In January 2006 I was asked to write a position paper for discussion by my Faculty Graduate Affairs Committee. A proposal had been made to introduce a quantitative system of assessment (Distinction/Pass/ Merit/Classification and percentage marks) within all Faculty of Arts taught MA programmes. I was asked to give reasons for maintaining the current Pass/Fail system in MA Fine Art. In doing so I’ve had to trace some relevant historical and background factors as well as presenting various threads of argument about knowledge and learning. As the paper raises a number of significant issues I’m circulating it as a contribution to the ongoing debate about assessment in the arts. I apologise for the length of the paper!

The paper raises a number of concerns about moving from the currently validated Pass/Fail grading system used within MA Fine Art, to a Distinction/ Merit/Pass/Fail system that necessitates the use of percentage marks – a move from a predominantly qualitative system that foregrounds verbal reporting on achievement, to a quantitative system that foregrounds numerical scores as a measurement of achievement.

I would like to argue in favour of retaining the qualitative system. In order to do this I’d like to raise a number of issues that arise in relation to learning, knowledge and interpretation in the field of art education, beginning with a brief historical survey.

Brief genealogy: competition, quality control & gatekeeping

Broadfoot (1996) and others (Holsten, 1979; Ball, 1992) have described how the development of assessment procedures in the nineteenth century was determined by a belief in the need to establish comparability in the rapidly growing professions and commercial institutions of the time: “the concern was reflected in the [introduction of] qualifying examinations for entry to particular professions and institutions […] The pressure of numbers, together with the need for comparability meant that such examinations were formal written tests” (Broadfoot, 1996: 31). This desire, driven by the demands of employers and professional bodies to impose strict selection regimes on the workforce, has continued, though with some changes of emphasis. For instance, Broadfoot (1996: 28) argues that,

as the competitive element of assessment has increasingly come to dominate over its role in the attestation of competence, content has tended to be determined by its legitimatory power rather than its relevance to particular tasks […] The preoccupation with the reliability of assessments has tended to eclipse concern with validity.

In other words, the pressure for ever more reliable, hence quantifiable, assessment systems has pushed aside the question as to whether such systems are valid or effective, let alone meaningful. This competitive model, so fundamental to capitalism in which educators’ act as gatekeepers for entry into the higher echelons of commerce and professional employment, continues to dominate at all levels of education.

In a recent authoritative series of papers on assessment from the LTSN Generic Centre, (Brown 2001: 6) the three main purposes of assessment are given as:

- to give a licence to proceed to the next stage or to graduation;
- to classify the performance of students in rank order;
- to improve their learning.

Note the importance of the gatekeeping function, and the classificatory and competitive imperatives displayed in the first two bullet points. Note also that the improvement of learning is third in this list! Brown points out that these purposes “may overlap or conflict”: there is evidence that quantitative and summative assessment does not improve ‘deep’ learning to the extent that formative and qualitative assessment does, indeed there is some evidence that it impedes deep learning and encourages surface learning. (see below)

The reliance on quantitative assessment data in fields in which the body of knowledge is largely quantitative and clearly determined, may be justified or even necessary – though the number of subjects where such conditions pertain is very small: eg. mathematics, ‘hard’ sciences, statistics and maybe aspects of technology, engineering and medicine – though even here there is much that is unquantifiable and contested. To transplant or impose such quantitative methods on other fields is, however, not justified, effective or necessary (except to satisfy the need for gatekeeping, competition and selection as outlined above).

There is no intrinsic educational value to such methods, and indeed they seem to fly in the face of government educational rhetoric that currently prioritises ‘student-centred learning’, ‘creativity’, ‘choice’, ‘life-long learning’, and ‘widening access or participation’ – all of which seem to be at odds with the privileging of kinds of knowledge that are quantifiable and suitable for statistical analysis. The marginalisation of non-measurable, or difficult to measure qualities and aptitudes is only one of the negative effects of over-reliance on quantitative assessment.

As Broadfoot (1996: 8) points out, formal assessment particularly in summative and quantitative modes is now so integral to mass education that “any attempt to release education from the constraints of the educational paradigm […] is likely to need a fundamental rethink.” Nevertheless, it seems to me we should avoid the use of such modes where possible, and retain or privilege formative and qualitative modes at every opportunity.

