What does good education research look like? Some beginnings and debates

or, why education research is not as straightforward as finding a cure for cancer...
When I begin in this way, it is not because I believe everyone is already an expert on the issue of good research, that they have nothing to learn; but nor is it about showing that they have somehow got it wrong, that they do not know what they are on about. Rather, I think that this initial snapshot of the ideas and values people bring to discussions about education research sets up something important about what type of a ‘thing’ this research activity is – at least in a field like education.

The first thing that can be said in relation to my starting point is that the instruction ‘give me an example of a piece of good research you’ve come across’ is a meaningful one. Whether or not they can explicate their reasons, in giving me examples people show that they have some criteria by which they decide what counts as ‘research’, and some further criteria by which they judge some research as ‘good’. Moreover the question is meaningful both to those who have spent time studying research and research methodology and also to those who have not. A second evident point is that people come up with very different actual examples of ‘good research’: their answers are diverse. What to some members of the class seems like an example of good research (gathering older women’s stories, for example), to others can seem like second-rate research or as not ‘research’ at all.

A straw poll of how a varied audience of non-experts think about education research is one useful starting point for recognizing the context in which education research operates. This is not a field in which the quality of one’s work is judged only by one’s academic peers; neither is it a field in which the success of research is established in only one way (finding a cure; making a bomb). Public debates rage about what issues matter in education, and these mediate and are mediated by political decisions, funding, emphases in academic appointments. And politics and values are part of these debates: people have different views about whether a topic is important; whether an argument or finding is an ‘advance’.

But neither are the examples of good research that people furnish simply free-floating expressions of personal preferences. Not everything counts as ‘research’. When I ask people to say a bit more about why they chose their particular example of good research, three themes tend to recur. Sometimes people consciously choose all three; sometimes they emphasize only one or two of them. The three themes are:

1 that the research was technically good: it did something very systematically, was ‘tight and convincing’, or was impressive in its design; was ingenious and creative in its methods;

2 that the research made a contribution to knowledge: that it established something that was not previously known in this way; for example, it proved something convincingly; or it showed something that changed our understanding of things; or it successfully put on the agenda a new
type of question or set of questions; it changed our way of looking at
certain things;
3 That the research achieved something that mattered – either universally,
or specifically to the person giving the example (for example, it had
obvious benefits to health; or it disrupted racism; or it generated useful
evidence about the value of one approach to a particular area of
teaching compared with other approaches).

Judgements about education research involve judgements about
research and about education. They commonly involve some con-
sideration both of the methodological (how well it was done) and of the
substance (what it achieved). At the heart of the fierce debates that rage
currently about education research, both in the community and in the
research literature, are attempts to make claims for particular criteria or
particular standards in relation to both of these. The debates and dis-
agreements in themselves are not signs of the impoverishment of
education research, but of the kind of thing it is. In the next section, I
want to argue this point further by taking some very popular (and to
many, uncontroversial) positions where people have tried to define ‘what
is good education research’ in terms of a single particular criterion, and to
show why I consider that these do not work. It is not, precisely, that these
ideas, which have a lot of ‘common sense’ appeal, are wrong. It is more
that they obfuscate, or that they try to tie up an answer by setting up one
aspect of the broad agendas of education research as if it could adequately
legislate for the whole. But the claims themselves are useful illustrations
of some of the ways that judgements about good research in education
are made, and allude to practices apparent in different contexts that will
be discussed in the second part of this book.

Claim 1: We can measure ‘good education research’ by its
contribution to learning

A few years ago, the Director of the Australian Council for Education
Research made this claim:

The purpose of medical research is to create and disseminate knowledge
and tools which can be used to improve human health . . . The purpose
of education research is to create and disseminate knowledge and tools
which can be used to improve learning.

The improvement of learning is the objective that drives (or should
drive) all education research.2

This statement of a criterion for good education research has a very
powerful appeal. Who would possibly disagree that in education we
should be trying, ultimately, to improve learning? It certainly has had an
immediate appeal to people in classes I have taught – and you can see how it would appeal to politicians and the public. It seems straightforward, practical, getting to the heart of the matter, proposing a criterion which lends itself to clear measures of whether progress is being made. Needless to say, it accords nicely with the currently fashionable ‘evidence-based’ view of what should drive education, that I will discuss further shortly.

My concern about this statement is not that ‘the improvement of learning’ is an improper criterion, but about the consequences of treating it as the sole or ultimate criterion. My own work in education has mainly focused on inequalities and social change and its implications, so I am concerned about the way it works to narrow education to the type of individual operation one might test adequately in a laboratory. Here are some examples of the type of things it leaves out:

- Schools do not just teach students things, they select and sort and produce differentiated outcomes. We can improve all students’ learning without changing the fact that in relative terms some will fail – and indeed some aspects of the current concerns about boys losing out relate to this: overall retention and achievement even for boys has improved over time – but in some respects their relative achievements and retention have declined (Yates 1997a; Arnot et al. 1999). The selecting and sorting are not just an accidental side issue of systems of schooling, or there would not be such fierce debates about forms of examination and university entry.

