

Dumbing Down: Some Ethical Questions in Assessment

Andrew Johnson
Monash University

Andrew.Johnson@CeLTS.monash.edu.au

Abstract: Assessment raises a number of ethical questions for markers, no less than for students. Considerable attention has been paid to the ethical, or more accurately, supposedly unethical behaviour of students in recent years related to the increase in incidences of plagiarism. Less attention has been paid to the competing ethical demands faced by assessors in an educational environment that has undergone rapid and radical transformation. The insistence on maintaining standards, on not dumbing-down courses, or passing underperforming students are among the stronger imperatives which educators at all levels are expected to follow. And yet, the models of assessment assumed in this imperative (normative) may be at odds with models of assessment, such as criterion or competency based assessment, practised in tertiary education. This paper explores the ethical questions that arise in the tension between different notions of assessment, from the perspective of what might constitute a 'fair' assessment practice for students.

Keywords: assessment practice, ethics, standards, fairness

Introduction

The phrasal verb, 'dumbing down', which I've used as the title for this paper is the index of a schism at the heart of education today. Academics, teachers, curriculum designers and administrators can frequently be heard to talk about an inexorable diminishment of the capabilities of students and consequently, the slackening of the intellectual demands education is able to place on students. At the same time we hear parents, or at least those who purport to speak for parents, politicians, employers and other community groups bemoaning falling standards in education. The irony is that the two groups look as though they might be, or should be talking about the same thing, but may in fact be talking about quite different phenomena.

For perhaps obvious reasons, the issue of assessment lies at the heart of the debate: assessment is the point in the educational process at which a student can be measured against a standard, and the point at which standards across a cohort, or across the community as a whole, can be identified and reported. Assessment is also seen as the 'driver' in modern education: it is the factor which motivates students to learn, and the focus of learning from early primary school through to the highest levels of tertiary education (McKeachie et al., 2001; Coffin et al., 2003).

Dumbing down, accordingly, can be understood to mean a lowering of the assessment bar, a reduction in expectations of student performance in assignments or exams. From another perspective, students' performance on a given assessment task is not so much the issue as the content of the education tested. In other words, it's not how the student performs on the given test, but what is on the test itself that matters. Between these two senses of 'dumbing down' there lies a considerable distance. And while the two 'sides' identified above explicitly differ in their account of the causes of falling standards, they often diverge more significantly in their views of the role of education in society, in their understanding of what contemporary values and standards are, and finally, in their views of the nature and purpose of assessment.

Thus, it can be argued that beliefs and claims about falling standards are often based in misunderstanding of ethical dimension of contemporary assessment practice. The debate over standards, the concern about 'dumbing down', necessitates better understanding of ethical issues in assessment.

A fair assessment?

Fairness is paramount in assessment, but much of the discussion of fair assessment is focused internally on the mechanisms of assessing student performance such as the reliability or validity of certain tests, rather than on the broader contexts in which assessment takes place. Liz Hamp-Lyons' (in Kroll, 1990) review of research on assessment

issues in second language writing identifies the first part of the problem succinctly: 'Of course' she writes, 'reliability is important, but as every introductory testing textbook tells us, reliability is a necessary but not sufficient condition for an ethical test' (Kroll, 1990; 73). An ethical test, Hamp-Lyons goes on to argue, is not only reliable, but has content, criterion and construct validity, and takes into account potential sources of unreliability and invalidity in the construction of the task, the situation of the writer, the scoring procedure used and the situation of the reader (or assessor). This comprehensive approach to establishing an ethical assessment practice is enlightening – but before looking more closely at some of these conditions, closer examination of ethical issues raised by the practice of assessment in general is required. In other words, it is necessary to establish the sufficient conditions for an ethical, or 'fair' assessment in more detail.

At the broadest level, the assessor has numerous obligations to the assessed: ordinarily this would involve avoiding bias or prejudice, making exceptions where exceptions are warranted (for lateness or other uncontrollable forces impacting on the students' performance) and so on. Such obligations might be summarised under the single heading of 'objectivity', which is in turn equated with fairness. Objectivity, if it means the exclusion or at least bracketing of interpersonal relations and subjective judgments, is not necessarily

Degrees of fairness and objectivity would appear to differ in various forms of assessment, such as formative and summative, but it does not appear possible to argue, *prima facie*, that one form of assessment is superior to another in respect of ethics, at least where ethics is construed as fairness. An initial confusion needs to be clarified here: confusion of the notion of 'standards' and the notion of 'values': the latter are arbitrarily determined levels, whereas the former, if they have any meaning at all, are seen as necessary and inevitable.

Certainly, it can be argued that some forms of assessment are more or less likely to relate to standards; but neither formative nor summative modes of assessment could be said to exercise a monopoly on values.

Assessment without 'norms' in some shape, is hard to imagine, even criteria based assessment, the student's performance can be judged relative to other students in a cohort, other students in general, or the community as a whole.

Reliability and validity are nonetheless, key issues (Hamp-Lyons;) and these are based in the theory of 'measurement' developed in the sciences, and in particular, psychology (see Murphy and Davidshofer, 1994).

Consideration of reliability and validity has a bearing on the debate between formative and summative assessment, with the former often not seen as being under the demands of reliability and validity to the same degree (feedback can be interpersonal, subjective and individualistic without necessarily being unethical, indeed developments in contemporary ethics tend to privilege the interpersonal and subjective response over the utilitarian notion of the 'common good') Moreover, from this ethical standpoint two further points can be made: 1. no assessment is ethical as such, the question is not how one tests but why, and in whose interests assessment is conducted. 2. education is not necessarily bound to assessment, or assessment is part of the social and institutional context of education in industrialised countries, but removed from issues of authentic 'learning' or 'teaching'. (cf. Pirsig, 1972). Dewey, *Philosophy of Learning* – measurement and standards are a societal and institutional necessity, but not an educational necessity. To put it another way, feedback is an essential part of learning, but grading is not.

