. In determining quality assurance procedures a lot of emphasis is placed on criteria or attributes that maximise profit. Things such as speed, economy and efficiency tend to take precedence. Another way of describing this conception of quality is 'value for money' (Harvey and Green, 1993). 'Fitness for purpose' has gained such wide acceptance that it is applied today in more complicated contexts – universities for example. The British, Hong Kong, Netherlands, Scandinavian and Australian quality assurance agencies make use of it. But 'fitness for purpose' begs an extremely important question. 'Whose purpose?'. In industry the answer is easy: the shareholders or owners' purpose. But who owns the university? And what is its purpose?

Perhaps we can illustrate the problem of how different people can see different purposes for the same activity by means of an anecdote. The vice-chancellor of a large technical university received complimentary tickets from Volvo for a performance of Schubert's unfinished symphony. Because he had a prior engagement he gave them to his total quality management officer, a man called Sven whom he had recently recruited from the car maker's staff development section. The next day in the coffee room he asked Sven what did he think of the quality of the performance. 'Well' said Sven, 'I thought you might ask, so I sent you an email about it'. When the boss opened his email there was Sven's message under the heading, *My thoughts on Schubert's unfinished symphony*. 'For considerable periods of time', he wrote, 'the four oboe players had nothing to do. The number should be reduced and spread over the whole orchestra thus eliminating peaks of activity. All of the twelve violin players were

playing identical notes. This seemed an unnecessary duplication and the staff of this section should be cut drastically. No useful purpose is served by repeating with horns the passage that was already played by the strings. If all the redundant passages were eliminated the concert could be reduced by half. Had Schubert attended to these matters he would probably have been able to finish his symphony after all’.

An important point of the story is that the quality expert saw the performance in terms of his own prevailing paradigm, a paradigm that he accepted unthinkingly after years of trying to improve production in a large car manufacturing plant. His paradigm suited Volvo. It might even suit the student enrolment process at a big university where speed and efficiency are desirable characteristics. But Sven’s paradigm did not suit a musical performance. Nor, we would argue, would it suit the quality assurance of teaching, learning and assessment at a university. Sven had very little expertise when it came to judging music. Judgement about the quality of a musical performance is better left to a music critic. Just as judgement about the assessment of learning is better left to teachers and learners.

‘Fitness for purpose’ is only helpful if we can clearly define the purpose of the university and the various activities within a university. Assessing students, for example, is a different sort of activity from enrolling them and the quality of each process has to be defined by different sorts of criteria. In the past, when universities were elite institutions rather than a part of mass education, there was consensus about its purpose. The university’s purpose was to pursue new knowledge, pass on established knowledge and serve the wider community. Ideally it pursued these goals in an atmosphere of academic freedom. Today this ideal lives on in the mission statements of most universities. The large technical university, mentioned above, declares that ‘All of our activity is to be characterised by academic freedom, responsibility and the will to achieve results of the highest international quality’.

In industry, ownership and purpose are intertwined. This is not so in the modern university where both ownership and purpose are contested issues. Governments, especially around election time, insist that they are the representatives of the people and since the people’s taxes pay for the university’s infrastructure and pay for the universities’ running costs the people, in effect, own the university. Since industry shares the cost of university research with government (about 50-50 these days) they too can lay claim to part ownership. The university council, especially if it is endowed with private funding, feels it owns the university. And the administrators, teachers, researchers and students who make up the university feel that they are the rightful owners.

If all these groups agreed on the university’s purpose then defining quality as fitness for purpose would not be so problematical. Unfortunately there is disagreement both between and within these groups. Each group has its own view on the role and purpose of the university and it changes over time. The government acting on behalf of its electorate says that the purpose of the university is professional training and useful research. Industry, which benefits from both, tends to agree. University councils might pay lip service to academic freedom but in a world of increasing competition and shrinking budgets, they see their purpose, more and more, as balancing the books, even making a profit.

