

Qualitative rather than Quantitative

The assessment of arts education

John Danvers is Deputy Head of the School of Art and Performance and a Teaching Fellow at the University of Plymouth. He is leader of the MA Fine Art programme and teaches fine art practice and critical studies at graduate and post-graduate level.

He has recently published a book entitled *Picturing Mind: Paradox, Indeterminacy and Consciousness in Art and Poetry*, published by Rodopi.

This paper addresses concerns about quantitative systems of assessment on a taught MA programme in Fine Art. It is clear, however, that the issues raised have significance and application across a broad spectrum of art, design and media subjects and levels.



// *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 employ a quantitative system of assessment (Fail/Pass/Merit/Distinction & 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!*

This paper registers 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 (Hoskin, 1979; Ball, 1992) have described how the development of assessment procedures in the nineteenth century was determined to a large extent by the need to establish competences in the rapidly growing professions and commercial institutions of the time: “this 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 development 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 legitimacy 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 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 ladders of selection, hierarchies of achievement and 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 constrictions of assessment procedures [...] would be likely to result in the collapse of the system itself”. 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, constructivism, 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 »



Above: Students at Cumbria Institute of the Arts.

advocate qualitative enquiry, discourse analysis and perspectivism, should employ modes of assessment that are rooted in positivist 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 reproduced 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 to learn* and are therefore more capable of critical independence and self-direction.

While formative assessment informs and energises learning, summative assessment often distracts 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 and skills), learning objectives determined by, and for, others (hurdles to be jumped), and the meeting of arbitrary deadlines that take little account of differences in the speed at which individuals learn, and that reinforce short-term 'surface' learning rather than long-term 'deep' learning.

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

Quantitative & qualitative assessment

Observation, evaluation and measurement in the fields of art and learning are not precise or objective processes. They are value-laden subjective processes involving two or more, often unequal, centres of power – most obviously student and assessor. Qualitative assessment methods usually comprise verbal descriptions and analyses 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 achievement of pre-specified outcomes or criteria, and which provide comparative data for ranking students against each other in a given cohort, or even across cohorts, year groups or different subjects. As far as the arts are concerned, in the latter case qualitative interpretations or judgements are somehow translated into numerical scores. It is obvious that there are profound differences between measurement and critical evaluation and interpretation. As I understand it 'assessment' comes from the Latin root, *assidere*, meaning 'to sit beside' – in our case, 'to sit beside the learner' – observing, reflecting upon and commenting upon, what is done, how it is done and what is produced in the process of learning.

It is itself a narrative episode in the continuum of learning. A verbal record or report can convey the nuances, complexities and provisional quality of such a narrative, in a way that a numerical score cannot.

Within art(s) education, numerical scores or grades are usually accompanied by verbal reports and feedback, and this is often used as an argument in mitigation of the negative effects of quantitative assessment. However, the value placed on the scores within the institution and, inevitably, within the student group who have been compared and ranked in a way that is absolute and fixed, marginalises and devalues the qualitative commentary.

The zone of interpretation

If, as Barthes, Eco, Dickie, Rorty, Danto, Gadamer¹ and many others would argue, the audience/observer is fundamentally implicated in the making of meaning in art, if the artwork is both the material event or object and the unfolding of interpretations that accompany it, then we are all participants in the making of the work. Therefore we cannot remove ourselves from this implication, we cannot divorce ourselves from complicity in what is a hermeneutical process that, by definition, is unfixed and provisional. There is no terminus to interpretation, and no measurement that can be made that could constitute a summative view. Likewise there is no place to stand outside the zone of interpretation, no neutral position from which to make measurements. Just as Heisenberg and Bohr argue that the observer affects what is observed (in relation to sub-atomic particles) so it can be argued, perhaps with more certainty that this is true in the field of art and learning (where learning can also be considered as a site of interpretation, or, as Ricoeur puts it, "a conflict of interpretations"). To put it bluntly, every time we attempt to measure or quantify a process of learning as manifested in a set of behaviours, a text or an artwork, we are attempting to measure a process in which we are deeply implicated. We cannot separate ourselves from the mutuality of learning – a process of interdependent dialogues, interpretations and actions.