- In schools, students do not just learn the things that schools and teachers set out to teach them, and that are measured in their final exams – and this too affects their future. In my own recent longitudinal research project, we followed young people from different backgrounds at four different schools from the ages of 12 to 18, and we found many examples of different values and aspirations and self-assessments that they learned at those different schools (Yates and McLeod 2000; Yates 2001b; McLeod and Yates 2003). For example, in two ordinary high schools, with demographically comparable student populations and comparable participation and higher school certificate results, and a broadly similar formal curriculum, we found that one school tended to produce students who were quite practical and vocationally oriented, and who saw their future as their own responsibility. When we interviewed these young people the year after they had finished school they were very much involved in their post-school courses or jobs, and planning what they would do over the next few years. At the same time, this school produced a strong sense of ‘keeping up with the Joneses’ and a considerable lack of sympathy with bullying, racism and the unemployed. At the other school, the young people at the end of school were as likely to be drop-outs as not: they were still ‘finding themselves’, not strongly career active. But this
was also a school where the school leavers we interviewed expressed appreciation for what their teachers had done for them, in taking an interest in them as individuals and giving them second chances. It was a school that had developed in the students we interviewed a general sensitivity to and acceptance of difference, values where they would speak up about racism even after they left school. These two schools had had a short- and long-term effect on their students, but it was not a simple picture of the kind one could measure by the end of school statistics. This form of qualitative longitudinal study gives a rather different perspective on ‘school effectiveness’ than researchers who use that term or who work with a focus on ‘learning’ and medical models tend to think about. But it is researching effects that are a relevant part of assessing what schools are doing, both short and long term, to the people who go there.

• What counts as ‘learning’ (that is, what is to be learnt) is a debated, contested issue. The Director’s statement encourages us to treat the end-point as a given, and suggests we are mainly working with technical questions. A moment’s thought or reading of the newspapers would show that this is not the case. In recent times, while there are certainly outbursts of concern about the effectiveness of learning or the standards being reached in a particular subject and in a particular country (the International Education Achievement (IEA) studies are designed to produce just such a reaction), there has been equal attention to concerns about what young people should be learning today: How much do they need to study contemporary culture as well as older literature? Given the pace of technological change and change in the form of work, what are the learning foundations for entry to work in the future? What story of the nation, the globe, social values is to be approved for the compulsory years of schooling?

• In relation to technical, vocational and professional education, while some issues are rightly about ‘learning’ (how to promote competency, for example, or how to develop ‘new workers’ with the right dispositions to be flexible, self-disciplining, lifelong workers (Chappell et al. 2003); there are many equally pressing questions about the wide-ranging effects of different ways of providing education and training (the relationship between certain forms of certification and what pay and conditions can be claimed, for example).

The idea that good education research must be directed to ‘improvement of learning’, just like other attempts to insist that it must be directed to ‘effectiveness’ or ‘employability’ or ‘developing every child’s potential’, is an attempt to define and restrict what can count as education research, whereas the field of education itself comprises a broad arena of practices, institutions and problems. It is possible, and observably so, for govern-
ments or universities or education departments or funding systems to make a political decision that at a particular time and context only certain ways of addressing education problems will be eligible for funding, but this is not the same as establishing that, in principle, only certain problems are part of the agenda. The latter statement appears to be making a claim about the scope of the field of education research, but is in fact making a claim for particular agendas of research or kinds of research within the field relative to other kinds. The analogy to medical research is a common vehicle used in such arguments about education research and what it should be doing, and I will return to that shortly.

Claim 2: Good education research must make sense to/be usable by teachers (or instructors or parents or the lay reader)

A different starting point in many debates about education research begins not with what good research does, and what topic it should be directed to, but by talking about why much education research is not good. Some of the tropes of this discussion are widely shared: the problem with education research is that it is irrelevant, too academic, poor quality, jargon ridden. It is not producing new knowledge that speaks to teachers or instructors. It is not useful.

All of these are salutary comments. They direct our attention to what sort of a field this is: a field that is not simply characterized by some abstract search for knowledge but a field of practice where different players have their own sense of what is needed or desirable. These comments air the questions of how, in this particular arena, we are going to judge traditional research issues such as ‘significance of outcomes’, ‘contribution to knowledge’. These are good questions, but Part 2 will try to show that they are not at all straightforward, and are given quite different enacted meanings in different parts of the education arena.

Taking the ‘usability’ criterion first, most people are aware that the questions and knowledge relevant to a teacher or instructor interacting with their class are not identical with those of a school administrator or a system policy maker. To some extent then, the claim about how too little of the research seems to speak to practitioners is a complaint about the relative funding and attention given to large-scale concerns rather than small-scale ones, or to the interests of some practitioners in the field (policy makers and administrators) rather than others (classroom practitioners).

