SHOULD EDUCATORS BE WARY OF EVIDENCE?

JOHAN MULLER, UNIVERSITY OF CAPE TOWN

INTRODUCTION

My principal focus in this presentation will be on technical knowledge produced by education researchers and practitioners. I am interested in innovations. I will not refer here to reflective and hermeneutical approaches to knowledge, valuable as they are. I am interested here in technical knowledge for policy. What is the status of evidence-based policy research in education? Should educational researchers be wary of evidence? For complicated reasons, many of them are indeed wary. Paradoxically, partly for this reason, policy-makers have come to be wary about educational researchers and the products of their endeavours; that is, they have come to be wary about the evidence they produce. The internal reservations about evidence on the part of the education research fraternity has, in one of those unintended consequences so often found in the interface between knowledge and policy, come to fashion a stick to beat them with. The upshot is that there is a *very weak bond of trust* between policy-makers and the educational research community. This is an international phenomenon, but is certainly marked in South Africa.

It should go without saying that this is a most undesirable state of affairs. I will make the case in this presentation that if education researchers are to re-establish a strong role for education research in policy, they will have to find ways to construct a rigorous basis for their evidentiary claims. This will be a complicated business, for institutional reasons that I will go on to discuss. However, I also hope to make it plain that it is achievable; indeed, that there are encouraging signs of progress in this regard.

I will begin this presentation with an analysis of the status and position of education research as a knowledge sector, and compare it to other knowledge sectors. This will reveal some of the inherited possibilities and limits for advance which, I would like to stress, cannot be laid at the door of the present generation of researchers. I will discuss some progress made in buttressing education research in general and with respect to research into effective practices in schools, ending with a practical example with which I have been involved that shows that it is possible to make fairly strong evidentiary claims about education that have direct relevance for policy. I will conclude that this is the only way that education researchers will regain the confidence of policy-makers and hence re-establish a productive basis for engagement with policy.

KNOWLEDGE SECTORS, MODES AND SPILLOVERS

Knowledge growth is a profoundly collective enterprise. Even when a single individual achieves a breakthrough, this is almost always on the basis of what has gone before: we stand, as both Newton and Google Scholar have said, on the shoulders of giants. It follows that the ways in which knowledge sectors organise this collective accumulation is critical. Knowledge sectors differ as to their rate of knowledge accumulation. At the high accumulation end of the spectrum, we find sectors such as biotechnology, parts of medicine and transport, for example. At the low accumulation end of the spectrum, we find management, urban development and education. Low accumulation sectors show a relatively slow rate of production and dissemination as compared with high sectors. The issue is to understand the basis for this difference. Following Dominique Foray & David Hargreaves (2003) I will discuss this difference in terms of their different modes of institutionalisation of knowledge production and transfer, and in terms of their richness in spillover potential.

MODES OF INSTITUTIONALISATION

We can distinguish two modes of institutionalisation of knowledge accumulation, namely: experimental mode and learning-from-practice (See Table 1).

Table 1: Modes of institutionalisation of knowledge accumulation

Experimental mode (science in technology)	**Humanistic mode Learning from practice mode**
Controlled experimentation	Situated learning by doing
Knowledge-technology link	Embodied practitioner knowledge
Off-line – strong institutions of application (such as R&D labs)	On-line – applications on site (such as classrooms)
Highly codified, hence easily transferred,	Often not codified, hence weak transferability

disseminated and absorbed	and dissemination
Strong incentives to produce and disseminate	Weak incentives to produce and disseminate

KNOWLEDGE SPILLOVERS

Sectors differ also as to the speed or rate at which knowledge becomes publicly accessible and absorbed by those other than the originator. This is referred to as spillover. Spillover rate is strongly conditioned by competition and cooperation. Competition in the form of incentives to produce new knowledge creates a strong impulse for researchers to raise their game by developing 'absorptive capacities' (that is, processes for monitoring what else is going on in the sector and maximising the possibility of learning from the innovations). Codification and dissemination is clearly critical to this. Competition thus, by increasing spillover, pushes up the general level of activity in the sector. Where competition is weak, as in education, dissemination, at least to practitioners, is largely by way of 'reforms'. Take-up is weak, the dosage of the treatment correspondingly weak (as I will show presently), and growth and innovation also correspondingly weak. Cooperation among researchers is a second means of fostering spillover.

Sectors that are spillover-rich generally combine competition and cooperation.

