Editorial

John MacBeath · Lejf Moos

Received: 13 April 2010 / Accepted: 16 April 2010 /
Published online: 12 May 2010
© Springer Science+Business Media, LLC 2010

The four articles which appear in this edition of the journal explore quite different issues in a number of different national and cultural contexts, although bound by a common thread of comparative performance, outcomes and accountability. Two of the papers are set in Nordic countries, one in Japanese higher education, and one visits both the Australia and the United States to compare respective approaches to the contentious issue of standards and standardisation.

That students are best placed to evaluate the quality of the teaching they receive seems unarguable, yet it is never quite that simple as the paper from Peter Burden reveals. The problems which he discusses raise a number of critical issues—the cultural context and prior history of evaluation, the impact of policy context, evaluation purposes both implicit and explicit, the summative or formative nature of the process, and the consequences of the activity for teachers of being subject to student scoring of their performance. Not to mention the reliability and validity of the instruments used.

Burden refers in this paper to a ‘buyer’s market’, which has given the impetus to the introduction of student evaluation of teaching. To what extent then does casting students in the consumer role reinforce a teacher-centered approach which positions teachers as the providers and measures student performance as the sole responsibility of the teacher? To what extent do evaluation instruments take account of the role of students in managing their own learning? Would a teacher who asks too much of his or her students receive a worse score than one who spoon fed them and led them by the hand to good exam performance? There are resonances here with Proitz, who, in the second paper, reminds us of Elliot Eisner’s statement that outcomes are partly dependent on the student, partly on the subject in question and partly on the teacher.

J. MacBeath (✉)
University of Cambridge, Cambridge, UK
e-mail: jecm2@cam.ac.uk

L. Moos
Danish School of Education, Aarhus University, Copenhagen, Denmark
e-mail: moos@dpu.dk

Given that Burden's paper is set in Japan, it is important to be alive to deeply embedded cultural values and how they combine with a policy cadre which is always looking over the wall to international comparison and cherry picking of practices from elsewhere. Student evaluation is 'a good thing' but to make it a good thing requires that teachers are aware of the purpose of the evaluation, that they are consulted on its design, its use and its administration. The same principle applies to students. It is simply good practice for students to be involved in the discussion of purposes, uses and ways in which quantitative survey data can be used as a tin opener to further discussion and suggest potential changes in pedagogy.

In many ways this paper offers an object lesson in what not to do if the purpose of such evaluation policy is indeed about improvement in pedagogy. Or may there be a less benign intent? As a punitive measure to remove tenure in tertiary establishments with diminishing student rolls? Nor can such perceived threat endear the process to teaching staff. And if there is a note of defensiveness among teachers interviewed in this study it is perhaps not too surprising, especially if ratings are seen as linked to contract renewals and or dismissal. Improvement and accountability purposes have never been comfortable bedfellows.

How would improvement and accountability purposes be served without 'standards'? Without standards how could we compare students' achievements, entry for higher education or employment? Without standards what would happen to comparative measures of school performance, league tables and decision on resources, funding and policy priorities?

How are standards set, maintained and applied fairly across large student populations? How many people does it take to set a standard and evaluate the quality of a student's work? At what point does subjective judgment become objective or at least optimally reliable? MacCann and Stanley's paper takes us through this minefield, raising questions as to differing procedures designed to articulate certain levels of academic achievement and how to and to assign students appropriately to different levels, or 'performance bands'.

'The use of standard setting for reporting educational outcomes is being introduced in many education systems around the world' write the authors and illustrate their argument with reference to New South Wales and the United States. We are introduced (at least for some of us) to the concept of 'cut scores', a concept nonetheless familiar to everyone who has complained of a grade which fell just marginally below a pass or a merit, a 'good' 69 rather than an 'excellent' 70, a B plus rather than an A. It is the delicate point of division between a sheep and a goat, between success and failure.

Given the high stakes nature of this process, how to set the standards which assign students to performance band for a given subject area requires more than individual subjective judgement, not even the collective subjective judgment of a small group but, as the authors of the paper argue, a group large enough to arrive at a valid consensus. 'A small group of judges, the performance band standards set could differ considerably from the standards that may have been set had a larger number of potential judges been used.' The smaller the group the greater the effect of the differing backgrounds and assumptions that individual judges bring to their decisions.

The thesis explored is that small group judgments may be more stringent or more lenient than when a wider expert community is involved, less prone to inconsistency

and fluctuations from year to year. Data from their comparative study upholds the thesis, finding ‘vastly more stable judgements from year to year for the very large judging group’ as compared to a smaller panel of judges.

This leads the authors to the conclusion that with dramatic improvements in technology we are now in a position which allows a much larger group of judges to participate in the process. The value added when a wide constituency of teachers are involved in standard setting would ‘no longer be perceived as an arcane process performed at ‘Head Office’ by the bureaucrats and technocrats. There would be more trust in the results of the procedure. Related to this, the procedure would be seen to be transparent’.