Formative & summative assessment, deep & surface learning

Given the importance of qualitative enquiry, experiential learning, inter-subjective dialogue and creative practice in arts education, it is surprising and seemingly inconsistent, that, when it comes to assessment, quantitative modes are prioritised. This is even more surprising when one considers the rhetoric of many contemporary critical discourses (as taught within most HE institutions) which place emphasis on hermeneutics, constructivisms, pluralism and relativism – all of which point to the conditional nature of knowledge and the provisional nature of interpretations and judgements. It seems odd that programmes of study which, for instance...
features

16/17

advocate qualitative enquiry, discourse analysis and perspectivism, should employ modes of assessment that are rooted in positive beliefs in objective measurement and statistical data.

It is widely accepted in educational development circles that ‘deep’ learning is what educators should be developing in their students, as opposed to ‘surface’ learning. The latter is characterised as the passive accumulation of information, to be memorised and regurgitated at assessment points. The former, characterised as active understanding - the ability to identify underlying principles and patterns, and to apply knowledge to new situations. It is self-evident, (and research supports this view), that a reliance on, or the giving of too much value to, summative assessment, particularly of the quantitative kind, engenders and reinforces surface learning, while formative, qualitative assessment tends to promote deep learning. Surface learning involves reproducing information or opinions; it promotes a convergent learning process, largely determined by the teacher and the requirements of assessment. Deep learning is about making sense and meaning. A more dynamic, open and divergent process largely determined by the learner. Deep learners also tend to learn how and are therefore more capable of critical independence and self-direction.

While formative assessment informs and energises learning, summative assessment often detracts individuals from their learning. Individuals can become alienated from the learning process and side-tracked by the pursuit of false goals - including the acquisition of marks or grades (rather than understanding skills) - learning objectives determined by, and for, others (hurdles to be jumped); and the meeting of arbitrary deadlines that take little account of other circumstances in the life of the individual learner, and that reinforce short-term ‘surface’ learning rather than long-term deep learning.

Given these widely acknowledged correlations between summative/ qualitative assessment and surface learning, and between formative/ qualitative assessment and deep learning, it is widely accepted that summative and qualitative modes dominate the education system.

Quantitative & qualitative assessment

Observation, evaluation and measurement in the fields of art and learning are not precise or objective. They are value-laden subjective processes involving two or more objects of power - most obviously student and assessor. Qualitative assessment methods usually comprise verbal descriptions and analysis of student behaviour and production (spoken and written), providing a critical commentary, advice and other feedback, useful as a formative aid to learning. Quantitative assessment comprises numerical scores or grades that are intended to measure relative achievements, with the understanding that such scores can be compared. The indeterminacy and unpredictability of learning is very apparent in art and learning. Assessment procedures will likely be less certain of the fairness of the marks that have been awarded - even though it is not unusual for there to be major differences between the two sets of marks.

The zone of interpretation

I believe that in art, as in all learning, time, indeed the whole of a single mark - the hole of an assessment. In some ways, the process would much more transparent and informative to the student and evaluators. Assessment procedures that provide comparative data for ranking students against each other in a given cohort, or across cohorts, year groups or different subjects.

Broadfoot (1996: 8)

Formal assessment is now so integral to mass education that any attempt to release education from the constraints of assessment procedures would be likely to result in the collapse of the system itself.

Assessment: advocacy, debate and enforced consensus

I have observed, and reluctantly participated in, too many summative assessment meetings to believe they are anything but erratic, inconsistent and, at times, absurd. Such meetings reflect the impossibility demands of two opposing systems of assessment, the qualitative and the quantitative, and they highlight the inherent difficulties in translating qualitative interpretations and provisional judgements into quantitative scores and absolute measurements. Participants arrive with more or less certainty about the fairness of the marks they wish to give to each student’s work. On most occasions they leave the meeting more or less certain of the fairness of the marks that have been awarded – even though it is not unusual for there to be major differences between the two sets of marks.

These differences emerge as the result of the adversarial process of advocacy and argument that characterises most assessment meetings. This process is a mixture of negotiation, rational argument and peer pressure, centred on subjective opinions about the degree to which students have achieved particular learning outcomes, as manifested in the artwork or texts presented for assessment. It is not unusual for two markers to present initial marks related to one student’s work that differ by 50 percent (say 55-5%). I’ve been at meetings at which the discrepancy has occasionally been from 45 to 65 percent). After much argument, counter-argument and compromise the mark finally agreed might well be 60% - a mark that neither of the markers originally thought appropriate and which now hovers on the borderline between grades/classifications rather than firmly within one.