To sum up so far:

- Sectors differ therefore as to their relative combination of modes, and their spillover-richness or poverty.
- No sector can productively rely on only one mode. This is particularly evident in sectors such as biotechnology and medicine, where learning from practice is an intrinsic part of the innovation system, even though the scientific mode is dominant.
- In low accumulation knowledge sectors, not only does a weak relation between modes occur, but also antagonism between them. In each case, this weakens the sector even further in terms of:

    - Its capacity for growth and accumulation.
    - Its capacity for dissemination and take-up.
    - A consequent further erosion of public trust.
    - A consequent further reduction of money for research.
    - A further impairment of the potential for virtuous impact on policy.

EDUCATION AS A SECTOR

For structural and historical reasons, education has developed an antagonism between its modes, its scientific and practitioner wings. Neither mode is strong in South Africa, either in terms of its evidentiary claims or in terms of innovation. Historically, the scientific mode has, in South Africa, come to be identified with political conservatism, and its practitioner mode with progressivism. Besides being wrong, this creates a further impediment to their rapprochement.

The climate of the knowledge society has placed pressure on low accumulation knowledge sectors to strengthen their basis for evidentiary claims, and in education globally, there have been some signs of definite movement in this regard. I will discuss this further, but first I would like to take a particular sector of educational research as an example, namely, research into effective practices in schools.

EFFECTIVE PRACTICES IN SCHOOLS

School effectiveness research, including in South Africa, is characterised by considerable classroom-based learning from practice innovation, very little of which is formally codified. This is in contrast to other sectors, such as law, engineering and architecture, where specific-case learning from practice success is frequently codified and widely disseminated. Classroom-based innovations tend to remain context-specific, if not site-specific. There is little formal assessment of the innovations, hence little opportunity for assessment of success and the conditions for it. There is consequently weak spillover and low levels of accumulation and growth.

This latter is further retarded by the small number of trained researchers relative to other sectors (which is changing), the lack of a shared theoretical/technical language, and the paucity of funding. Large-scale randomised designs in naturalistic settings such as schooling are vastly expensive, and large sample studies have until recently been rare partly for this reason. They have also been rare because of the reluctance to

conduct formal experiments, as already discussed. There are weak incentives to disseminate learning from practice knowledge compared, for example, with the incentives for doctors to disseminate their innovations.

It is little wonder that Foray and Hargreaves (2003: 13) conclude: 'The system in education for innovation and the rapid spread of ideas and practices is deeply flawed'.

Again it is worth stressing that these researchers are not saying that the research is not good; they are saying that the institutionalisation of innovation and dissemination is weak. As already mentioned, the pressures for rapprochement between the different epistemic cultures have been increasing. This rapprochement will take a different form in education compared with medicine, for example, because in this latter sector there is a dominant mode and a complementary subordinate mode, whereas in education there are two rather weak antagonistic modes. In education, thus, the evidentiary base for both modes requires strengthening. I will discuss the strengthening of the scientific or experimental mode.

THE CLAREMONT DEBATE

In 2003, the US Department of Education's Institute of Educational Sciences declared its commitment to experimental and some quasi-experimental designs over the plethora of 'soft' methods in the sector of evaluation research in education. The American Evaluation Association (AEA) issued a statement, signed i.a. by Michael Scriven, which sharply criticised the move. A group of senior AEA members issued a counter-statement signed i.a. by Mark Lipsey supporting the US Department. To mediate the standoff, Claremont Graduate University in 2004 organised a debate between Scriven and Lipsey to find common ground.

"Somewhat surprisingly, Lipsey and Scriven agreed that randomised trials are the best method currently available for assessing programme impact (causal effects of a program) and that determining program impact is a main requirement of contemporary (education) program evaluation" (Donaldson & Christie 2004).

Both Scriven and Lipsey recognised the undeniable theoretical advantage of using a randomised design in education programme evaluation, but recognised too that, for practical and ethical reasons, it was not always desirable or feasible (though it is chastening to note that the Campbell Collaboration Social, Psychological, Educational, and Criminological Trials Register includes nearly 13 000 such trials and is growing fast). Nevertheless, Scriven at least believes that quasi-experiments with sound designs could be used to determine valid programme effects. This conclusion is echoed by two publications of the US National Research Council of the National Academies (NRC 2002, 2005). In these two reports, titled *Scientific Research in Education* and *Advancing Scientific Research in Education* respectively, as in the Claremont debate, we see the US education research community girding its loins to toughen up its evidentiary base and the training requirements at graduate level that this will require. This debate is also beginning in South Africa.