Alternatively, the claim here might be seen as speaking from one position within a range of paradigm differences and points at issue in the field of research itself: how much of what happens to participants in an education system is a result of large resource allocations, accountability systems and other processes and checks that can be instituted at system
level; how much is it explicable by social changes and movements and relationships beyond the field itself (processes perhaps not immediately accessible at a ‘common-sense’ level); how much is it revealed by laboratory-based experiments on the mechanics of cognitive processes; and how much are outcomes essentially produced by what happens in particular and complex relationships between teachers and learners? Some of these paradigms claim that ‘useful’ knowledge about teaching/learning process can begin with serious research on basic mechanisms that are later applied in ways that can be used by teachers (or used to manage teachers); others say that real insights are only found in research that works in real-life contexts with the collaboration of the teachers.

Now consider the criterion that good research must, at least, ‘make sense’ to a broad class of readers, that research that is jargon-ridden and ‘too academic’ is poor, and helps account for the poor reputation of research in this field. Later in this book, I illustrate two different dimensions of why this accusation might be less transparent than it seems (which is not to say that it is never warranted). Firstly, personal testimonies from practitioners about what they find meaningful and useful indicate considerable diversity in this (see Chapter 7). Some teachers choose to do doctorates on formal and jargon-laden academic themes, and later claim that these areas of intellectual interest have been relevant to their own practices, while other teachers have little time for research which takes the forms necessary if it is to be deemed respectably technically ‘valid’ as research. Some practitioners look for research to provide new ideas rather than ways to do it; others are critical of research not directly framed to the latter ends. And some ideas which once were the jargon of a few researchers can eventually enter the ‘common sense’ of a wide range of practitioners or even the broader domain of public discourse.

Equally importantly, university-based researchers do not simply choose idiosyncratically to conduct and write research in ways that practitioners dismiss as academic: it is frequently a requirement of how they must operate if they are to gain a doctorate, hold down a university job, win money to do research. The specific ways that ‘contribution to knowledge’ or ‘significance’ are judged in contexts discussed in Chapters 3, 4 and 5 (the thesis, publishing in refereed journals, winning competitive research grants) force choices that may well compromise the achievement of a researcher in relation to these same criteria as judged in the contexts discussed in Chapters 6 and 7 (commissioned research; schools, teachers and other practitioners). Some of the tensions here are likely to be exacerbated by the movement to ‘evidence-based’ policymaking and its corollary of favouring for funding only the research that meets ‘scientific’ standards of design, particularly large controlled trials and comparisons. This latter direction of course draws on another widely
Claim 3: Quality education research must be scientifically-based research

In the USA, a number of federal initiatives since the late 1990s have been enacting in legislation the requirement that research funding for education be ‘scientifically-based’: that only SBR (scientifically-based research) designs be eligible for federal funding, and only initiatives based on such research be eligible for the billions of dollars of federal aid. Subsequently, the Education Sciences Reform Act (2002) was passed, creating a new Institute of Education Sciences to replace the previously named Office of Education Research and Improvement in order to ‘advance the field of education research, making it more rigorous in support of evidence-based education’. It specifically aims to promote ‘more rigorous’ and more focused randomized trials, and other effectiveness studies, and to circulate the results of such research evaluation and review through mechanisms such as its ‘What Works Clearinghouse’.

In other countries too, there are some parallel moves to encourage a research organization and set of criteria that attempt rigorously to review the state of knowledge on a particular area, and fund only projects that meet standards of rigorous research design to take this further. In the UK, for example, the Department for Education and Skills (DfES) works in collaboration with the National Foundation for Education Research (NFER) and the EPPI Centre (the Evidence for Policy and Practice Information and Co-ordinating Centre of the Institute of Education, University of London) to produce a database on current education research in the UK (CERUK). The EPPI Centre’s publicity emphasizes its ‘systematic’ approach to reviews and research appraisal, in contrast to ‘traditional literature reviews’ and suchlike. Systematic reviews ‘use explicit methods to identify what can reliably be said on the basis of these studies’.

These moves are different from the longstanding practices of having a range of ‘experts’ in a field assess the quality of work, as discussed in Chapters 4 and 5. The particular developments here are moving to prescribe particular technical qualities that research must meet to be considered legitimate.

The case in favour of such a move is heard repeatedly across national contexts and different arenas. Like the ACER Director’s argument earlier, the arguments make lavish use of analogies with medical research (its immensely greater achievements compared with education; its much higher funding; its higher public reputation), and promote an approach...
that elevates controlled comparison as the basis for rigorous knowledge.

Here is one version of the case that indicates the ready appeal of this
movement to improve education research by making it operate more like
research that has been successful in other areas:

This process [of Scientifically Based Research] could create the kind of
progressive, systematic improvement over time that has characterized
successful parts of our economy and society throughout the 20th cen-
tury, in fields such as medicine, agriculture, transportation, and
technology. In each of these fields, processes of development, rigorous
evaluation, and dissemination have produced a pace of innovation and
improvement that is unprecedented in history. . . . Yet education has
failed to embrace this dynamic, and as a result, education moves from
fad to fad. Education practice does change over time, but the change
process more resembles the pendulum swings of taste characteristic of
art or fashion (think hemlines) rather than the progressive improve-
ments characteristic of science and technology. . . . If Rip Van Winkle
had been a physician, a farmer, or an engineer, he would be unem-
ployable if he awoke today. If he had been a good elementary school
teacher in the 19th century, he would probably be a good elementary
school teacher today.

