

Editorial

John MacBeath · Lejf Moos

Received: 8 May 2009 / Accepted: 11 May 2009 /
Published online: 19 May 2009
© Springer Science + Business Media, LLC 2009

‘I never realized how important it could be to talk to my students about their experiences as learners in my classroom’.

This statement, quoted in Martin, Scantelebury and Martin’s paper, reveals much about an education system obsessed with the assessment, evaluation and accountability yet one which so frequently fails to get a measure of students’ experience and what learning means to them. It prompts the question ‘How valid are the tools we use to gain insights into the quality of learning in classrooms, in a climate so saturated with targets, performance and accountability that is difficult to see the educational wood for the political trees?’

This volume of EAEA contains a number of papers with a focus on instruments and strategies designed to provide insights into the effectiveness of learning and teaching both with reference to schools and to higher education. In reading across these papers we are asked to consider the extent to which tools and strategies have a common function in these differing contexts, and the generalisability of the principles which underpin them.

The cognitive interview has been developed to address some perceived shortcomings of the survey as a data gathering instrument. Despite the generally accepted processes of survey design to provide results that are valid, reliable, sensitive, unbiased and complete, instruments, no matter how standardized, they rest on certain assumptions, write Wildy and Clarke. The nature of the survey assumes that each item has the same meaning to all respondents, that respondents are able to understand the items, that they have the information being sought and that they are willing to answer the items. In an international climate where recruitment and retention of school principals/headteachers work is of growing concern. Wildy and Clarke’s research focuses on what new incumbents regard as the most challenging

J. MacBeath · L. Moos (✉)
Aarhus University, Aarhus, Denmark
e-mail: moos@dpu.dk

tasks in their early years of appointment and whether they believe they were prepared for these challenges. The International Study of Principal Preparation is a collaborative project among twelve countries (Australia, England, Canada, China, Jamaica, Kenya, Mexico, New Zealand, South Africa, Tanzania, Turkey, USA).

The challenges presented by using standardised questionnaire loom particularly large when applied in twelve different country contexts, each with different linguistic conventions and underpinning cultural reference points. This is where the cognitive interview offers a diagnostic tool which can help to remove, or lessen, the ambiguity in survey items, the interview providing an opportunity to tune into the way respondents process thoughts, feelings, beliefs or experiences. The key purpose is to understand what information and processing respondents engage with as they reflect on their response to each item. This strategy comes into its own with particular force in surveys which involve conceptually abstract terms or which require careful thought and self-reflection. The value of the cognitive interview thus identifies and ambiguities with items before researchers in each participating country go to the expense of producing their own language versions of the survey, and so produces an answer to Karabenick's question 'Do they think what we mean?'

In the case of an international survey to be administered across 12 different countries in which English is not always the first language, the nature of challenge should not be underestimated. These issues resonate strongly with the editors of this journal, a Dane and a Scot who laboured intensely over a common survey instrument to be used in seven different countries. Once having agreed that we had got the language right about as right as we were going to get it, when it came to analysing how teachers had interpreted the questionnaire statements it took literally years to unravel, requiring verbal feedback and discussion, similar to what is implied by cognitive validity and co-generative dialogue. In this volume, with four papers from the U.S and one from Australia, we find ourselves struggling with a common language that we cannot refrain from editing at times. Terms such as administrators, for example, refer to quite different categories of people in the U.K. and the U.S. The word 'instruction' which appears in all of these American papers sits uneasily with readers in other countries and we have at times used the word teaching for the benefit of that international readership.

International collaboration and cross country comparisons can both challenge and inspire but also carry a health warning. Asking educational practitioners to answer the same survey across cultural and political borders assumes that there is inherent comparability. We take it as read that we are talking about the same phenomena. The same is true of research findings that emanate from unfamiliar cultures and from foreign political contexts. The assumption that education is education, that schools are schools and that words carry common meanings is a trap for the unwary. As authors, editors and readers we add a note of caution as to what we can legitimately infer from the instruments in which we invest our faith.

Cogenerative dialogues have emerged as a methodological framework to engage a range of stakeholders in meaningful conversations about teaching and learning experiences. These may take place among teachers or between teachers and students but their context is one of '*being-in*', that is, not detached from the situation through objective observation, teacher rating schedules or post facto questionnaires. The authors of this paper argue that the approach becomes particularly salient where

there is a diversity of background and ‘the teacher population becomes differentiated from the student population’, for example in inner-city schools where there may be social, class, ethnic or gender differences in the perspectives that children and adults bring to classroom encounters.