One might imagine the lecturers, researchers and students who compose the university would have a common purpose. But this is not always so. For some the purpose of the university is to graduate innovative and independent thinkers, carry out pure research, challenge existing

paradigms and create new ones. For the more pragmatic the purpose is to produce graduates acceptable to the various professional bodies that will register them, and carry out commissioned research. From the above discussion it is clear how complicated it is to decide on 'purpose'. In defence of those who want to pursue an 'esoteric' debate on the concept of quality, it could be argued that deciding on the purpose of higher education and what fits it, can be just as interminable as what is excellent or good in university education.

The quality of assessment

In the rest of this paper we want to argue that if the purpose of a university is about initiating, fostering and promoting learning then a focus on the quality of a university's assessment system will help us to better understand both the notion of 'quality' and how we can assure it. Our position is based on the fact that assessment drives much of what happens in a university. It involves the same stakeholders we have mentioned above, although this may not be obvious at first glance. A focus on assessment also allows us to discuss some of the definitions of quality we have mentioned so far (excellence, the best of its kind, a desirable standard, fitness for purpose, value for money) and introduce a few more (continuous and/or transformative improvement).

The definition 'fitness for purpose' forces us ask 'What is the purpose of assessment in higher education? The simple answer should be to discover if students undertaking a course of study really understand and can apply what they have learnt. This is the pedagogical purpose. But of course it can have a political or a social or an economic purpose. Exactly what those purposes are depends on whom you ask. The government, the taxpayer, industry, the university, the dean of a faculty, the teacher or the student.

Their answers will be reminiscent of those given above in regards to the purpose of the university itself. Government and industry would like lecturers to assess students so that their marks reveal whether or not they can be employed to do a job as engineers, teachers, social workers, medical staff and so on. If graduates are graded this is even better since it can be used in the job selection process. Assessment procedures that produce statistics which can be compared across institutions are particularly appreciated. Summative, norm referenced examinations in a secure, controlled environment are ideal for this.

The more pragmatic teachers mentioned above are often in agreement with government and industry. So are those who lack job security or are keen not to rock the boat. Researchers who have some teaching duties also tend to 'go with the flow', especially when the flow is along a deep channel dug out by generations of closed-book, end-of-course exams. Assessment for many of these people is about grading students and complying with institutional and government requirements. The system works, why change it.

Teachers who are fascinated with their subject and are interested in promoting learning for learning's sake tend to favour formative, criteria referenced assessment. The more varied and more continuous it is then the better it fits their purpose. Of course this takes more time, breaks with tradition and is not always popular with other teachers who may have an unwritten agreement with their students that traditional exams are the simplest way of grading or gaining accreditation. The latter would argue that closed-book, end-of-course exams provide an even playing field in what is basically a competitive system. They often feel that essays, projects and take-home exams open the door to cheating. Such fears are fed by the fact that there are already internet sites that are willing (for a price) to help students do just that.

'Fitness for purpose' forces us to define the group we are talking about. It also forces us to see that quality assessment for one group may be poor quality for another. But definitions like 'goodness, excellence or the best of its kind' take us a step further. What is the best reason for assessing learning? Which sort of assessment is the best of its kind? If the answer is assessment that tests for a deep understanding of a subject, that encourages meaningful, transferable learning, that diagnoses gaps or mistakes in a student's knowledge, that builds a learner's confidence, that discovers innovative or creative answers then we can begin to list some of the defining qualities of good assessment.

A defining quality of excellent assessment, given the above, is that it should be formative. In other words students should get immediate feedback and be given the chance of learning from their mistakes. Other qualities are that it should be continuous or at least spread evenly over the whole course. It should be valid. In other words it should serve the purpose we intend, whether that is formative or summative. In the case of an achievement test it should provide a fair sample of what has been learnt. It should also be reliable. A similar assessment taken at different times ought to yield similar results.

