Evidence, Criteria, Policy and Politics: the debate about quality and utility in educational and social research

Chapter 34, pp. 766- 795, in Denzin N. & Lincoln Y. (Eds 2018) *The SAGE Handbook of Qualitative Research* 5th edition, Sage

Harry Torrance, Manchester Metropolitan University, UK
h.torrance@mmu.ac.uk

Introduction

A good deal of the day-to-day business of educational and social research, the "practice architecture" of a discipline or a field of endeavour, as Kemmis (2012) calls it, revolves around debates about theory, methodology, and ethics. How should a particular social issue be theorised? How should an empirical investigation be framed? How should the fieldwork be designed and conducted? These are the intellectual questions which define a field of inquiry. But a field of inquiry is enacted and operationalized in material and institutional settings. Equally important questions, indeed perhaps more so at the present time, are who has the right to decide these matters? In whose interests is social research undertaken? Who decides what counts as 'evidence'? And how is independent critical research, particularly critical qualitative research, to be funded and sustained in the face of government insistence on direct relevance to, and immediate impact on, social policy and intervention programs? Thus what we might term the 'political economy' of social research – what gets funded and why – is as important to the overall delineation and development of social research as the particular theoretical and methodological debates that frame a disciplinary field at any particular time. Intellectual questions and material contexts interact to produce the practice of social research at any particular historical moment.

So it is at present. Educational and social research in general, qualitative approaches to educational research in particular, have been under sustained criticism for fifteen years and more, particularly from government but also from many researchers themselves who see an opportunity to advance their particular vision of social science. The argument is that qualitative approaches to educational and social research have not provided a sufficiently cumulative and robust evidence base for the

development of policy and practice, and in particular have not produced sufficient experimental data to allow policy makers to evaluate policy alternatives. Educational research in particular is criticised for being too much of a 'cottage industry', producing too many small-scale, disconnected, non-cumulative studies that do not provide convincing explanations of educational phenomena or how best to develop teaching and learning.

Responses have included significant philosophical arguments rebutting the claim that straight-forward empirical evidence can be produced to identify cause and effect in social research; detailed arguments about the complex interaction of research and policy, again rejecting any straight-forward linkage; and the development of a range of arguments for and examples of mixed method approaches to social research. Nevertheless the calls for further development and expansion of experimental methods in social research continue. They are linked to arguments about the need to concentrate research resources (funding) on fewer centres of excellence undertaking much larger scale investigations and interventions, and to arguments about the need to focus on some (policy relevant) issues and topics at the expense of others. The debate is not just about method, it is also about the scale and focus of inquiry. This chapter will review key elements of these debates, reflect on the implications for the field of qualitative research and its relationship to policy, and look to delineate the ground on which critical qualitative research can continue to be undertaken.

The call for 'better' evidence

Calls for the development of evidence-based policy and practice in education and social research raise many questions about the nature of evidence and the relationship between research, policy, practice, and the democratic process. Such calls seem self-evidently reasonable – who would argue against the use of evidence? Who could argue in favour of superstition-based practice? Yet the production and use of evidence is not straightforward and policy-makers can cite 'the evidence' when it suits them, and ignore it, invoking other political exigencies, when it does not. As Winston Churchill once famously remarked when discussing the relationship between science and democracy "Scientists should be on tap, not on top" (1949, quoted in Leach 2013, p. 298).

Evidence-based practice also appeals to ideas of rational planning and the self-interest of some sections of the research community, privileging research knowledge above what is often characterised as traditional, ineffective professional practice. With respect to education, such calls ask 'where is the secure research-derived knowledge-base of teachers and teaching?' - simultaneously castigating the teaching profession for not having one, and the research community for not providing it (Hargreaves 1996; Slavin 2002). Comparisons are often made with medicine, which, it is claimed, does indeed have a well-proven knowledge base from which to select and apply treatments. Though within medicine itself, the debate is rather more complex and nuanced, especially when it comes to the interaction of treatments and patient decision-making – not all patients take their medicine (Barbour and Barbour 2003).

The debate about how educational research and, more generally, social research, might better serve policy is not a new debate, and has been revisited many times since the inception of educational and social research as established university-based activities (e.g., Lagemann, 2000; Nisbet & Broadfoot; 1980; Weiss, 1972, 1980). However, it has been addressed with new vigor since the late 1990s as successive governments in the USA, the UK, and elsewhere have looked for better value for money from research, and more particularly looked to research for legitimating and supportive endorsements of their policies. The debate carries particular import for those working in the broad field of qualitative inquiry since it has tended to privilege so-called 'scientific' approaches to educational and social research, by which is meant empirical investigations of educational activities and innovations, oriented to the identification of causality, explanation, and generalization (e.g. National Research Council, 2002). 'Scientific' research has been extensively defined in US legislation, and includes reference to "measurements or observational methods that provide reliable and valid data across evaluators and observers… evaluated using experimental or quasi experimental designs…with a preference for random assignment experiments…" (NCLB, 2002, 115 STAT. 1965). The argument of critics is that qualitative research is not scientific, or not scientific enough, and cannot produce definitive evidence about 'what works' in social policy interventions.

Implicitly, therefore, and sometimes quite explicitly, qualitative approaches to research are marginalized. The debate seems to reflect both long-term changes in what we might call the 'terms of trade' between science and policy, along with more specific short-term jockeying for position amongst particular researchers and government officials/advisers at a particular point in time.

The intensity and focus of the current debate in the UK can be dated from a speech in 1996 by David Hargreaves (then Professor of Education at Cambridge University) to the Teacher Training Agency (TTA—a government agency regulating teacher training). Hargreaves (1996) attacked the quality and utility of educational research, arguing that such research should produce an "agreed knowledge base for teachers" (p. 2) that "demonstrates conclusively that if teachers change their practice from X to Y there will a significant and enduring improvement in teaching and learning" (p. 5). Subsequent government-sponsored reviews and reports took their lead from this speech and produced what might be termed a mainstream policy consensus that the quality of educational research was low, particularly because so many studies were conducted on a small scale and employed qualitative methods, and therefore "something had to be done" (Hillage, Pearson, Anderson, & Tamkin, 1998; Tooley & Darby, 1998, Woodhead, 1998). That such claims were disputed need not detain us here (but see for example Hammersley, 1997, 2005; MacLure, 2003). It is worth noting, however, that subsequent analyses of papers published by the *British Educational Research Journal,* the leading UK journal of the British Educational Research Association, and of educational research projects funded by the UK Economic and Social Research Council (ESRC), demonstrated that critics had misrepresented the field and that in fact a wide range of methods were and are employed in British educational research, including-large scale quantitative analysis, experimental design, and mixed methods (Gorard & Taylor, 2004; Torrance, 2008).

The parallel intervention to Hargreaves in the USA is probably the National Research Council Report (2002) "Scientific Research in Education," though this in turn was produced in response to already extant policy debate and legislation identifying what would be defined as "research" for purposes of federal funding—specifically the Reading Excellence Act, 1999, and the No Child Left Behind Act,

2001 (see Baez & Boyles, 2009, pp. 5 ff., for illustration and discussion of these acts). A huge literature has been prompted by this legislation, subsequent attempts to delineate the boundaries of "scientific research in education" and responses to those attempts. For example extensive reviews were published in *Educational Researcher,* 2002, vol. 31, no. 8; *Qualitative Inquiry,* 2004, vol. 10, no. 1; *Teachers College Record,* 2005, vol. 107, no. 1. More recently the debate has continued with responses seeking:

to re-state some of the philosophical issues in identifying cause and effect and establish actionable knowledge in human affairs (Maxwell 2012, Morgan 2014);

understand some of the pressures that policymakers are under when trying to collect and evaluate evidence (Donmoyer 2012);

make visible some of the problems of publishing qualitative research in these new times (Ceglowski, Bacigalupa and Peck 2011);

link Randomised Control Trials (RCT) into more mixed method research designs (Christ 2014, Hesse-Biber 2012);

note the parallels between the evidence-based policy movement and colonialism (Shahjahan 2011);

and reassert the distinctiveness and social justice aspirations of qualitative research (Denzin 2010).