Quantitative assessment excludes the nuances of multiple interpretations and evaluations in favour of an absolute unitary measurement. Any attempt at 'objective' measurement (quantitative assessment) is intrinsically flawed and it inevitably leads to reification, abstraction and generalisation – the opposite of what is probably intended (namely, specific and precise data based on empirical evidence).

Feyerabend raises another issue that may be relevant here. In, *Against Method*, he discusses the incommensurability of many scientific methods and theories – "the lack of a common measure". It is arguable that there may also be incommensurability between the work, ideas, actions and understandings, of individuals with different notions of what art practice is, what it is for, how it should be done and therefore how it can be interpreted and evaluated. These individuals may be students (peer-to-peer), staff (assessors in a position of power) and students interacting with staff in an assessment event.

Formal assessment is now so integral to mass education that any attempt to release education from the constrictions 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 impossible demands of two conflicting 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 finally 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 may differ by 5 to 10 percent – say 55-65, (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, the need to arrive at a definitive single mark – the holy grail of quantitative assessment. In some ways the process would be much more transparent and informative to the student if the marks of each assessor were published and a cluster of marks were awarded for each unit of assessment – not one! This would reflect the variety of evaluations and suggest that the process, and the mark, is conditional rather than absolute.

The continuum of learning: indeterminacy & divergence

If learning is a continuum of cognitive processes, manifested in actions and constructs, then the outcome of learning is more learning, a continuance of action, construction and reflection. Outcomes may well be unpredictable, unknown at the outset of an activity or only become apparent long after the supposed period of learning. If assessment is to engage with, and be indicative of, this dynamic continuum then describing 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 competences, 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 inevitably privileges and »

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



reinforces outcomes-based learning, and outcomes-based learning tends to develop convergent thinking at the expense of divergent thinking. I have argued elsewhere (Danvers 2003: p. 50–51) that divergent learning and teaching, develop and promote divergent thinking, and are very distinctive characteristics of art education:

Learners are encouraged to progressively extend the arena of possibilities within which they operate, not to seek enduring solutions or answers but to open up unfamiliar territory and new ideas. By encouraging divergent thinking, trying out different ways of doing and making, and exploring different meanings and interpretations, learning is experienced as a continuum of changing opportunities for revision, renewal and self-constitution. Individuals explore and articulate a range of different ideas and material constructs within a framework of collective experimentation, risk-taking and mutual responsiveness. Outcomes are sought which are more rather than less unpredictable. The emphasis is on inventiveness, innovation and going beyond the status quo. Individuals and groups within a particular cohort may develop radically different modes of learning and signification grounded in divergent beliefs and values. In contrast to convergent learning in which learners are drawn towards a common body of knowledge, beliefs and values – towards definite conclusions and pre-established solutions – in which differences of opinions, ideas and practices may be discouraged, and risk-taking minimised.

This quote is extracted from a paper that seemed to articulate commonly held views amongst academics in the arts. One of these views was that indeterminacy and improvisation were two other important characteristics of learning in art (and design):

Art and design practices often tend to manifest high levels of indeterminacy, and make use of improvisatory modes of thinking and action. On many occasions artists may have no clear objective in mind when they embark on a piece of work – other than to produce ‘something’ or to ‘see what happens’. While making use of established patterns of production and ways of thinking, they respond to all kinds of stimuli and changing circumstances. Both the responses and the stimuli may be unpredictable – indeed the unexpected is something that is actively sought. The focus and ‘content’ of the work may emerge in the process of making rather than as a pre-determined objective. Deterministic, goal-orientated ways of thinking and making are often counter-balanced by periods of activity in which outcomes cannot be determined, and open-ended ‘play’ is a more accurate description of what takes place. Playing with ideas, processes, images and materials, the individual may suspend critical, analytical and rationalistic abilities in order to ‘see what happens’, to let things develop in ways which accommodate chance, randomness and intuition. When something emerges that is interesting or unexpected, or with a strong sense of ‘rightness’, it is only then that critical reflection is re-engaged and an understanding of what has happened may develop. Periods of working ‘in the dark’, or when ‘not sure of what is happening’, can be as exciting and productive as periods of lucid control.