(Slavin 2002: 16)

At the heart of a wide range of research is the attempt to build in some
systematic way on what has gone before, and there is an intuitive appeal
to current moves to try to pull together more systematically and on a
larger scale what research has so far established, especially given that
such moves have been accompanied by increased government funding
for education research. The current debates in the education research
community about the moves are also of interest because they make
explicit some of the implicit benchmarks that are widely held about good
research, particularly the appeal to medical breakthroughs, and attempt
to set down what it would mean to do similar quality research in edu-
cation. But therein lies the problem: what, precisely, is the ‘scientific’
characteristic of research of those fields that have made widely-recognized ‘research’ breakthroughs? What would it mean to do this in
education? And how appropriate is it to take such an approach as the
single benchmark for good research in education?

The first point to notice is that the move to prescribe certain forms of
‘scientific’ research as the benchmark for good research in education, is
not one that has been reached by a developing consensus within the
research community, but one that comes into being only because much
of the education research community has apparently come to different
conclusions, and has developed directions that are not sufficiently
science-like. In the UK, for example, the debate sparked by David
Hargreaves and James Tooley (Hargreaves 1996; Ball and Gewirtz 1997;
Tooley 1998, 2001) in the late 1990s was that the research community itself had lost its way, that ‘peer-reviewed’ journals were accepting research articles that do not meet good standards of research (as defined by Tooley).

In the USA, developments were driven by decision making in the political sphere. For example, the 1998 legislation to allocate a large sum of money for school reform on the condition that funds were allocated only to models ‘proven’ in terms of experimental-control comparisons on standards-based measures, was an initiative of two congressmen (Slavin 2002). The subsequent and much cited ‘No Child Left Behind’ Act (2001) mentioned ‘scientifically-based research’ 110 times, defining this as ‘rigorous, systematic and objective procedures to obtain valid knowledge’ which includes research ‘that is evaluated using experimental or quasi-experimental designs’, preferably with random assignment (Slavin 2002). But this particular operationalization of what it meant to do ‘scientific’ research preceded the commissioning of a report by scientific experts on what they thought doing ‘scientific’ research entailed (the report on Scientific Principles in Education Research produced by a subcommittee of the National Academies of Sciences, of Engineering and of Medicine) (Feuer 2002 et al.). And, despite the legislative intent to fund only scientifically-based initiatives (SBR), Slavin, an advocate of the new directions, notes that the bulk of the federal money on school reform initially at least has gone to approaches that do not in fact meet high standards for evidence-based approaches, partly because so much existing research does not fit the technical form required, and partly because ‘state officials who review CSR [Comprehensive School Reform] proposals still have broad discretion’ (Slavin 2002: 16) – that is, they may have a different way of judging what they think is quality research.

So what does good ‘scientifically-based’ research look like? In Acts like ‘No Child Left Behind’, and in bodies associated with the US Institute of Education Sciences and the UK EPPI Centre, it means research that has a particular methodological form: research that looks like an experiment or a quasi-experiment or clinical trial because it uses careful controlled comparisons. Only such research is deemed ‘valid’ or ‘proven’ and only such research is taken account of when panels are commissioned to review research in particular areas of education to identify what is ‘known’ and ‘not known’. Yet when scientific experts are asked to say what it means to be scientific, it appears that the issue is not nearly so clear cut.

According to the report by scientific experts from the National Academies in the USA, ‘scientific’ research describes a ‘culture of inquiry’. It involves certain norms or principles, and self-monitoring by the research community. It does not involve ‘an algorithm, checklist, or how-to guide’ (Feuer et al. 2002: 7). The report lists the norms or principles shared by ‘all sciences’ as follows:
• Pose significant questions that can be investigated empirically,
• Link research to relevant theory,
• Use methods that permit direct investigation of the questions,
• Provide a coherent and explicit chain of reasoning,
• Yield findings that replicate and generalize across studies, and
• Disclose research data and methods to enable and encourage professional scrutiny and critique.

And even here it qualifies its list by noting that it ‘is very unlikely that any one study would possess all of these qualities although a successful programme of research is likely to embody all of them’ (Feuer et al. 2002: 7).

There is nothing in this list to declare that only controlled comparison studies measure up to these principles, and it is at least arguable that all streams of education research that participate in the culture of peer-reviewed journals do work with these norms, depending on how one defines ‘replicate and generalize’. My argument, discussed further in Chapter 4, would be that even postmodern articles could be interpreted as meeting these standards in that they are accepted for publication only if they are seen as linking to and building on some existing lines of this developing body of theory and empirical analysis.