It is not my intention to review all of this work here, though I will return to the debate about mixed methods later in the paper. However, one quotation from the debate is worth highlighting, since in many respects it summarizes the "scientific" case, particularly the case for using not just a broadly quantitative empirical approach, but a specifically experimental design. Thus, Robert Slavin (2002), a leading proponent of the scientific method in the USA and inaugural Director of the Institute for Effective Education at the University of York, UK, argues that "the experiment is the design of choice for studies that seek to make causal conclusions, and particularly for evaluations of educational innovations" (p. 18.). And, in a turn of phrase that is directly reminiscent of Hargreaves's (1996) speech, Slavin suggests that policy makers want to know "if we implement Program X instead of Program Y, or instead of our current program, what will be the likely outcomes for children?" (p. 18).

Thus we would appear to have two research communities talking past each other with little constructive engagement over legitimate criticisms or potentially common concerns. Proponents of experimental design are clearly in the ascendancy however and, as such, perhaps do not feel the need to engage. This certainly seems to be the case in the UK where the debate has been given renewed vigour by the publication of two recent government policy papers advocating far more use of RCTs in UK social policy and educational research. The papers were commissioned by the Cabinet Office Behavioural Insights Team (Haynes et. al. 2012) and the Department for Education (Goldacre 2013). In many respects Goldacre's paper on Education might be seen as 'Hargreaves: 2.0'. In his own words, Goldacre's paper is a "call to arms" (p.16). He sets up a rhetorical binary between educational research(ers) and proponents of randomised controlled trials. Educational researchers are positioned as ignorant (of RCTs), incompetent (in research design and methods), and uninterested in improving teaching and learning in the classroom; proponents of randomized controlled trials are positioned as knowledgeable, skilled, only looking to identify what's in the best interests of children. Large parts of the paper draw on examples from medicine (again), and are completely (wilfully?) ignorant of the debates going on in the UK and USA. The paper ends, as many of these sorts of interventions tend to do, with a disciplinary 'land grab' for resources. Goldacre (2013) concludes "We need academics with quantitative research skills from outside academic education departments – economists, demographers, and more, to come in and share their skills…" (p. 18). Ah, the economists, thank goodness for the economists, whose models and analysis have so helped us to produce such an effective banking system, and to develop our economies over the last few years. Their RCTs have really helped with that. Leaving aside the rhetoric however, the Cabinet Office and Goldacre papers have reignited the debate about RCTs in the UK policy context and influential research funders are now routinely looking for such designs in research proposals (e.g. the Education Endowment Fund: http://educationendowmentfoundation.org.uk/apply-for-funding/).

However, it is also important to recognize that these criticisms are not restricted to the USA and/or UK policy contexts, nor indeed are they restricted to educational research. Reviews of and attacks on the quality of educational research,

and particularly the quality of qualitative educational research, have impacted debate in Australia and New Zealand (Cheek, 2007; Middleton, 2009; Yates, 2004), and are emerging in Europe (Besley, 2009; Bridges, 2005, 2009; Brown, 2003, Depaepe 2002; Frederiksen and Beck 2010). Similar debates about issues of relevance, utility and their relationship to definitions of research quality have been noted in other disciplines, for example Business Studies (Caswill and Wensley 2007), Social Policy (Brown 2010), and Anthropology (Mills and Ratcliffe 2012). Critiques have also been leveled against social research more generally. In a speech to the UK Economic and Social Research Council in 2000, titled "Influence or Irrelevance" the then Secretary of State for Education, David Blunkett (2000), asserted that

> Many feel that too much social science research is inward-looking, too piecemeal, rather than helping to build knowledge in a cumulative way, and fails to focus on the key issues of concern to policy-makers, practitioners and the public, especially parents.

More recently Schorr and Farrow (2011) note that the use of RCTs in social policy research is now advocated by the US Office of Management and Budgeting and they review the general trend across both government and philanthropic organizations to focus on a narrow range of experimental evidence when designing, implementing and evaluating a broad range of social intervention programs.

Thus, when the previous edition of this Handbook was published in 2011, the legislative concern to promote "scientific research in education" and particularly the place of RCTs in educational research was a fairly specific American phenomenon. Now it would appear to be far more widespread, both geographically and in terms of its reach across the full range of social policy research and research funders. Educational research, qualitative approaches to educational research, but also qualitative approaches to social research more generally have all come in for criticism and, taken together, suggest that qualitative inquiry is facing a global movement to reassert broadly empiricist and technicist approaches to the generation and accumulation of social scientific

"evidence" for policy making. The focus, worldwide, is on both methods and topics, seeking evidence to inform policy making, particularly evidence about "what works." Elements of such a movement will differ in their origins, orientations, and specific national aspirations. But equally they do seem to represent a concerted attempt to impose (or perhaps reimpose) scientific certainty and a form of center-periphery, research, development, dissemination (RDD) system management on an increasingly complex and uncertain social world.

Long-Term Trends: Whither/Wither Science and Government?

Part of the backcloth to the current debate is the uncertain status and legitimacy of both science and government at the present time. The role, purpose, and utility of science and scientific research is less agreed upon and less secure than it once was, and with respect to this, just as educational research can be seen to be situated in a wider debate about social research, so social research can be seen to be located in a wider debate about scientific research and the role of science in society. In the UK, for most of the 20th century, the relationship between science and government was determined by the so-called "Haldane principle" (after Viscount Haldane, an influential liberal politician who chaired the committee that articulated the principle in 1918). This settlement essentially resolved that university-based science would be funded from the public purse to pursue fundamental research, which would in turn produce unpredictable, but nevertheless substantial, long-term scientific and technical benefit—i.e., "basic" research would, over time, produce the platform for more "applied" technological developments and benefits. The central tenet of this position is that the quality of basic research is grounded in the independent and disinterested pursuit of knowledge. This was even characterized as the creation and operation of the "independent republic of science" by Michael Polyani (1962, cited in Boden, Cox, Nedeva, & Barker, 2004). The Haldane principle has a direct parallel in the United States with the publication of Vannevar Bush's "Science: The Endless Frontier" (1945). This argued, on the back of scientific successes apparent in the Second World War, for the federal government to significantly expand support for

scientific research on the basis of a similarly "arms length" linear model of "basic" research eventually leading to technological benefit. This led to the setting up of the National Science Foundation (NSF) in 1950 (see Greenberg, 2001).

More recently, however, government calls for much more short-term responsiveness and utility have pervaded policy debates and aspirations on both sides of the Atlantic and elsewhere—e.g., the Clinton focus on science and technology policy in the 1990s (Greenberg, 2001), and the current UK government concern to identify and evaluate the "impact" of research through its new Research Excellence Framework (DBIS, 2009; Higher Education Funding Council for England [HEFCE], 2009, 2014), the successor to the Research Assessment Exercise (Torrance, 2006). These calls for research to produce social and economic impact are in turn located in debates about the role of research in promoting economic growth, particularly in the context of globalization and constrained government budgets after the 2008 banking crash and global economic recession. Research is now expected to serve the development of each nation state's 'knowledge economy'. To this end, selectivity and concentration of research resources are particularly being pursued in the UK. A recent Government White Paper (i.e. a policy statement framing legislation) stated quite clearly that "we intend to maximize the impact of our research base on economic growth" (p. iv). It went on: "To compete effectively the UK must harness its strengths in…research…and its expertise in areas such as design and behavioural science…" (DBIS, 2011, p. 6). In essence the White Paper argues that investment in research should be oriented to those areas that promise most economic return, with "behavioral science" being deployed to understand and change people's behavior in relation to key threats to economic development such as poor health and global security. Moreover the White Paper goes on to assert that major social and economic challenges "can only be resolved through interdisciplinary collaboration" (p. 20) and thus government will "actively support strong collaborations" (p. 8) across disciplines and institutions. In turn, the UK's main social science funding body, the Economic and Social Research Council (ESRC) in its most recent 'Strategic Plan' includes a section on "Facilitating Partnerships and Realising Impact" which states:

> We act as a broker to ensure that the academic community is fully aware of the ways that social science can help meet the needs of government, business and the third sector (ESRC 2015, p. 14)

Research, including social research, will be marshaled and directed in the national economic interest.