There are strong resonances here with Proitz paper. ‘Systems of examination for providing students with credentials at the conclusion of a programme are dependent on society’s acceptance, trust and confidence to be viable. In a situation where the collective does not agree on or accept the system of examination the whole concept of credentials and certification for further education and employability would probably collapse’, she writes. These are commonly linked with ‘outcomes’, the subject of her paper. The term has become so embedded in the daily discourse that it is difficult to see it afresh. Like standards, outcomes are now the defining measure of schools and schooling and students’ success and status. As Proitz points out ‘scholars play a key role in defining, discussing and problematising educational concepts’. That is, of course, what we do. Trying to make life difficult for policy makers who appropriate terminology and make language their own, is our stock in trade. But who bequeathed us this powerful term which has now captured us and won’t let go?

The term ‘learning outcome, Proitz writes, is anchored in the ‘objectives movement’ at the beginning of the past century as well as in the theories on ‘mastery learning’ and the works of Benjamin Bloom of the 1950s. A learning outcome ‘makes possible a refined understanding of the learning process and thus permits a drawing of relatively precise implications for the design of instruction’. So said Gagne, but in its development it has come to serve a dual purpose. It is now put to the service of measuring institutional effectiveness within a competitive ‘league table’ concourse. Proitz offers a helpful analytical framework within which to locate the diverse purposes and functions of outcomes—countering planning and curriculum development on the one hand with accountability on the other, and, on the other axis, process as against a result-oriented end point. It may serve as a dialogic ‘tin opener’, as she concludes, ‘attentiveness to the dominant perspective and orientation is essential to open debate on what constitutes valuable learning.

That this paper comes from Norway it reminds us just how much the Anglo-Saxon world has colonised the language and politics of school reform. As Proitz writes—‘The linguistic factor of the English language as the prime language in the international scholarly debate concerning learning outcomes may represent a barrier for less dominant voices in non-English-speaking countries to be heard. This may overshadow linguistic nuances in how the term is being defined, discussed and understood as well as differing and important cultural and historical aspects that could enrich the debate’.

We remain in Nordic countries, and with aspects of culture, in the fourth article in this edition of the journal. It is one of those five Nordic countries which claims pride

of place in the OECD international league tables, Finland. In the PISA studies, Finnish pupils continue to achieve the highest Nordic (and global) scores in reading, science and maths. While much paper and printing ink has been devoted to justifying, or demythologising Finland's ranking as 'the world's best in terms of both quality and equity', Orsted's paper puzzles over why there should be such a large achievement gap between two nations as culturally similar as Finland and Denmark. Both are, writes Orsted, 'identical' in the basic welfare state model, the GNP per capita, size of population, educational structures and features concerning family life which include birth and divorce rates.

All the Nordic countries share a child-centred approach and a so-called 'class teacher' system whereby a class of pupils remains under the primary care of one teacher for at least 3 years, often much longer, and there is a complete absence of streaming or setting by ability. The one teacher model allows variation in teaching, paired and group work plus wider scope for special projects, storylines, feature days, 'play-to-learn' and hands-on activities.

The beauty of this paper is in its rigorous analysis of the respective classroom contexts and cultures of these two apparently similar systems while identifying deeply significant aspects which separate them. The use of Czikzenmihalyi's concept of 'flow' and its applications is both novel and compelling.

While the presence and degree of 'flow' measured in the Danish and Finnish classrooms are somewhat similar in this study, the contexts in which they occur are markedly different and assume a quite distinctive character in each place. Time, place, relationships (pupil-pupil, teacher-pupil(s), locus of decision-making, pacing of activity and break times all loom large as significant variables. In the Finnish schools, each 45-min lesson is followed by a 15-min compulsory outdoor break during which children are active in initiating and organizing a variety of traditional outdoor games all involving a lot of exercise. In Denmark, there are fewer and much shorter breaks in which the children can decide for themselves if they want to go out or stay inside the classroom. In Finland the school canteen serves a free two-course meal including free milk or juice in the lunch break Breakfast is often available too.

Among the many insights into the importance of health, diet, exercise and modulation of activity, the most salient finding of all is that the key difference lies in how these routines impact on the lowest achieving 25%. In Denmark the free flowing pupil-directed activity, very often outside the classroom in hallways and social spaces, over a 90 min period can prove difficult to manage productively for the least able who need the attention, formative feedback and support that Finish teachers are more likely to offer. Long periods of independent, paired, or group work can, it is suggested, produce anxiety, cause boredom or drudgery for some of the Danish pupils, in particular pupils from low SES backgrounds.

The conclusion of the study which appears to carry weight and to add significantly to the Finland debate is that:

1. The Finnish school system pays more attention to the class as a community of pupils; pupils who must function together and take account of each other.

2. Finnish teachers put a great deal of effort into establishing joint attention in the classroom with focus on common goals and activities.
3. At the same time, much energy is devoted weak learners or pupils with learning difficulties.
4. Feedback and evaluation is mainly the responsibility of the teacher, while in Denmark it is more common that pupils evaluate themselves; a challenge which the weaker pupils may not be able to meet.

While an article dealing with comparisons between two neighbouring Nordic countries might easily be bypassed, this would be to lose the hugely valuable insights which this article offers. It not only helps to explain much of the Finnish ‘mythology’ but also helps to shed light on the perennial debate about standards, outcomes and accountability.