Nevertheless, scientifically-based research in education is normally interpreted as ‘random or matched controlled studies’, and the case where Feuer, Towne and Shavelson explain what these principles look like for education helps to show why this is the case. In particular, the argument constantly blurs and shifts between (a) a consideration of how scientists in other fields operate; (b) what people say about the reputation of education research; and (c) the problem of whether it is the education research ‘peer’ community itself, outside research experts such as the National Academy, politicians, or people in general who are relevant judges of good practice here. I now want to look a bit more closely at what types of appeal are being made, what source of legitimation the arguments refer to at different points.

Scientific culture vs technical standards for methodology

The National Research Council (NRC) report emphasizes that scientific method is not a single set of methodological techniques as are defined in the earlier legislation. For one thing, scientific method requires attention to and methods appropriate to the contextual specificity of the phenomenon being investigated: ‘No method is good, bad, scientific, or unscientific in itself. Rather, it is the appropriate application of method to a particular problem that enables judgments about scientific quality’ (Feuer et al. 2002: 8).
The specificity of education as a field is drawn on heavily by critics of the imposition of new standards for what ‘valid’ research is, and here the authors and advocates of the case for a ‘scientific’ approach in education appear to share at least some of the starting points of such critics. They say, for example, ‘many have argued – and we agree – that the complexities of education are not analogous to the physiology of disease and thus the expectation that any single intervention could adequately “cure” an education “ill” is misplaced’ (ibid. 2002: 12). But the authors bury this particular point in a footnote, and, having noted it, retreat immediately from their own opinion as experts, to the court of appeal of what politicians and the community think: ‘We include the reference here as an illustration of public perceptions and policy rhetoric; we do not intend to take on the underlying substantive issues’ (Feuer et al. 2002: 12, fn 14).

Nevertheless, among education researchers and in the practitioner community there is an ongoing debate about the implications of the specific nature of education for methodological approach (for example, Loughran 1999; Lagemann 2000; Rist 2000; Berliner 2002; Edwards 2002; Erickson and Gutierrez 2002). Among the reasons for the widespread shift in the late 1970s and beyond to qualitative and case-study based work was that this seemed to address better the complexity of the real-life classroom situation, and that it is ‘methodolatory’ to insist on laboratory-based or artificially simplified experimental programmes that may well generalize and build on themselves in other similar laboratory or other artificially simplified contexts, but that are not ‘generalizable’ in real-world settings. Others, such as Berliner (2002), point to a range of empirical evidence from past research and education reform that would question a faith in quasi-experimentalism and controlled comparative trials as a definitive answer:

- Findings do not stay stable over time (for example, the starting point assumptions of researchers about gender or race in the late twentieth century are different from those of ‘good research’ done earlier in the century);
- Attempts to replicate a programme of research have frequently produced inconsistent results;
- Within broad ‘scientific’ findings about education there are frequently many individual classrooms or schools that show directions of effect in confounding directions, that is, context and specificity appear to be of as much interest as the ‘general’ pattern.

There are practical and ethical problems too as to whether or how far education research can replicate controlled methods, even if it wanted to. Some of the problems are described by one of the enthusiasts for these new standards, recounting his own difficulties in setting up research that is appropriately controlled. In one case he found, not surprisingly, that
schools were not prepared to participate in a study where they might be randomly assigned to experimental or control groups – even though they were being offered $30,000 to participate. The researchers had to increase these incentives enormously before schools would participate, meaning that the study became so expensive that it was very unlikely ever to be replicated (Slavin 2002: 18).

Another point emphasized in the NRC report (a report produced by Academies with experts from science, engineering and medicine) was that the scientific work is not guaranteed by the techniques used but by the peer scrutiny that takes place by the members of the research community in that field, and that this is what makes possible ‘disciplined, creative and open-minded thinking’ and produces researchers who ‘can engage differing perspectives and explanations in their work and consider alternative paradigms’. However, in the case of education, this soon produces a circular problem for these experts: taking peer scrutiny as the court of judgement appears to be acceptable only if the community of peers is already acceptable. Once again, the scientific experts retreat and appeal to the fact that ‘education research is perceived to be of low quality’, not by themselves (that is, they decline to make a pronouncement on this matter), but by ‘lawmakers’, community and some education researchers. This perception is used to justify certain standards of method or structure being imposed as the criteria of good quality.

**Analogies and courts of appeal**

Medical research is repeatedly invoked in arguments about education research, usually in ways reminiscent of the old song, sung by Rex Harrison as Professor Henry Higgins in *My Fair Lady*, ‘Why can’t a woman be more like a man?’ Why can’t education research have the prestige, the obvious history of progress of medical research? Why can’t it attract the same money and produce an aura of awe and expertise? Why can’t it produce the same outcomes and breakthroughs? Why can’t it look the same?