Similar debates about purpose and level of government expenditure on social research can be observed in the USA in relation to calls to restrict NSF expenditure on the Social, Behavioural and Economic Sciences (SBE; Coburn 2011; Cantor and Smith 2013). Such calls have not gone unchallenged (Lempert 2013, Wilson 2013), though these responses tend to adopt the UK government's position by arguing that restricting Federal funding to science and technology is short-sighted - the social sciences are central to understanding the interaction of humans with natural and technical systems, and thus are needed in order to understand issues of climate change, implementing technological change, and so forth. Such arguments accept the basic premise that science should serve policy in a much more direct way than hitherto. Thus science in general, and social science in particular, is now expected to serve government policy and economic development very directly. This clearly begs questions about how to define quality and utility.

Equally, however, government itself is under pressure to "deliver," especially in areas of public policy. Since the first oil crisis of the 1970s put severe pressure on public spending, especially in the UK, and with the development and implementation of monetarist critiques of government spending in the 1980s, and the collapse of the Soviet Communist Bloc in 1989, there has developed a severe crisis of confidence and legitimation with respect to the role of government itself, especially with regard to the provision of public services: Are they really needed? If so, could they be better and more efficiently provided by other mechanisms and stakeholders? What reasons are there for state intervention in the lives of ordinary citizens? Such questions underpin the development of neo-liberalism (Lemke 2012, Peters and Olson 2009) and can certainly be seen in the animus of critiques such as Coburn (2011). In this respect, government demand for "evidence" is as much a demand for material to justify its

own existence, as it is a demand for the evaluation of particular policy alternatives. What is at stake is the legitimacy and efficacy of policy intervention *per se.*

Experimentalism: Part of the Solution or Part of the Problem?

Advocates of experimental design have inserted themselves into this uncertain nexus. Given such uncertainty, it is understandable that governments and policy makers will look to research for assistance. Research, or more generally, 'science', is still largely regarded as independent of government and thus able, at least in principle, to provide disinterested evidence for both the development and evaluation of policy, despite recent moves toward the development of a closer and more utilitarian relationship. The attraction of the sort of evidence that Hargreaves (1996), Goldacre (2013) and Slavin (2002) claim can and should be provided is easy to appreciate. It sounds seductively simple. When charged with dispensing large amounts of public money for implementing programs and supporting research, one can understand that policy makers might value this sort of help—at least as long as the answers to the questions posed are clear and not too radical or expensive (Donmoyer 2012).

But here's the rub—the answers to questions of public policy and program evaluation are often not very clear (nor indeed are the questions sometimes). More circumspect proponents of experimental methods, specifically randomized controlled trials (RCTs), acknowledge that in order for a causal relationship to be established, even within the narrow terms of an RCT, very specific questions have to be asked. Thus, for example, Judith Gueron (2002) argues that while "random assignment . . . offers unique power in answering the 'Does it make a difference?' question" (p. 15), it is also the case that "[t]he key in large-scale projects is to answer a few questions well" (p. 40). In the same edited volume of papers, produced from a conference convened to promote "Randomized Trials in Education Research," Thomas Cook and Monique Payne (2002) agree that

> most randomized experiments test the influence of only a small subset of potential causes of an outcome, and often only one. . . . even at their most comprehensive, experiments can responsibly test

> only a modest number of the possible interactions between treatments. So, experiments are best when a causal question involves few variables [and] is sharply focused. (p. 152)

What these observations mean is that RCTs can be very good at answering very specific questions and attributing cause in terms of statistical probability. What they cannot do is produce the questions in the first place: That depends on much prior, often qualitative, investigation, not to mention value judgments about what is significant in the qualitative data and what is the nature of the problem to be addressed by a particular program intervention. Nor can RCTs provide an explanation of *why* something has happened (i.e. the underlying causal mechanisms at work). That, likewise, will depend on much prior investigation and, if possible, parallel qualitative investigation of the phenomenon under study, to inform the development of a theory about what the researchers think may be happening. Thus RCTs, even within their own paradigmatic terms, are actually only one part of a much longer chain of necessary and integral research activities, as indicated below in Figure 1:

Figure 1: Logic and Sequencing of RCT research design

Initial exploratory research	Possible explanations of phenomena: iteration of data and theory	More focused pilot studies developing a specific intervention	RCT	Wider implementation and further refinement (assuming the RCT is successful; often it is not)	Large scale dissemination and policy implementation

Thus without a reasonable understanding of why particular outcomes have occurred, along with identifying the range of unintended consequences that will almost inevitably accompany an innovation, it is very difficult to generalize such

outcomes and implement the innovation with any degree of success elsewhere. A good example of such problems is provided by California's attempt to implement smaller class sizes off the back of the apparent success of the Tennessee "STAR" evaluation. The Tennessee experiment worked with a sample, whereas California attempted statewide implementation, creating more problems than they solved by creating teacher shortages, especially in poorer neighborhoods in the state. There simply weren't enough well-qualified teachers available to reduce class size statewide, and those that were tended to move to schools in richer neighborhoods when more jobs in such schools became available (see Grissmer, Subotnik, & Orland, 2009). Interestingly, in this respect, Cook and Payne (2002) continue,

> The advantages of case study methods are considerable. . . . we value them as adjuncts to experiments. . . . Case study methods complement experiments when . . . it is not clear how successful program implementation will be, why implementation shortfalls may occur, what unexpected effects are likely to emerge, how respondents interpret the questions asked of them, [and] what the causal mediating processes are. . . . qualitative methods have a central role to play in experimental work. (p. 169)

Similarly, in their more recent review of evaluating social programs, Schorr and Farrow (2011) note that:

> Leading public and philanthropic funders are constructing a framework for what is considered credible evidence…however we suggest that the boundaries which the prevailing framework draws round acceptable evidence too greatly limit the knowledge base available…Programs and practices that are proven through experimental methods are an important component…but…are best seen as a take off point rather than a destination…[ ]…The problems that face us today tend to be caused by such complex forces that their course cannot be changed by isolated interventions…[ ]…too much potential for innovators and for improved outcomes will be lost if we continue to define credible evidence too narrowly (pp. iii, iv & vi).

One is tempted to ask, "So what's all the fuss about?" Why is some RCT advocacy so strident and exclusive? Of course, different researchers will vary in the importance they give to qualitative methods, and it is both puzzling and irritating to have qualitative methods reduced to an "adjunct" or a "complement" to experimental approaches, or as some activity to be undertaken before the "real" scientific work begins (see Shavelson, Phillips, Towne, & Feuer, 2003, p. 28). But it does seem as though those whose work actually involves the conduct of social science experiments have a well-informed view of the strengths of qualitative research, along with clear understandings of the limitations of experiments, as opposed to those who just engage in uninformed criticism of qualitative methods and advocacy for RCTs.

There is not enough space here to go into all the potential problems of conducting randomized experiments in the "natural" (as opposed to laboratory) setting of the school or the classroom. Extensive philosophical and practical critiques (and rejoinders) about the nature of causality and the place of RCTs in understanding social interaction and evaluating human services have been published by Erickson and Gutierrez (2002), Howe (2004), and Maxwell (2004, 2012), among many others. The debate goes back at least as far as Campbell and Stanley's (1963) classic paper on 'Experimental and Quasi-Experimental Designs for Research on Teaching', and they in turn, acknowledge in their introduction McCall's 'How to Experiment in Education', published in 1923. Current practitioners such as Gueron (2002) and Cook and Payne (2002), cited above, provide comprehensive accounts of the challenge of undertaking experiments "in the field." The real problem with experimental methods, however, is that even if conducted as effectively as possible, they often don't actually answer the "Does it make a difference?" question. Generating statistically significant results which, within the RCT paradigm are taken to 'prove' that an intervention works, is very difficult:

> Like a steady drip from a leaky faucet, the experimental studies being released this school year by the federal Institute of Education Sciences are mostly producing the same results: "No effects," "No effects," "No effects". The disappointing yield is prompting researchers, product developers, and other experts to question the

> design of the studies, whether the methodology they use is suited to the messy real world of education, and whether the projects are worth the cost, which has run as high as $14.4 million in the case of one such study. (Viadero, 2009, p. 1)

It is interesting to review the 'What Works Clearinghouse' website with such issues in mind. After 10 years and more of Federally mandated RCTs, does the Clearinghouse *actually* identify what works? The website is vast and a full review is beyond the scope of this paper, though that in itself is perhaps an indication of the unwieldy nature of much of the 'evidence' produced. Click on the What Works Clearinghouse website and the introductory copy states:

> **We review the research** on the different *programs, products, practices,* and *policies* in education. Then, by focusing on the results from *high-quality research*, we try to answer the question **"What works in education?"** *Our goal* is to provide educators with the information they need to make **evidence-based decisions**. (http://ies.ed.gov/ncee/wwc/ accessed 22/10/14, original emphases)

At the top of the page is a counter stating "10,498 studies reviewed" (and it does change – it stated "10,360 studies reviewed" when accessed two days previously – i.e. 138 studies were additionally reviewed, or at least added to the website, in two days).