THE KHANYA EVALUATION

I conclude with an example from work I have recently been involved in. It is often not possible to conduct a true experiment in education settings, largely because interventions are started before the researchers arrive on the scene (Taylor, Muller & Vinjevold 2003). Researchers are therefore mostly not able to assign treatment and control conditions randomly to a sample population. Nevertheless, it is usually possible to randomly select from within the treatment population and from within a matched control population. This form of investigation, called a quasi-experiment, is increasingly common in education, and although it does not permit making causal inferences, does at least allow fairly indicative correlational inferences.

In 2003, a colleague and I were commissioned by the Western Cape Education Department (WCED) to establish whether a rather expensive investment in IT was worth the time and expense. A chosen sub-set of schools, all of them serving poor disadvantaged neighbourhoods, had been given computer laboratories, their teachers specifically trained, WCED support staff specifically allocated to help them, and software provided for a range of subjects. We decided to investigate the Mathematics software, called Master Maths (MM), and the question we sought to answer was: Does the administration of MM software to a random sample of treatment schools measurably improve learner performance in maths, when compared to a random sample of matched control schools? The study found that there was a strong correlation between the amount of time spent, and number of sessions spent, logged on to MM (dosage) – that is time on task – and improvement in maths performance from grade 11 to grade 12. The relationship between amount of time spent on MM and improvement in maths performance is positive, statistically significant and moderate in strength (see Figure 1 and Table 2).

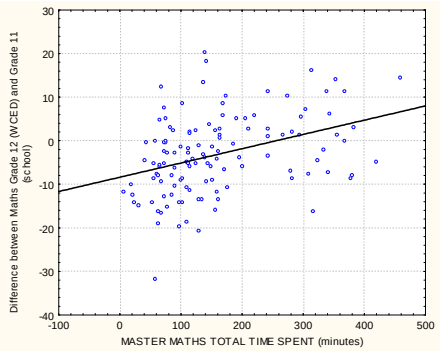

Figure 1: Relationship between time spent by learners on Master Maths and their improvement between Grade 11 and Grade 12 final mathematics examinations

Table 2: Pearson product moment correlation coefficients for the relationship between time spent on Master Maths and improvement in mathematics performance

	Difference between results on Grade 11 and Grade 12 mathematics examinations
Time spent on Master Maths	$r = 0.37$ ($p < 0.001$, $n = 125$)
Sessions on Master Maths	$r = 0.36$ ($p < 0.001$, $n = 125$)

The study also found that the intervention was delivered at widely differing strengths, and overall at an alarmingly low dosage level: treatment learners were logged in for an average of two and a half hours per six months. This low dosage level is a common feature of IT interventions, and of education interventions in general, and is a confounding factor when assessing the impact of any education intervention.

The result gives a firm message to the WCED that their investment is paying off, and that it would pay off even more if they took steps to ensure that the dosage of the intervention was maximised. Did they heed the message? It is hard to say for sure. The expensive Khanya project continues, but whether the dosage has increased we cannot say. Did the WCED fully appreciate the findings? Again, it is hard to say, but we suspect the answer is equivocal. The relative unfamiliarity of education policy-makers with reading the implications of hard data will have played a role. This simply underlines the extensive public role of researchers to make sure that their results, and the policy implications, are properly grasped by the public custodians of social services provision, but first they must have a valid, reliable evidentiary base with which to bargain. Anything else leaves them stranded in an increasingly suspect ivory tower.

REFERENCES

Donaldson, S. & Christie, C. 2004. 'The 2004 Claremont debate: Lipsey vs. Scriven. Determining causality in program evaluation and applied research: should experimental evidence be the gold standard?' [Online] Available at: www.cgu.edu/pages/465.asp.

Foray, D. & Hargreaves, D. 2003. The production of knowledge in different sectors: a model and some hypotheses. *London Review of Education*, 1(1): 7–19.

Louw, J. & Muller, J. with Tredoux C. 2004. *Learner Performance and MasterMaths*. Khanya Evaluation Project. University of Cape Town.

National Research Council. 2002. *Scientific Research in Education*. Committee on Scientific Principles for Education Research. R.J. Shavelson & L. Towne (eds). Centre for Education. Division of Behavioural and Social Sciences and Education. Washington, DC: National Academy Press.

National Research Council. 2005. *Advancing Scientific Research in Education*. Committee on Scientific Principles for Education Research. L. Towne, L. Wise & T. Winters (eds). Centre for Education. Division of Behavioural and Social Sciences and Education. Washington, DC: National Academy Press.

Taylor, N., Muller, J. & Vinjevold, P. 2003. *Getting Schools Working.* Cape Town: Pearson Education South Africa.