In most assessment meetings there is an alternating pattern of convergence and divergence of opinions, values, interpretations, assumptions, prejudices and insights - energised by the particular dynamics of the group. However this rhythm of debate and open-ended exchange is subject to a strictly enforced necessity for convergence, that is to arrive at a definitive single mark - the hole of quantitative assessment. The process would much more transparent and informative to the student and evaluators. Assessment procedures that provide comparative data for ranking students against each other in a given cohort, or across cohorts, year groups or different subjects.

The continuum of learning: indeterminacy & divergence

If learning is a continuum of cognitive processes, manifested in actions and constructed, then the outcome of learning is a combination of action, construction and reflection. Outcomes may be unpredictable, unknown at the outset of an activity or only become apparent long after the supposed period of learning. Assessment to engage with, and be indicative of, this dynamic continuum then describes qualitative processes of change, transformation and unfolding possibility is likely to be more useful and achievable than attempting to measure the quantity of accumulated knowledge or competencies, let alone more abstract qualities such as creativity and imagination.

The indeterminacy and unpredictability of learning is very apparent in art education, and in other subjects in which creative practice is at the centre of the curriculum. Outcomes-based assessment inherently privileges and...
Features 18/19

“It is very difficult to identify any significant educational value that can be ascribed to quantitative and summative modes of assessment in arts education.”

Each perspective needs to be considered on its merits, as shedding light from a different angle, and in relation to other perspectives, as providing a more rounded picture. No perspective should be considered as definitive or as representing the final word on a particular topic.

Perspectives

Two other views that have wide currency in philosophy and critical theory also have a profound bearing on assessment: perspectivism and reivodicity (or what Rorty sometimes calls “fallibilism”). Perspectivism involves a belief that knowledge is always partial, and that there can be no absolute, objective or complete view of any subject, topic, idea or issue. Our learning is always informed and guided by earlier learning, by our needs, intentions and expectations, and by our beliefs and values. Each perspective needs to be considered on its merits, as shedding light from a different angle, and in relation to other perspectives, as providing a more rounded picture.

Reivodicity

Given the relative, fluid and perspectival condition of knowledge, it follows that all views, theories & opinions are subject to revision. Indeed, effective learning, if it is to avoid dogmatism, prejudice and eventually bigotry, involves a constant willingness to revise, re-think and re-formulate – to be open to new facts & ideas, and to seek out alternative perspectives that are challenging and revivifying. This inherent revivodicity of knowledge has implications for our thinking about assessment. Judgements can only ever be tentative and conditional, subject to continuing revision over time. Assessments are made from a particular perspective, at a specific moment in a continuum of changing views. Any representation or notation of this process (for example, by representing a particular judgement as final and summative, or as a fixed measurement or quantificational “fact” rather than as a qualitative opinion) sought not to be unchallenged.

As Esser-Hall puts it, “Interpretation has no final result and each ending holds a new beginning.” (Esser-Hall 2000; 289) Hence, a continuum of exchanges of interpretations, none of which can be identified as summative. So also, with learning and art-making, and the assessment of these: there can only be a process of reiteration, translation and unfolding of understandings, interpretations and provisional judgements – always open to revision.

It is not surprising that contradictions and tensions are likely to arise from the adoption or imposition of assessment regimes which do not reflect current ideas about knowledge and learning.

Conclusion

It is the belief that the dominance of quantitative and summative modes of assessment in arts education is largely the result of institutional demands for statistical accountability, quality-control accountability and hierarchical ladders of progression (and exclusion). It is very difficult to identify any significant educational value that can be ascribed to them. Consequently we should resist the deployment of such modes whenever possible, and certainly we should not acquiesce to these kinds of demands without questioning their validity.

All of the above concerns, and the educational beliefs and values from which they arise, lead me to consider the use of a threshold mode of summative assessment (Pass or Fail), with the focus on a written report, to be preferable to a hierarchical grading system that focuses on the numerical scoring of quasi-measurements.

References


Feyerabend, P. (1975): Against Method, Verso