Does any of this provide sound, informative and interesting information or "evidence"? On the front page, under the banner headline "Find What Works", key categories of substantive topics are listed. The first button is "improve literacy skills in 3rd graders". Click on this and you find: "Results. 57 Interventions found ", with a complex filtering system to search the 57 results – not exactly user-friendly for a busy teacher or administrator. Of the first three studies visible in the partially open search window, all state their "effectiveness rating", as "potentially positive effects", but with the "extent of evidence" listed as "small". Even this categorisation isn't very transparent however, since the actual numbers of studies reviewed in each case varies, from 12 to 100.

The front page of the website also has a button for “Intervention Reports”. Click on this and 553 results appear (accessed 20/10/14 & 22/10/14, so it would appear that none of the additional 138 studies noted above were “Intervention Studies”). In total, of the first ten studies on this Intervention site,
3 showed no evidence of effectiveness,
3 showed mixed outcomes (some “potentially positive” evidence and some not)
4 showed “potentially positive” evidence. No studies showed wholly “positive” results.
“Potentially positive” is defined as:

> At least one study shows a statistically significant or substantively important positive effect, AND No studies show a statistically significant or substantively important negative effect AND fewer or the same number of studies show indeterminate effects than show statistically significant or substantively important positive effects.

So no reports showed completely “positive” results and only four out of the first ten studies listed under “Intervention Reports” on the WWC website could report even “potentially positive” effects; and this rating could have been based on as little as one study. I could go on. Much of the material sampled is similarly ambiguous in its claims and limited in its usefulness. Of course this is not a representative sample of the whole WWC website and it is important not to treat the site unfairly; but equally it is perhaps not an untypical sample in-so-far-as it represents what a busy teacher or administrator might do in trying to find out what, does indeed, ‘work’. The site simply does not do this – it is far too complex and unwieldy, and most of the studies reviewed provide equivocal evidence at best.

Of course it could be argued that it is important to know when something doesn’t work, as well as when it does, or where there is no evidence one way or another. This is how ‘normal science’ operates (Kuhn 1962): it takes time, usually progresses in very small incremental steps, and sometimes not at all. But this hardly matches up to the overblown rhetorical claims that current ‘evidence initiatives’ definitely identify what *does* work. My point is not that there is anything necessarily wrong with providing this sort of data base, but that it just doesn’t do what it says on

the tin – it doesn't tell you, simply, easily, as promised, 'what works'. Moreover the rhetoric privileges one form of research design and activity over others, and this would seem to be the real purpose. As Lather (2010) observes:

> I am not against the scientific study of education. My issue is how the narrowly defined sense of science-based evidence in this effort at the federal level works to discipline educational research. Calls for policy research that support neo-liberal governmental initiatives must be challenged for what they are: bad science for bad politics (p.37)

Nor are such issues restricted to the WWC website. The UK Cabinet Office issued a 'What Works' summary document in November 2014 (https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/378038/What_works_evidence_for_decision_makers.pdf , accessed 09/02/15). The document summarises evidence from the government's 'What Works Network' covering social policy issues including crime reduction and social care, as well as education. The document claims it will lead to a "step-change...in government" (p.5) but actually provides similarly equivocal evidence when it comes to the specifics of intervention. Thus for example one of the highlighted areas in Education states:

> Students in a class with a teaching assistant do not, on average...perform better than those with only a teacher. However...teaching assistants can have a positive impact if they are trained to support pupils...in well-structured interventions (p.13).

This rather begs the question of how teaching assistants might be trained, by whom, for how long, and in what kinds of "well-structured interventions".

The Education evidence used in the UK Cabinet Office summary derives from the Education Endowment Foundation (EEF) website (http://educationendowmentfoundation.org.uk/). The EEF is a charity set up by government to fund intervention studies in education. It links with an independent charity The Sutton Trust. The website includes a "Toolkit" which reviews interventions and rates them, similar to the What Works Clearinghouse. In its own words:

> The Sutton Trust-EEF Teaching and Learning Toolkit is an accessible summary of educational research which provides guidance for teachers and schools on how to use their resources to improve the attainment of disadvantaged pupils.
> http://educationendowmentfoundation.org.uk/toolkit/about-the-toolkit/ accessed 09/02/15.

The Toolkit rates each intervention by cost, evidence and impact, defined in terms of additional months of schooling that the intervention produces. One of the most effective interventions listed in the Toolkit in terms of impact is "Early years intervention":

> early years and pre-school intervention is beneficial with above average levels of impact (a typical impact of six additional months' progress)...

However, the Toolkit also reports that "Early years intervention" is very expensive and goes on to note that:

> In most studies, the impact on attainment tends to wear off over time...Early years and pre-school interventions are therefore not sufficient to close the gap in attainment for disadvantaged children...
> http://educationendowmentfoundation.org.uk/toolkit/early-years-intervention/ accessed 09/02/15

Under a sub-heading 'How secure is the evidence' we are informed that:

> There are a number of systematic reviews and meta-analyses which have looked at the impact of early childhood intervention. Most of these are from the USA however...Evaluations of Sure Start in the UK do not show consistent positive effects
> http://educationendowmentfoundation.org.uk/toolkit/early-years-intervention/ accessed 09/02/15

So once again, even apparently strong evidence of 'what works' is hedged around with caveats, and begs many questions about the detail of the interventions.

Two other sites similar to the What Works Clearinghouse exist in the UK, based at the Institute for Effective Education (IEE) at the University of York and the Centre for Effective Education (CEE) at Queen's University, Belfast. IEE's inaugural director was Robert Slavin and he is still listed on its website as a current member of

staff. The IEE website lists 15 pages of 'intervention programs' with evidence ratings, a total of 90 programs in all. They are rated by the allocation of red dots – three dots ("proven"), two dots ("moderate"), one dot ("limited") and no dots. Only the first seven entries out of ninety score three dots (proven), with the next 11 studies categorised as two dots (moderate evidence), and the next 3 as one dot (limited evidence). The remaining 69 studies are categorised as "not evaluated". (http://www.evidence4impact.org.uk/programmes.php#search_results accessed 10/10/14). So, to reiterate, only seven out of ninety studies (7/90) are rated as "proven". The rating system defines "proven" as "has been shown to work in multiple well-controlled studies. This intervention has a good chance of improving your pupils' outcomes if it is implemented as designed". Yet again, this rather begs the question of what is meant by "a good chance" and how the intervention could indeed be "implemented as designed". Other evidence, of how to implement the intervention, and what the practical problems might be, is clearly needed.