There are, however, numerous ethical problems with the blurring of distinctions between formative and summative assessment, whether out of a kind of resistance to participate in summative assessment for pedagogical reasons (which can also be understood as irresponsibility), or a lack of knowledge of the distinction which often happens in assessment of complex, integrated or highly synthesised writing tasks – tasks which do not lend themselves readily to 'summative' assessment procedures.

That said, within summative assessment practices, the nexus between reliability and validity suggests criterion referenced assessment is generally preferable to normative, standardized assessment. The latter in the form of standardized testing may score highly with respect to reliability, but can lack validity (specifically, content validity and in terms of admitting the individual context of the learner, as a variable – cf GSA.) the former, criterion referenced, looks to have high degrees of construct, content, and predictive (or criterion) validity but can lack reliability.

This framework, to the extent that it is practiced by most educators to some extent or another (often intuitively in higher education, and in assessment of writing) sheds light on some confusions (deliberate or otherwise) that contribute to 'standards' debates, but also suggests that the real issue is not so much how to 'maintain' standards but how to manage ethical assessment given the particular, and peculiar situation of education today.

I am thinking here of a number of variables, contexts, situations that have put increasing pressure on assessors.

Competing notions of the role and function of education, and changing ‘values’ eg. competitiveness vs. equity / and postmodern vs traditional accounts of value.

Educators today generally share a view of the liberal democratic tradition in which Education, as Dewey argues, serves the purpose of reproducing values, such as tolerance, critique of institutionalised power and privilege and so on.

Perhaps to an increasing extent, the dominant values in our culture, however, are derived from a laissez-faire or Friedmanite capitalism in which ‘winning’ the competition (so long as one does so within the rules) is more important than participation in a community with shared values. A closer look at the Nelson paper (2004) outlining present governments notion of values and standards warranted.

For liberal democratic tradition, education based in formative and criterion referenced assessment will no doubt seem more appropriate, whereas the latter sanctions high-stakes standardized testing which ranks participants reliably, if unfairly.

Look also at challenges to content and curriculum in the debate over critical literacy, and changes to the curriculum in VCE English, also from perspective of the ‘understanding’ of the role of English literature as a university discipline.

Globalisation and Internationalisation of the Curriculum, the Shift from Elite to Mass education and Professionalisation

This creates difficulties in terms of validity related to content and student context. Expectations of prior learning variables are a particular issue in higher education. Example of the GSA, test of graduate attributes, is particularly revealing in this regard.

Language and literacy standards

Here I discuss further how linguistic insight into the dynamic nature of language, world English, descriptive as opposed to prescriptive grammars, whole language vs. traditional grammar impact on assessment procedures.

Conclusion

Analysing the phenomenon of dumbing down from the quantitative perspective of whether standards are falling or not would appear to be fraught with difficulties. What I am arguing is that it is also of secondary importance – the real issue is recognising the ethical imperatives inherent in the process of assessment, understanding the forces that are currently shaping education, and constructing appropriately responsive assessment practices.

References

Agnew, Eleanor (1995). Rigorous grading does not raise standards: It only lowers grades. *Assessing Writing*, vol. 2, no. 2, pp 91-103.

- Chanock, K., Clerehan, R., Moore, Tim & Prince, A. (2004). Shaping university teaching towards measurement for accountability: Problems of the graduate skills assessment. *Australian Universities Review*, vol. 47, no. 1, pp 22-29
- Cumming, Alister (2002). Assessing L2 writing: Alternative constructs and ethical dilemmas. *Assessing Writing*, vol. 8, no. 2, pp 73-83
- Coffin, Caroline, Mary Jane Curry, Sharon Goodman, Ann Hewings, Theresa M. Lillis & Joan Swann (2003) *Teaching Academic Writing: A Toolkit for Higher Education* London: Routledge.
- Freeman, R. & Roger Lewis. (1998). *Planning and implementing assessment*, London, Kogan Page.
- Hamp-Lyons, Liz (2002). The scope of writing assessment. *Assessing Writing*, vol. 8, no. 1, pp 5-16.
- Hamp-Lyons, L. (2004). Rev of Mark D. Shermis and Jill C. Burstein, Editors, *Automated Essay Scoring*, 2003 in *Assessing Writing*, vol. 9, issue 3, pp 262-265.
- Johnson, L. (2004). Rev of. B. Broad, *What We Really Value: Beyond Rubrics in Teaching and Assessing Writing*, 2003 in *Assessing Writing*, vol. 9, issue 2, pp 181-185.
- Kroll, Barbara (1990) *Second Language Writing: Research Insights for the Classroom* Cambridge: Cambridge and London.
- Murphy, Kevin R. & Charles O. Davidshofer (1994) *Psychological Testing: Principles and Applications*. New Jersey: Prentice Hall.
- Pirsig, Robert M. (1974) *Zen and the Art of Motorcycle Maintenance* London: Transworld.
- Sadler, D. R. (1996) Criteria and standards in student assessment. *Different Approaches: Theory and Practice in Higher Education*. Proceedings HERDSA Conference 1996. Perth, Western Australia, 8-12 July. <http://www.herdsa.org.au/confs/1996/sadler.html>
- McKeachie, W., N. Chism, R. Menges, M. Svinicki & C. E. Weinstein (2001) *Teaching Tips: Strategies, research, and theory for college and university teachers*, 9th ed, D.C. Heath and Company, Lexington, MA.