To demand formative assessment in its true sense flies in the face of the competitive nature of higher education, at least as it is currently constituted. The present system is designed to sort the sheep from the goats. With formative assessment learners compete against themselves rather than others. One could argue that industry, government and universities should set their own summative exams – this already happens with university entrance exams. Let the employer assess capability for the job or the government set an annual exam to decide on the funding of student places. Teachers could then get on with providing 'quality' assessment, in the sense that it fits the excellent purpose of enhancing learning.

It is interesting to note that agencies like HSV or QAA would probably agree with our definition of quality assessment. They even ask for evidence of it when they carry out their audits. Ironically the audit itself smacks of summative rather than a formative evaluation of quality. The intention is to give feedback to the particular department and university under review. But this purpose is muddled by the need to compare universities. Because audits have to be published a bad report will damage a university's reputation whether or not it rapidly repairs the defects identified in the report. Audits, like end-of-course exams, are highly motivating. But they also encourage a type of compliance that looks more like rote learning than deep understanding. Like their own students, university staff prepare for their particular exam by looking at previous audits, by asking other departments what they did to pass and by anxiously gathering as much evidence as possible to ensure the highest grade from the quality assessors. Like their students they want to give the auditors what they want. And, like other summative examinations, the audit's main function becomes one of inspection and grading.

Conclusion

Few universities or the teachers who work in them dare defy this system. In today's academic world the one who pays the piper calls the tune. And the tune today is economic accountability not academic freedom. Increasingly in higher education, decisions about quality assurance depend on a government minister and his/her departmental bureaucracy. Today quality is defined and monitored by those outside the university. Too often, despite the stated policies, it is quantity not quality that is measured. In the research area this means the number of refereed journal articles and the size of research grants. In teaching the number of graduating

students and the grades they attain. This in turn steers the sort of assessments that are used in universities.

The danger of this is that those within the university will abrogate their responsibility for defining and assuring the quality of assessment. External reviews are here to stay. It is important that university staff be involved with them in a positive way. Quality audits can improve the quality of teaching, learning and assessment. Nevertheless, researchers, teachers and students - those who have most at stake, given the espoused purpose of the university - often feel disempowered by external quality assurance, especially if their purpose tends to be political rather than pedagogical. In this paper we have made a plea for giving experts in the field of higher education (teachers and students) a chance to agree on the main purpose of assessment. A chance to identify what makes particular assessment practices excellent or the best of their kind. Our hope is that pedagogical ideals such as the meaningfulness and transferability of learning will outweigh political ones. That researchers, teachers and students in universities will comply with intrinsic standards of excellence rather than with those imposed from outside.

References

- Ball, C (1985) What the hell is quality? In D Urwin (ed), *Fitness for Purpose: Essays in Higher Education*, SRJE and NFER-Nelson, Guildford.
- Boyle, P and Bowden, J A (1997) Educational quality assurance in universities: an enhanced model. *Assessment and Evaluation in Higher Education*, 22 (2), 111-21
- Harvey, L and Green, D (1993) Assessing quality in higher education: a transbinary research project. *Assessment and Evaluation in Higher Education*, 18 (2), pp 143-148
- Högskoleverkets rapportserie (2001:9) *From quality audit to quality assessment: the new evaluation approach for Swedish higher education*. Stockholm: Högskoleverket
- Newton, J (2000) Feeding the beast or improving quality? *Quality in Higher Education*, 6, 153-163.
- Pirsig, R (Perennial Classics Edition 2000) *Zen and the art of motorcycle maintenance: an inquiry into values*. Harper Collins, New York.

Copyright © 2002 Michael F Christie and Thomas Stehlik: The authors assign to HERDSA and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to HERDSA to publish this document in full on the World Wide Web (prime sites and mirrors) on CD-ROM and in printed form within the HERDSA 2002 conference proceedings. Any other usage is prohibited without the express permission of the authors.