The Centre for Effective Education website claims that it is "Transforming Education Through Evidence. Education for transformation" (http://www.qub.ac.uk/research-centres/CentreforEffectiveEducation/# accessed 11/10/14). Unlike WWC and IEE the Centre does not list large numbers of reviewed studies, but reports its own studies. Under "Current Projects" the website states: "The Centre for Effective Education has an impressive portfolio of research projects; currently running 10 major studies, of which 6 are randomised controlled trials". However, of those projects listed under 'Current Projects…Randomised Controlled Trials', no reports are available, including from 4 projects which ostensibly finished in 2012. Of those listed under "Completed Projects…Randomised Controlled Trials" the most recently completed (in 2013) is "The DELTA Parenting Programme Evaluation: DELTA evaluation a randomised controlled trial". The findings are summarised as:

> Overall this randomised controlled trial showed that parents who took part in the Early Years DELTA parenting programme reported increased feelings of confidence (parental efficacy), specifically in relation to three of the nine outcome areas measured, namely parents' confidence in: their knowledge of

> their child's development and needs; their self-acceptance as a good parent, and; disciplining and setting boundaries for their child... This trial provided no evidence of change in the remaining outcome areas (p.6) http://www.qub.ac.uk/research-centres/CentreforEffectiveEducation/Filestore/Filetoupload,421854,en.pdf (accessed 11/10/14)

Thus even an apparently positive study indicates that the intervention made no positive difference to six out of nine of the intended outcomes. The study also notes that use of local libraries was impacted negatively. The report includes extensive discussion of the limitations of the study, among them being that the measures used involved parents of young children self-reporting through completing questionnaires as to whether or not they *felt* "increased feelings of confidence" and so forth. What difference the intervention made to the *actual* way in which parents interacted with their children would have required a different (more qualitative, observation-based) research design. The report is to be commended for being appropriately cautious about what it can claim, but equally the claimed 'gold standard' legitimacy of an RCT is largely illusory.

Similar observations can be made about other reports on the CEE website. They read like many publically available technical research reports produced by many scholars over many years - basically saying 'trust us, we're the experts, we've done the stats'; classic 'scientistic-style' reporting. Overall there is nothing wrong with these reports – they are unremarkable evaluation reports that any competent university department or commercial consultancy could have produced. But this, of course, is the point – they are unremarkable. They cannot be said to be "Transforming Education through Evidence" so one has to wonder about why there is so much focus on RCTs, why now, and why they are surrounded by all the self-promotional rhetoric? I have, of course, answered my own question with this last observation – what we are talking about here are not arguments about 'better' or 'worse' approaches to research design, but rather ways of positioning oneself in the market place of competitive research bidding. It is also interesting to note that a very significant focus of many of the prominent studies across all these websites seems to involve early years

interventions of various kinds – changing disruptive behavior, improving parenting skills, improving early literacy and numeracy, and so forth. One is moved to wonder why might this be – it is a relatively small aspect of what we might include in any overall definition of 'education', albeit an important one; and can such studies really tell us any more than we already know from many years of both qualitative and quantitative research on early years education? Poverty is the issue in this context, rather than, or at least in addition to, the effectiveness of early years education.

An important issue here is the difference between 'research' and 'evaluation'. RCTs can provide some limited evidence in relation to specific programs and interventions, but it is puzzling as to how such an approach to evaluation has come to dominate policy discussions of 'research' *per se*. We should not, however, be surprised that the evidence produced by RCTs is so equivocal. Campbell and Stanley (1963) noted that there have been regular periods of RCT advocacy and RCT disillusionment in educational research as the clear cut results that RCTs promise have been unforthcoming. It was precisely the confounding problems of diverse implementation and interaction effects that produced so many "no significant difference" results in the 1960s in the context of curriculum evaluation studies. Reflections on such results prompted the development and use of qualitative methods in evaluation studies in the first place, in the1970s and 1980s (Cronbach, 1975; Cronbach & Associates, 1980; Guba & Lincoln 1981, 1989; Hamilton, Jenkins, King, MacDonald, & Parlett, 1976; Stake, 1967, 1978; Stenhouse, 1975; Stenhouse, Verma, Wild, & Nixon, 1982). Indeed, in one mixed method study of the "problems and effects of teaching about race relations" (as issues of race were called in the UK in those days), it was reported that 60% of the sample student population became less racially prejudiced as measured by attitude tests after following a particular program, but 40% became *more* prejudiced. As the author himself mused, what on earth is one supposed to do with such a result (Stenhouse et al., 1982)?

It is interesting in this respect to note that Schorr and Farrow (2011), in addition to arguing that complex social problems cannot be changed by one-off "isolated interventions" (p. iv) however well evaluated, also call for "more responsive, sensitive and cost effective ways of learning in real time…" (p.2). It was

Bob Stake, of course, as far back as 1973 who called for the development of "responsive evaluation":

> To be of service and to emphasize evaluation issues that are important for each particular program, I recommend the responsive evaluation approach. It is an approach that sacrifices some precision in measurement, hopefully to increase the usefulness of the findings to persons in and around the program...Responsive evaluation is based on what people do naturally to evaluate things: they observe and react...An educational evaluation is responsive evaluation (1) if it orients more directly to program activities than to program intents, (2) if it responds to audience requirements for information, and (3) if the different value-perspectives of the people at hand are referred to in reporting the success and failure of the program. (Stake 1973, pp 4-5)

Beyond Single Studies: Systematic Reviewing

One response of those interested in unpacking the problems with RCTs highlighted above would probably be to conduct further detailed investigation of the program as implemented, using a range of other methods. I shall return to the issue of mixed methods research designs below. However, a different approach has been advanced by those committed to experimental design but who acknowledge the potential weakness of relying on single studies—that of so-called 'systematic reviewing'. Advocates of systematic reviewing argue that evidence to inform policy should be accumulated across studies, but not just any studies, rather, only those that pass strict tests of quality. And those tests of quality have until relatively recently involved focusing on large-scale samples and, ideally, experimental designs (Gough & Elbourne, 2002; Oakley 2000, 2003). The case for developing systematic reviewing is based on transparency of process and clear criteria for including and excluding studies from the review. The case derives from critiques of so-called 'narrative reviewing', which, it is claimed, focuses on summarizing findings, in relation to a particular argument, rather than reviewing the whole field dispassionately and "systematically" so that the reader can be confident that all relevant prior knowledge in a field has been included and summarized. Arguments in favor of conducting such reviews reflect the critiques of social and educational research outlined earlier: that

the findings of empirical studies are often too small-scale, non-cumulative, or contradictory to be useful. Advocates are closely associated with the Cochrane Collaboration in medical and health care research and the Campbell Collaboration in social science, both of which favor the accumulation and dissemination of research findings based on scientific methods, particularly randomized controlled trials. As such, systematic reviewing is very much located within the international "evidence-based policy and practice" movement (Davies, 2004; Davies & Boruch, 2001; see also Mosteller & Boruch, 2002, p. 2, for evidence of the close networking of this international movement).

The original criteria of quality employed by systematic reviewing clearly derived from the medical model, but it is interesting to note that even as some researchers continue to argue the relevance of an RCT-based medical model to educational and social research, it has been criticized as inadequate in the field of medicine itself. Medical researchers understand that many issues of patient treatment and care require the design of qualitative as well as quantitative studies, in order to understand the ways in which patients respond to diagnosis and treatment, and substantial developments have tried to find ways of integrating the findings of qualitative studies into systematic reviews (e.g., Barbour & Barbour, 2003; Dixon-Woods, Booth, & Sutton, 2007; Dixon-Woods, Fitzpatrick, & Roberts, 2001). As Major and Savin-Baden (2011) note: "myriad approaches to the synthesis of qualitative research have now arisen" in the "health and medical professions" (p. 646). Such developments indicate that qualitative data is appreciated as important in understanding the conduct and impact of medical processes and interventions.