Nowhere is this more sharply and poignantly illustrated than in the Palme d’Or winning French film *Entre les Murs* (The Class) in which teacher and students alike struggle to find a cultural reference point and language so that there is some common ground for learning (on both sides) to occur. The English translation of the title fails to capture the significance of the French title (within these walls), a reference to the constraining milieu of the classroom in which the space between words can be difficult, or impossible, to bridge. In the final scene of the film, at the end of the school year, a girl waits behind to talk to teacher sitting wearily at his desk. ‘Sir, she says, reflecting on a whole of a year’s teaching ‘Im afraid I have learned nothing’. There are strong resonances with this in Martin, Scantelebury and Martin’s cameo of a teacher’s dialogue with his students in their paper. It is not only a graphic illustration of a communication lacuna between teacher and students but also highlights the need to address that disjuncture openly through a genuine search for the common territory on which teaching and learning may meet. The authors of this paper add: ‘What’s really interesting about the cogenerative dialogues, in addition to getting input from students about my teaching, is that I get a chance as the instructor to offer my suggestions for how *they* can improve their learning. Instead of having a student just fail, I have a way to reach out now and say, how can we each do this better?’

Teaching effectiveness is also the focus of three papers, two in a U.S. context (Phelps and Ding), the other in Norway (NN). Researchers have for decades struggled to identify measures which would give an accurate, valid and more or less reliable reading of what distinguishes more and less effective teaching, and while there is a wealth of data on the subject, the case remains open. Student feedback, teacher knowledge, classroom observation, performance monitoring, test scores, all have their advocates and critics.

The problem posed by Geoffrey Phelps is what criteria to apply in assessing the effectiveness of teaching reading. He refers to Lee Schulman’s work which drew attention to a special amalgam of content knowledge and pedagogy- Pedagogical Content Knowledge (or PCK). Understanding how students engage with what the teacher is trying to convey to his or her charges relies less on the teacher’s literacy level, suggests Phelps, than a continuous quest for the most useful forms of representation in any given circumstance, the most powerful analogies, illustrations, examples, explanations, and demonstrations—‘in a word, the most useful ways of representing and formulating the subject that make it comprehensible to others’. also includes an understanding of what makes the learning of specific topics easy or difficult”

This theme is taken forward in Cody Ding’s paper with the rhetorical question ‘Can bad data can provide good conclusions? Yet, this is in many instances what happens when instruments used to evaluate teaching effectiveness rely on measures such as student outcomes. What we are able to learn from such data is at best minimal and at worst misleading, that is if we are genuinely interested in knowing what kind of teaching leads to what kinds of learning. Data which connects what

teachers have done in the class to what students have *learned*, as opposed to temporarily reproduced, are, says Ding, rarely available in such a way that valid links can be made. The problem which Ding addresses, in common with other contributors to this volume, is the lack of relevant informative data, or the means by which to access it. Many studies are constrained by the fact that evaluation becomes a kind of post-hoc test based often revealing a mismatch between the goals of the educational program and the data on which the evaluation rests.

In deciding the constructs to be measured, therefore, suggests this author, evaluators should carefully consider the issues of *effect durability* (how long the effects last), *effect differentiability* (for whom and on what it has an effect), and *dosage of the program* (how intense the intervention should be) - the triple Ds of evaluation. The conclusion is that only through sustained and nuanced forms of longitudinal evaluation that these inherent flaws will be rigorously addressed.

Ding's question 'Can bad data provide good conclusions is highly pertinent for Eyvind Elstad's treatise on Norwegian reform, the issue which he explores in the final paper in this volume. Norway has, in common with a number of its European neighbours, been frightened into raising standards by the 2006 OECD report which placed Norway unexpectedly low in the international league tables of student achievement. Elstad's paper describes the response of schools to the attack of the media, delighting in league tables of performance and naming, blaming and shaming of individual schools. As he finds, schools improve when there are high stakes consequences and high profile media coverage of 'the worst schools in the county'. The question that remains is what does 'improvement' mean and to what extent do aggregated student attainment reflect long term capacity building and professional development? Whatever the longer term consequence and the durability of reform, as discussed by Cody Ding, the deleterious effect on teacher and student morale in Norwegian schools is not in dispute, and the jury has to remain out as to the real benefits of such high powered accountability policies. That other administrations have backed off naming and shaming as a formative and professionally empowering policy is perhaps a salutary warning.

We have brought these papers together in this edition because there are strong thematic links among them and the issues which they address are perennially pertinent to assessment, evaluation and accountability. 'We are still only in the foothills of our learning' wrote David Perkins some years ago. Today he is perhaps more optimistic but it seems to us, in re-reading these various contributions, that we have still have some way to travel, often uphill, to bring together what we know and what we do.

John MacBeath

Lejf Moos