One approach in these debates has been to keep exploring the analogy until you find the illustration that will support the direction you wish to argue for in education. On the one hand, Robert Slavin, quoted earlier, argues that the take-up of scientific methods in medical research produced ‘a pace of innovation and improvement that is unprecedented in history’ and tends a Rip Van Winkle contrast to show how little teachers have been advanced compared with doctors (Slavin 2002: 16). In riposte, Frederick Erickson and Kris Gutierrez counter with the example of thalidomide:

> We are concerned that premature conclusions about ‘what works’ in the short term without careful consideration of side effects that may appear
downstream, can provide false warrants for the education equivalent of thalidomide. That was a medical treatment that was shown scientifically (i.e. by means of randomized trials) to have clear positive effects. [...] What a tragic irony; thalidomide prevented morning sickness very effectively but it was also effective in causing deformities in the fetus growing in the mother’s womb. The latter effects were only discovered after the babies were born, and it took years to trace the cause of the deformities back to the mothers’ use of thalidomide. Will our current desperate attempts to discover ‘what works’ to raise standardized test scores in the short run have analogous effects on our children and teachers in school...?

(Erickson and Gutierrez 2002: 23)

Some use medical research to justify only experimental laboratory work or randomized clinical trials; others point to the way medicine itself has been impacted on by movements that bring in more complicated criteria, theories and forms of research to take account of doctor–patient relationships in new ways: for example, from the women’s movement, from consumer rights, from cross-disciplinary developments in psychology, from studies of minorities and power.

What is interesting is why medical research is such a ubiquitous point of reference. The fact that breakthroughs have been made is undeniable, but, as the NCR expert report makes clear, this is only part of the story, even in medicine. Most of the arguments explicitly use medicine to associate arguments about method and design quality with arguments about status and money. But if the phenomenon is not the same, there is no guarantee that replicating the methods will have such an effect – either that it will produce similar outcomes or that, even if it did, it would acquire the same prestige or attract the same funding. Every aspect of education (training, salaries, conditions of practitioners and so on) is funded differently from medicine – not just research. And funding is committed or not committed (both by governments and by individuals) for many reasons in addition to a rational calculation that it will produce particular outcome effects. To give a simple example of this, in the mid-1990s state government elections in Australia were dominated by politicians promising that they could spend less on schools than they were currently doing. State premiers scrutinized each other’s education funding and it was a matter of shame and a task for reform to be spending the most per head. Efficiency and small government rather than education spending were the dominant discourses. After some period of this and some unexpected losses by parties that had most successfully run with this platform, the discourse has changed. In more recent elections, politicians compete to show how much they care about education, including how much they are prepared to spend on it. In an election campaign in 2003, one party began with material promising that certain
of its proposed education reforms would be ‘cost neutral’ but removed this part of the boast from its election material as the campaign progressed. Presumably, the party’s polling had told them that this boast about restrained spending on education was not helping them seem attractive to the voters.

It is also interesting how the ‘poor reputation’ of education research is frequently used as a court of appeal that establishes its low quality, rather than as something that needs to be investigated and unpacked. The issue of what produces ‘reputation’ is a far more complex matter than most of the debates assume, and an issue that would require some attention to structures, communication, multiple demands and how fields of research are or are not made visible.

*Like hemlines?*

We have seen that currently, in a number of countries, governments, research offices and organizations, and at least some researchers are getting excited about the potential for education research to be made more rigorous and systematic and to produce the results and transformations that have been apparent in other fields. In the passages above, I have focused particularly on the case being made by the advocates of such an approach, with some briefer references to other parts of this debate. Here is a slightly more extended case by two researchers, David Hamilton and Malcolm Parlett, against the controlled evaluations and trials that are increasingly favoured:

The most common form of agricultural–botany type evaluation is presented as an assessment of the effectiveness of an innovation by examining whether or not it has reached required standards on pre-specified criteria. Students – rather like plant crops – are given pre-tests (the seedlings are weighed or measured) and then submitted to different experiences (treatment conditions). Subsequently, after a period of time, their attainment (growth or yield) is measured to indicate the relative efficiency of the methods (fertilizers) used. Studies of this kind are designed to yield data of one particular type, i.e. ‘objective’ numerical data that permit statistical analyses. Isolated variables like IQ, social class, test scores, personality profiles and attitude ratings are codified and processed to indicate the efficiency of new curricula, media or methods.

Recently, however, there has been increasing resistance to evaluations of this type. The more notable shortcomings may be summarized as follows:

1. Education situations are characterised by numerous relevant parameters. Within the terms of the agricultural–botany paradigm these must be randomised using very large samples; or otherwise strictly controlled. The former approach entails a major data-collection
exercise and is expensive in time and resources. … The latter procedure – of strict control – is rarely followed. … is dubious ethically, but also leads to gross administrative and personal inconvenience. [And even if it was used] rarely can ‘tidy’ results be generalised to an ‘untidy’ reality. Whichever approach is used, there is a tendency for the investigator to think in terms of ‘parameters’ and ‘factors’ rather than ‘individuals’ and ‘institutions’. Again, this divorces the study from the real world.

2 Before-and-after research designs assume that innovatory programmes undergo little or no change during the period of study. This built-in premise is rarely upheld in practice.

3 … the concentration on seeking quantitative information by objective means can lead to neglect of other data, perhaps more salient to the innovation, but which are disregarded as ‘subjective’, ‘anecdotal’ or ‘impressionistic’. However, the evaluator is likely to be forced to utilise information of this sort if he is satisfactorily to explain his findings, weight their importance and place them in context.