The original "hard line" position of systematic reviewing in social research has been significantly modified, as it has encountered considerable skepticism over the last several years (Hammersley, 2001; MacLure; 2005; cf. also Oakley's 2006 response, and Hammersley's 2008 rejoinder). Different kinds of research findings, including those of qualitative research, are now routinely included in such reviews, though with caveats about the quality of evidence deployed. Attempts have also been made to appraise the quality and thus the "warrant" of individual qualitative research studies and their findings before inclusion However, this can lead toward absurdity

rather than serious synthesis as the complexity of qualitative work is rendered into an amenable form for instant appraisal. Thus, for example, Attree and Milton (2006) report on a "Quality Appraisal Checklist . . . [and its associated] quality scoring system . . . [for] "the quality appraisal of qualitative research" (p. 125). Studies are scored on a 4-point scale:

A No or few flaws

B Some flaws

C Considerable flaws, study still of some value

D Significant flaws that threaten the validity of the whole study (p. 125)

Only studies rated A or B were included in the systematic reviews that the authors conducted, and in the paper they attempt to exemplify how these categories are operationalized in their work. But their descriptions beg many more questions than they answer. The above scale simply provides a reductionist checklist of mediocrity. Even the most stunning and insightful piece of qualitative work can only be categorized as having "No or few flaws". The point about qualitative research of course, is that its persuasiveness depends on its insight, not adherence to a particular approach to fieldwork or analysis. And this is precisely the point at issue with respect to using research to inform policy: Standards and checklists *cannot* substitute for informed judgment when it comes to balancing the rigor of the research against its potential contribution to policy. This *is* a matter of judgment, both for researchers and for policy makers.

Systematic reviewing is also very expensive and inefficient in terms of time and material resources, given the little it often delivers in terms of actual 'findings'. This is a problem it shares with the conduct of individual RCTs of course. The results of systematic reviews can take many months to appear, and policy makers are as likely to ask for very rapid reviews of research to be conducted over a few days or weeks, and possibly assembled via an expert seminar, as to commission longer-term systematic reviews (Boaz, Solesbury, & Sullivan, 2004, 2007). However, the more general issue for this chapter is the impact of the 'scientific evidence movement' on qualitative research, and the above checklist produced by Attree and Milton (2006) well illustrates the contortions that some qualitative researchers are prepared to go

through in order to maintain the visibility of their work in the context of this movement.

Impact on Qualitative Research: Setting Standards to Control Quality?

Another major response to the evidence movement has been for organizations and associations to start trying to "set standards" in qualitative research, and indeed in educational research more generally, to reassure policy makers about the quality of qualitative research and to reassert the contribution that qualitative research can (and should) make to government-funded programs. However, the field of qualitative research, or qualitative inquiry, is very broad, involving large numbers of researchers working in different countries, working in and across many different disciplines (anthropology, psychology, sociology, etc.), different applied research and policy settings (education, social work, health studies, etc.), and different national environments with their different policy processes and socioeconomic context of action. It is not at all self-evident that reaching agreement across such boundaries is desirable, even if it were possible. Different disciplines and contexts of action produce different readings and interpretations of apparently common literatures and similar issues. It is the juxtaposition of these readings, the comparing and contrasting within and across boundaries that allows us to learn about them and reflect on our own situated understandings of our own contexts. Multiplicity of approach and interpretation, and multivocalism of reading and response, are the basis of quality in the qualitative research community and, it might be argued, in the advancement of science more generally. The key issue is to discuss and explore quality across boundaries, thereby continually to develop it, not fix it, as at best a good recipe and at worst a narrow training manual.

Nevertheless, various attempts at "setting standards" are now being made, often, it seems, with the justification of "doing it to ourselves, before others do it to us" (Cheek, 2007; see also the discussion by Moss et al., 2009). In England, independent academics based at the National Centre for Social Research (a not-for-profit consultancy organization) were commissioned by the Strategy Unit of the UK government Cabinet Office to produce a report on "Quality in Qualitative Evaluation:

A Framework for Assessing Research Evidence" (Cabinet Office, 2003a). The rationale seems to have been that UK government departments are commissioning policy evaluations in the context of the move toward evidence-informed policy and practice and that guidelines for judging the quality of qualitative approaches and methods were considered to be necessary. The report was produced under a different government and before the latest renewed focus on experimental design in the UK (Goldacre 2013, discussed above). Nevertheless it provides an interesting insight into what constitutes officially sanctioned qualitative research.

The framework is a guide for the commissioners of research when drawing up tender documents and reading reports, but it is also meant to influence the conduct and management of research and the training of social researchers (Cabinet Office, 2003a, p. 6). However, the summary "Quality Framework" begs many questions, while the full report reads like an introductory text on qualitative research methods. Paradigms are described and issues rehearsed, but all are resolved in a bloodless, technical, and strangely old-fashioned counsel of perfection. The reality of doing qualitative research and indeed of conducting evaluation, with all the contingencies, political pressures, and decisions that have to be made, is completely absent. The implication is that one would have to comply with everything in the framework in order for one's work to be regarded as high quality. The issues which are highlighted are indeed important for social researchers to take into account in the design, conduct, and reporting of research studies. However, simply listed as issues to be addressed, they comprise a banal and inoperable set of standards that beg all the important questions of conducting and writing up qualitative fieldwork. Everything cannot be done; *choices* have to be made: How are they to be made, and how are they to be justified?

To be more positive for a moment, and note the arguments that might be put forward in favor of setting standards, it could be argued that if qualitative social and educational research is going to be commissioned, then a set of standards that can act as a bulwark against commissioning inadequate or underfunded studies in the first place ought to be welcomed. It might also be argued that this document at least demonstrates that qualitative research was being taken seriously enough within

government at that time to warrant a guidebook being produced for civil servants. The Framework might be said to confer legitimacy on civil servants who still want to commission qualitative work in the face of the policy move to RCTs; on qualitative social researchers bidding for such work; and indeed on social researchers more generally, who may have to deal with local research ethics committees (RECs; IRBs in the USA), which are predisposed toward a more quantitative natural science model of investigation. But should we really welcome such "legitimacy"? The dangers on the other side of the argument, as to whether social scientists need or should accede to criteria of quality endorsed by the state, are legion. In this respect, it is not at all clear that, *in principle,* state endorsement of qualitative research is any more desirable than state endorsement of RCTs.

Similar guidelines and checklists have appeared in the USA. Ragin, Nagel, and White (2004) report on a "Workshop on Scientific Foundations of Qualitative Research," conducted under the auspices of the National Science Foundation and with the intention of placing "qualitative and quantitative research on a more equal footing . . . in funding agencies and graduate training programs" (p. 9). The report argues for the importance of qualitative research and thus advocates funding qualitative research *per se,* but equally, by articulating the "scientific foundations" it is arguing for the commissioning of not just qualitative research, but a particular form of qualitative research. Moreover when it comes to the basic logic of qualitative work, Ragin et al. (2004) do not get much further than arguing for a supplementary role for qualitative methods:

> Causal mechanisms are rarely visible in conventional quantitative research . . . they must be inferred. Qualitative methods can be helpful in assessing the credibility of these inferred mechanisms. (p. 15)

Ragin et al. (2004) also conclude with another counsel of perfection:

> These guidelines amount to a specification of the *ideal* qualitative research proposal. A strong proposal should include as many of these elements as feasible. (p. 17, emphasis original)

But again, that's the point: What is *feasible* (and relevant to the particular investigation) is what is important, not what is ideal. How are such crucial choices to be made? Once again, "guidelines" and "recommendations" end up as no guide at all; rather, they are a hostage to fortune whereby virtually any qualitative proposal or report could be found wanting.

A potentially much more significant example of this tendency is the American Educational Research Association (AERA) "Standards for Reporting on Empirical Social Science Research in AERA Publications" (AERA, 2006). The Standards comprise eight closely typed double-column pages and include "eight general areas" (p. 33) of advice, each of which is subdivided into a total of 40 subsections, some of which are subdivided still further. Yet only one makes any mention of the fact that research findings should be interesting or novel or significant, and that is the briefest of references under "Problem Formulation," which we are told should answer the question of "why the results of the investigation would be of interest to the research community" (p. 34). Intriguingly, whether the results might be of interest to the *policy* community is not mentioned as a criterion of quality.

As is typical of the genre, the Standards include an opening disclaimer that:

> The acceptability of a research report does not rest on evidence of literal satisfaction of every standard. . . . In a given case there may be a sound professional reason why a particular standard is inapplicable. (p. 33)

But once again, this merely restates the problem rather than resolves it. The Standards may be of help in the context of producing a book-length thesis or dissertation, but no 5,000-word journal article could meet them all. Equally, however, even supposing that they could all be met, the article might still not be worth reading. It would be "warranted" and "transparent," which are the two essential standards highlighted in the preamble (p. 33), but it could still be boring and unimportant.