These situations are highly complex and unstable, requiring flexible thinking and responsive handling of material processes. Meaning and making are in a state of flux, with countless possibilities rapidly presenting themselves. Decisions may have to be made with little time for conscious thought. Developing the ability to improvise (with ideas as well as materials), and to generate and make use of situations in which indeterminacy prevails, are key aspects of learning within art and design. *The need for time and opportunities to develop these abilities can run counter to the increasingly deterministic emphasis on goal-orientated behaviour in which linear systematic processes lead to predictable outcomes.* [my italics] (ibid: p. 53)

It is odd therefore that more resistance has not been evident in art education to the rapid increase in both outcomes-based assessment and quantitative assessment – the latter an apparent attempt to measure learning processes that are often indeterminate, and to impose summative judgements on open-ended enquiry.

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.

Perspectivism

Two other views that have wide currency in philosophy and critical theory also have a profound bearing on assessment: *perspectivism* and *revisibility* (or what Rorty sometimes calls ‘fallibilism’). Perspectivism involves a belief that knowledge is always partial, incomplete and contingent. 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. No perspective should be considered as definitive or as representing the final word on a particular topic. There can be no neutral, omniscient or ‘objective’ view. Multiple perspectives are to be welcomed. Diversity, difference and pluralism are factors to be affirmed in all educational contexts. While qualitative assessment can take account of different perspectives and articulate nuanced judgements or opinions (in joint reports and numerous formative feedbacks), single numerical scores or grades cannot.

Revisibility

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’ and ideas, and to seek out alternative perspectives that are challenging and revitalising. This inherent revisibility 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 mis-representation or reification of this process (for example, by representing a particular judgement as final and summative, or as a fixed measurement or a quantitative ‘fact’ rather than as a qualitative opinion) ought not to go 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 my belief that the dominance of quantitative and summative modes of assessment in arts education is largely the result of governmental and institutional demands for statistical accountancy, 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 wherever 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. •

While this paper articulates my own personal viewpoint it is informed by comments and concerns raised by staff and students on taught postgraduate courses with whom I’ve had contact over the years (particularly as an external examiner). A number of students made the point that a Pass/Fail system tends to avoid the artificial pressures of more complex quantitative systems usually used at BA level – they emphasised the usefulness of narrative feedback as opposed to the generally debilitating effects of relatively arbitrary numbers and a false sense of competition.

Footnote

1. I’m thinking here of: Barthes’ notion of the ‘writerly text’; Eco’s ‘open work’; Dickie’s ‘institutional theory’ of art; Rorty’s conception of art (and science) as descriptive narratives; Danto’s theory of art as an evolving social-historical construction; and the hermeneutical theories of Gadamer and Ricoeur.

References

Ball, C. (1992) Ladders and Links: Prerequisites for the Discussion of an International Framework of Qualifications, Wellington, N.Z., New Zealand Qualifications Authority

Broadfoot, P. (1996) Education, Assessment and Society, Buckingham & Philadelphia, Open University Press

Brown, G. (2001) Assessment: A Guide for Lecturers, *Assessment Series Guide No. 3*, Learning and Teaching Support Network, Generic Centre

Danvers, J. (2003) ‘Towards a Radical Pedagogy: Provisional Notes on Learning and Teaching in Art & Design’, *The International Journal of Art & Design Education* 22:1 (p.47-57)

Esser-Hall, G. 2000 ‘Perpetual Beginnings: The Role of Phenomenological Hermeneutics in Art Education’, *The International Journal of Art & Design Education* 19:3 (p.288-295)

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

Hoskin, K. (1979) The examination, disciplinary powers and rational schooling, *History of Education* 8 (2): 135-146