4 … tends to be insensitive to local perturbations and unusual effects.

5 Finally, this type of evaluation often fails to articulate with the varied concerns and questions of participants, sponsors and other interested parties. … diverts attention away from questions of education practice towards more centralised bureaucratic concerns.

(Hamilton et al. 1977: 7–9)

One reason I have chosen to report this particular argument by Hamilton and Parlett at some length is that although published in 1977, it can be read as a rejoinder to the case for concentrating on large controlled trials that are now being treated as such an exciting new discovery. Some of the current debates imply that the main reason education research has not been making progress is that it has not been bright enough to think about trying to be systematic and to do big studies. The sub-text is that the field had not been able to get its act together, has not been able to recognize what might be possible – that no one had previously come up with the idea that concentrating on big scientific studies would be a good thing. But reading the history of debates about research suggests something other than either ignorance or simple ‘fads’ is at work.

I have reported these arguments about scientifically-based research and the counter-arguments in favour of qualitative and anthropological styles of research not because I think the latter disposes of the former (or vice versa), or that there is no place for large controlled studies, but because I think attention to the different arguments and location of arguments helps to show what debates about education research look like. Education research might look like ‘fads’ or ‘changing fashion in hemlines’ to use Slavin’s words, but there is some logic in why different types of arguments keep recurring rather than one disposing of the other once and for all.
One obvious issue is that there are different end-point users for the research, and what some want to know differs from what others want to know – with implications for methodology and also for what a research design and a research publication should look like. Another is that the phenomenon itself has different conditions at different times. Parlett and Hamilton were writing to contexts where decentralized curriculum was common and school-based innovation was favoured. Slavin and others are writing in a context where governments are attempting to bring more centralized control and steering of what happens in the research community generally, not just in education. It is not the same as simply finding the mechanism that causes a disease.

Definitions and enactments

Many of those who participate in debates about education research, even where they disagree strongly about the importance of different types of work, allow that there is at least some place for large data-gathering studies that attempt to do some controlled comparison, and some place for other kinds of work – interpretive, philosophical, case study (though there is less agreement about this in the case of post-structuralist work and in arguments by post-structural researchers). But the debates about the moves to tie federal government funding to one particular form of work, even if they formally acknowledge that there is room for a range of other styles of research, are heated because funding rules can begin to define in practice what is able to be done as research, and to define what is counted as legitimate well beyond the context in which it is initially enacted.

Even in the account of the National Academies, this difficulty is apparent. Feuer et al. are clear that scientific education research is not to be read as being the only legitimate form of education research: ‘we do not intend to minimize the significance of humanistic, historic, philosophical and other non-scientific forms of study in education’ (Feuer et al. 2002: 5). But there is an ongoing slippage in how both the specialists and the broader community react if important bodies decide that only certain methodologies will be considered ‘scientific’ or ‘valid’.

Context and timeliness as issues in education debates

Examining recent debates about the quality of education research draws us into issues about the scope of the field, the questions and audiences (or courts of appeal) that are considered important, and about what analogies and types of arguments carry weight for whatever reason. They
draw attention to issues of context in what is being said and done, context both in the sense of historical context – what matters at this time – and in the sense of the particular and broader setting of the discussion. Here, using some further examples from my own involvements in research, I want to draw together a number of the themes that have been flagged in the preceding discussion as part of the ‘thing’ that makes up education research.

1 The phenomenon under study is not static, whether it is about students, teachers, schools, vocational training, higher education. One of the reasons Julie McLeod and I began the 12 to 18 Project was to investigate how teenagers see themselves, schooling and their future today. These young people have grown up with different cultural norms from their parents; different kinds of parents, technological artefacts, kinds of experience, school curriculum policies, even different words in the language. Research done on their parents’ generation when that generation was at school does not necessarily hold today in terms of who girls or boys are, what motivates them, what would best engage their interests in learning, or what trajectory they are following or need to follow to end up with a good job or life.

Similarly consider the old single-sex versus co-education debate. It is highly debatable whether ‘single-sex’ or ‘co-education’ is the same phenomenon where it exists now as it was in the 1950s. We can choose to see co-education as an abstracted ‘factor’, where research done several decades ago can give us a foundation for what to do today. But we can choose to see it as something that is, at least potentially, done differently and experienced differently according to what purposes teachers have in mind, how they actually set things up, what the girls or boys and their parents believe about it, even what the media says about whether or not it is a ‘good thing’ – and that might need some new research rather than taking old research on trust.

2 The range of questions thought to be worth answering are not static, though some questions do endure. At the beginnings of national state education systems, the major issue was often how to get good attendance in schools, and what types of pressure to put on families to achieve this. When nations become concerned about their international competitiveness, the emphasis turns to achievement comparisons of various kinds. At some times, and in some countries, there is a particular interest in identifying different types of aptitude and providing appropriately for different types of students; at other times and in other countries, levels of basic achievement, or mass outcomes, are the focus of concern. In vocational education, the concern has sometimes been with how to train effectively for existing jobs; currently, there is a concern about how to educate in a way that will allow people to go on to do other types of jobs
in the future. Sometimes the key interest has been how to operate a
system in a way that costs less. But some questions, of course, are more
enduring ones. What is an appropriate way to develop mathematical or
scientific or literacy knowledge and skills? What is the relationship
between particular ways of teaching and particular learning outcomes?