It is also interesting to note that words such as *warrant* and *transparency* raise issues of trust. They imply a concern for the very existence of a substantial data set as well as how it might be used to underpin conclusions drawn. Yet the issue of trust is only mentioned explicitly once, in the section of the Standards dealing with "qualitative methods": "It is the researcher's responsibility to show the reader that the

report can be trusted" (AERA, 2006, p. 38). No such injunction appears in the parallel section on "quantitative methods" (p. 37); in fact, the only four uses of the actual word *warrant* in the whole document all occur in the section on "qualitative methods" (p. 38). The implication seems to be that quantitative methods really are trusted—the issue doesn't have to be raised—whereas qualitative methods are not. Standards of probity are only of concern when qualitative approaches are involved.

Mixed Methods Research

A further response to current debate has been the development, or, perhaps more accurately, the rediscovery and redevelopment of mixed methods research. Mixing methods in social research and program evaluation has a long history. The argument has been that no single method could afford a complete purchase on the topic under study (Bryman, 1988: Denzin, 1970). Evaluations have routinely employed a range of methods to investigate the site-based specifics of program interpretation and adoption, alongside more general surveys of implementation and outcomes across sites (Greene, et.al. 1989). However over the last ten years or so, the 'Field' of 'Mixed Methods Research' (MMR) has increasingly been exerting itself as something separate, novel, and significant, such that proponents such as Tashakkori and Teddlie (2003) claim: "Mixed methods research has evolved to the point where it is a separate methodological orientation with its own worldview vocabulary and techniques" (p. x). Johnson et.al (2007) argue that "Mixed methods research…is becoming increasingly…recognised as the third major research approach or *research paradigm*…" (pp. 112, original emphasis).

More recently, as such views have been challenged, interrogated and augmented, the arguments have been modified. The claim to a distinct third paradigm is left open, not least because other MMR advocates have criticised the whole notion of paradigms somehow driving and determining research methods and have argued instead for a more grounded and pragmatic approach to understanding what researchers actually do, and how different approaches are actually combined in action (Christ 2009, Greene 2008, Harrits 2011, Morgan 2007, Tashakkori & Teddlie 2010). In addition to debates about mixed methods *per se*, it is also the case that mixed methods research has been alighted upon as a way to engage and modify the debate

about RCTs, and embed qualitative research in larger scale mixed methods studies. Thus for example Mason (2006) argues for qualitative methods to "drive" mixed methods research; Hesse-Biber (2010a,b) and Mertens (2007, 2010) argue for the use of qualitative methods to advance social justice issues in large scale investigations, and to enhance the "credibility" of RCTs (Hesse-Biber 2012).

The problem however, noted by many in the field (e.g. Hesse-Biber 2010a), is that qualitative research is too often allocated a subservient role in mixed method designs and also that the language of pragmatism (e.g. Morgan 2007, 2014) can shade into the treatment of research design as simply a technical matter, without reference to purpose, values or social justice. In this respect it does seem that 'Mixed Methods Research' (capital 'M', capital 'M', capital 'R') is being presented as a new and better form of science; a more complex approach to research designed to address the more complex problems that social research now faces (Caracelli, 2006; Creswell, et. al. 2011, Schorr and Farrow 2011). Whether such a position really benefits the development of qualitative research is a moot point. The subservience of qualitative methods would appear to be a particular issue in the context of government-sponsored evaluation studies, when arguing for the inclusion of qualitative methods is often predicated on what detail it can supply to flesh out the bones of a survey or an experiment in the context of a demand for empirical evidence of 'what works'. Thus for example, Valerie Caracelli (2006) in an article written from the perspective of the US Government Accountability Office reports several such instances. She argues for the inclusion of qualitative methods alongside the use of surveys and randomised controlled field trials "to assure contextual understanding" (p. 84). She states that: "Recently, there has been an acknowledgement about how ethnographic studies can inform agency actions and how it can be used to study culture in organisations" (p.87). Similarly in the UK Cabinet Office report reviewed above (Cabinet Office 2003b), one of the key quotes in the report used to justify the use of qualitative methods comes not from the epistemological or methodological literature, but from a civil servant, a government department "research manager":

> I often commission qualitative research when it's about users or stakeholders and . . . I want to understand . . . how a user is likely to respond. . . . I want to

> know how they see the world…it's a wonderful vehicle . . . if you want to understand the motives of people. (Cabinet Office, 2003b, p. 34).

So agencies and policy makers on both sides of the Atlantic value qualitative research for the insight it can provide into the "culture" of organisations and the "motives" of the people who work in them. It can certainly be argued that policy and its evaluation will benefit from being grounded in such data, but it is equally the case that such evidence could be used to monitor compliance with policy rather than to evaluate it. Qualitative data could even be used to allocate blame to individuals at the local level if implementation is found to be ineffective, rather than critique policy or dissemination strategies. Evaluation requires that policy and policymakers are themselves also rendered subject to scrutiny, not that (qualitative) social research methods are simply used to provide data for government.

Capacity Building, Professionalization, and the Retreat Into 'Science'

One response to the defining of standards and guidelines and the development of larger scale mixed methods research designs is simply to accept them at face value. In many respects they are unremarkable, very much 'business as usual' in the long history of social and educational research. And yet such documents, and renewed calls for mixed methods research, carry more import in current circumstances —they also legitimate a particular delineation and control of the discourse surrounding qualitative research. In so doing, and in combination with other interventions such as the increasing reach of ethics committees and government regulation of research activity (Department of Health, 2005; Lincoln & Tierney, 2004; Torrance, 2006), they are beginning to change the very social relations of research and the ways in which issues of research quality have hitherto been addressed. Pursuing and developing quality in qualitative research has always involved reading key sources iteratively and critically, in the context of designing and conducting a study, and discussing the implications and consequences with doctoral supervisors, or colleagues or project advisory groups. *Setting* standards in qualitative research, however, is a different enterprise. It implies the identification of universally appropriate and applicable procedures, which in turn involves documentary and institutional realization and compliance.

These developments are taking place in the context of moves towards what we might term ‘big social science’ – the concentration of research resources on large scale inter-disciplinary and inter-institutional programs of research. Governments around the world are seeking better value for money from their investment in research and this has involved restricting and focusing resource allocation. Governments, funding agencies and individual universities are now concentrating resources on fewer research units and programs, and are taking decisions to develop a 'big science' model of social science. This is being pursued by funders supporting fewer, larger projects, with explicit policy encouragement for researchers to develop cross-institutional, mixed method approaches, to address the supposedly 'big issues' of our time: health and well being, an aging population, sustainable growth, and so forth. These issues are indeed important, and research evidence should be produced to interrogate and inform public debate. But such issues are being presented as part of a common-sense, taken-for-granted trade-off of government funding in exchange for social scientists serving policy. Critique, diversity of perspective, and the insight into complexity which detailed qualitative studies can provide are potentially being marginalized. Social science is being reconceptualised as a technical service to government rather than developed as a democratic intellectual resource for the community.

Much of the activity associated with such moves goes under the heading of “capacity building,” certainly in the UK. As the government seeks to concentrate research resources in a smaller number of universities and extract maximum economic and social value from them, “centers of excellence” are being promoted, along with a concomitant obligation for these centers to link with and train in standard procedures those left stranded outside them (Department for Business, Innovations, and Skills, 2009; ESRC, 2009a, 2009b, 2015; National Centre for Research Methods, n.d.; Torrance, 2014, 2015). Similar aspirations also seem to be emerging in the USA (Eisenhart & De Haan, 2005; NRC, 2005; Walters, Lareau and Ranis 2009). It seems, then, that what is going on here is a struggle over the political future and bureaucratic institutionalization of social research. What we are witnessing is a crucial moment in the continuing professionalization of social research. Governments are looking to control and quality-assure the process of social research and in so doing are treating researchers as an almost directly employed category of government worker in the

"nationalized industry" of knowledge production for the knowledge economy. Concentration of research resources, coupled with large-scale collaborative programs of research projects, provides the context for and link into networked programs of capacity building and professional development for researchers. This provides a context for the development of a cadre of professional social researchers, oriented to funded policy priorities, and outside of or certainly additional to what might be termed the traditional scholarly route of disciplinary PhDs and individual monographs. The production of new textbooks in long-standing, but newly prominent fields such as Mixed Methods Research might be said to be a manifestation of the same phenomenon (e.g. Tashakorri and Teddlie 2003, 2010). What was once an ordinary approach to research design, has become a major intervention in the training of social researchers.