3 The material support to do or publish research changes, and along with it
there are different opportunities and pressures in relation to doing particular
types of research. In Australia, in the 1970s, the government made
available money for school-based partnership research, both through
Innovations funding, and through its Disadvantaged Schools Program.
This effectively encouraged styles of research that were local, action-
oriented and collaborative. Now major funding primarily comes through
national competitive grants programmes or through consultancies to
state and commonwealth authorities or national associations in different
fields of education. The former encourages research that is seen to be big:
to have national and preferably international impact. The latter
courages research that is policy oriented.

This changing funding and political context affects not just method-
ology and the types of research that seem desirable, but the substantive
agendas of that research and its presentation. In the 1970s, in both
Australia and the UK, attention to ‘innovation’ was encouraged, and
attention to disadvantage or inequality was also encouraged. In the 1980s
and 1990s, disadvantage went very much out of favour, and politicians
were enthralled by ‘effectiveness’. If you wanted to get your research
funded and listened to, it needed to be couched more in terms of winners
and effectiveness, and less in terms of elucidating the problems that losers
were facing. One side effect of this was that sociologists, in danger of
being removed from education faculties because they were seen as cri-
tical and not contributing to good teaching practice, began to rebadge
themselves as ‘policy researchers’. Similar changes affect publishing and
readership. In Australia, in the 1980s, books on girls and gender were
selling well; now they are not.

4 Changing social agendas and knowledge also directly affect the research
field and what counts as adequate or good research in terms of a thesis, or
a journal article, or a book. When I did my Masters degree in Bristol in
the mid-1970s, Paul Willis came and gave a talk on the study he was
doing in Birmingham, which later became the book Learning to Labour.
The advertised title of his talk was something like ‘An education eth-

nography of youth, and the reproduction of class’, and what I remember
from the discussion was Miriam David giving Paul Willis a hard time for
advertising his talk as being about youth (or young people), when in fact
his research was solely about boys. (In the end of course, Willis did take
account of this and his book is now remembered as much for its con-
tribution to understanding working-class masculinity as to class reproduction, the tradition where he began.)

Up to about the mid-1970s, it was quite legitimate for researchers to do research on all-male samples and to write about it as if it included everyone (Kohlberg, on moral thinking, was another well-known example, which spawned Carol Gilligan’s body of work to ascertain if the hierarchy looked the same if you took girls and women as your starting point). And the same thing applied to a lot of quantitative research on all sorts of topics published before the 1970s. Another way of putting this is that in that earlier period, across a whole range of different methodologies (quantitative and qualitative), research could be judged as good research without having any fine attention or sensitivity to gender – it was not picked out as an essential factor for investigation. Today, however, if you propose an education study and do not include gender as something you will mark or note, there is a reasonable chance your research will be seen as inadequate. Note that this does not mean that you have to be compellingly interested in gender, or to find that gender is an important differentiating feature in a particular area of investigation, but you are expected to be alert to the possibility that it may be, and to take account of this in your methodological design. This is not about whether you, yourself are particularly interested in gender issues: it is a sign that debates about this have become socially prominent in relation to debates about young people, and about what is happening in education. Whether you are dealing with quantitative research and factor analysis, or interpretive, case-study work, an awareness of gender as a possible aspect differentiating or influencing the phenomenon you are studying is expected. (However, empirically, one would have to say there seems to be no similar imperative to be required to take account of the theories and previous research that people who have worked on gender issues have produced!)

What good research looks like – reflecting on the debates

Looking at these debates, and taking education research as an arena, I hope it will be apparent why I think we are not dealing with something as simple and straightforward as a cure for cancer. It may indeed help to make progress if we devote greater resources to research, call for larger studies that attempt to pull together expertise and build on it, do more ‘rigorous’ research, replicate what seems to help progress in finding the building blocks for a cure for cancer. But in education we also have a field whose end-point parameters and questions are much more diverse, whose setting and agendas change. It is a field where an individual student (or teacher or administrator) is located in a bigger setting where
the culture and the things they are exposed to change; where parents’ and teachers’ thinking about what they are trying to achieve changes; where policies are introduced and change each time there is an election, often for reasons relatively unrelated to technical research results on ‘effectiveness’ and at least as related to how modes of doing education fit or conflict with the prevailing political philosophy of the day. Different participants have different demands on what types of research knowledge about education they need, and how this should be delivered. Researchers are engaged in debates both about what is possible and about what is desirable and whose interests should be served. Researchers are also positioned within decisions, processes, structures created by others, decisions that directly and indirectly define good research and the consequences for researchers of taking certain research paths rather than others.