This in turn provides threats and opportunities for researchers as they seek to position themselves as both independent and autonomous sources of disinterested (i.e., scientific) advice, but nevertheless trustworthy professionals who can be relied upon to focus on topics of interest to policy and deliver a high-quality product. Thus, some researchers are attempting to respond to the pressure of policy and the evidence movement by producing defensive documents that emphasize the need for professional standards and self regulation (i.e., the AERA Standards above). In so doing, they implicitly accept the charges leveled against them. As Rizvi and Lingard (2010, p. 6) observe: "Policies...proffer solutions to the problem constructed by the policy itself". Such responses also appeal to and attempt to reassert the independence of "science" and the scientific community as a self-regulating group which, while broadly inclusive, nevertheless has clear boundaries and not only can define and protect standards, but *will.*

Other researchers are seeing opportunities to redefine the field and their place within it (i.e., their status and access to research funding). This is similarly being pursued by an appeal to science, but it involves a much more exclusive and narrow interpretation of science—defined by method (RCTs and systematic reviews) rather than broad approach. However, this latter group seems increasingly out of step with government demands for utility. One-off interventions studies do not address the

issues of large-scale inter-disciplinary approaches to complex social problems; and in any case too many such studies simply show "no effects" or only "potentially positive" effects. They are also expensive, as are systematic reviews, and hardly represent value-for-money. Thus it would appear that such researchers are deploying the rhetoric of science as part of a competitive struggle with other researchers for resources, rather than in any direct response to the supposed needs of government.

Despite the current policy commitment to RCTs, continued funding for such a narrow version of science seems unlikely if results are not more useful. The irony here is that despite repeated calls to make educational and social research more like medical research, social policy research demands much quicker results and truncated timescales. 'Normal science' takes time. It progresses very slowly by the accumulation of many routine and piecemeal studies. RCTs in medicine constitute part of this routine. It is taken for granted that new therapies could well take 15-20 years to move from initial laboratory observations to 'proven' available treatments (cf Balas and Boren 2000). It could similarly be argued in education that many years of "no effects" RCTs will be required before a sound knowledge base of what does indeed 'work' can be produced. However social systems, including education systems, are dynamic and change over time – they cannot be held constant while the problems of today are finally solved 20 years hence. The problems, and our interpretations and understandings of them, will have long since morphed and changed and research activity must engage with this changing context. Social research and social policy are in an iterative relationship not a linear one.

Science Is Not Enough: Toward a Different Approach

Interestingly, just as we've been here before with respect to 1960s/1970s disillusionment with research results that regularly showed "no significant difference," so we've been here before with respect to the response of the research community. Barry MacDonald (1974/1987) identified similar tensions over what role the research community should play in evaluating educational innovations. He identified three ideal types of approaches to evaluation—*autocratic, bureaucratic,* and *democratic,* aligning autocratic evaluation with scientific research, bureaucratic

evaluation with technical collaboration, and democratic evaluation with providing information for the widest possible public audience in real time:

> Autocratic evaluation is a conditional service to . . . government. . . . It offers external validation of policy in exchange for compliance with its recommendations. . . . the evaluator . . . acts as expert adviser. . . Bureaucratic evaluation is an unconditional service to . . . government. . . . The evaluator . . . acts as a management consultant [and] the report is owned by the bureaucracy and lodged in its files. . . . Democratic evaluation is an information service to the whole community about the characteristics of an educational program. . . . The democratic evaluator recognises value pluralism and seeks to represent a range of interests. . . . techniques of data gathering and presentation must be accessible to non-specialist audiences. (pp. 44–45)

MacDonald (1974/1987) also argued that evaluation reports should not simply be technically competent scientific documents but rather: "The criterion of success is the range of audiences served. The report aspires to 'best seller' status" (p. 45).

Of course, times change and the parallels with current debates are not exact. In particular, the obviously favored stance of "democratic evaluation" still presupposes that data can be gathered and interests represented in a fairly straightforward, realist fashion. Such aspirations would be more complex to accomplish now. Yet such a formulation also resonates with contemporary issues around stakeholder involvement, voice, and the engagement of a wider community in deciding which research questions are important to ask and how best to try to answer them. It is now widely recognized from many different perspectives, including that of the empowerment of research subjects on the one hand, and policy relevance and social utility on the other, that an assumption of scientific disinterest and independence is no longer sustainable. Other voices must be heard in the debate over scientific quality and merit, particularly in an applied, policy-oriented field such as education. Thus, for example, Gibbons el al. (1994) distinguish between what they term Mode 1 and Mode 2 knowledge, with Mode 1 knowledge deriving from what might be termed the traditional academic disciplines, and Mode 2 knowledge deriving from and operating within "a context of application":

> [I]n Mode 1 problems are set and solved in a context governed by the, largely academic, interests of a specific community. By contrast, Mode 2 knowledge is carried out in a context of application. (p. 3)

Such knowledge is "transdisciplinary . . . [and] involves the close interaction of many actors throughout the process of knowledge production" (p. vii). In turn, quality must be "determined by a wider set of criteria which reflects the broadening social composition of the review system" (p. 8).

The language employed by Gibbons et al. (1994) and the assumed context of operation very much reflect an engineering/technology-transfer type set of activities, but they also mirror a far wider set of concerns with respect to redefining the validity and social utility of research. There is a clear orientation toward the co-creation of knowledge through collaborative problem-solving action—rather than the discovery of knowledge through centralized, "expert" experimental investigation, which then gets disseminated to "practitioners" at the periphery. Ideas about the co-creation of knowledge link with deliberative and empowerment models of evaluation (Fetterman, 2001; House & Howe, 1999), which in turn owe something to MacDonald's (1974/1987) original notion of "democratic evaluation" (explicitly so, in House & Howe's case). The concept of "Mode 2 knowledge" also reflects something of the arguments around indigenous knowledge (Smith, 2005) produced *in situ* for local use, and the many articulations and interrogations of how to identify and represent different "voices" in research (e.g., Alcoff, 1991; Fielding, 2004; Goodley, 1999; Jackson & Mazzei, 2009; Shahjahan 2011). Such arguments, coalescing into a diverse, contested, but nevertheless highly provocative and promising constellation of issues around the validity, utility, and ethics of social research, also bring us to the very limit of what it is currently possible to think about the relationship of qualitative inquiry to science, policy, and democracy. The challenge we face is how to sustain the tension between interrogating and reconceptualizing problems—"thinking the new"—while also addressing the "here and now" of the enduring social and political issues that face our society (see Lather, 2004, 2010). The issue is how to reconcile the (research) need to investigate and comprehend complexity with the (policy) urge to

simplify and act. To invert Marx, policy makers seek to change the world, but first they need to try to understand it, while involving others in both processes.

Social research faces profound challenges at the present time as decades of empirical research, stretching back to the Chicago School of the 1930s and indeed beyond, repeatedly rediscover social problems and re-describe them in contemporary terms, rather than solve them. Policy looks to social research for advice, even 'proof', but it cannot be provided in the form that is being sought. The current policy consensus seems to be that we need large scale concentrations of research power, involving multi-institutional and multi-disciplinary investigations of the major issues of our time: community poverty and early family intervention; health and well-being; globalised security, and so forth, underpinned by experimental design to identify which interventions will work best to change human behavior. As neo-liberal capitalism (re)creates the problems, governments and, increasingly, large scale philanthropic organizations which have been produced in the context of neo-liberalism, are trying to address them by squeezing more and more out of essentially the same model of social research as has been available for nearly a century. Meanwhile researchers in turn seek to protect and enlarge their funding sources and position themselves better in relation to the increasingly competitive funding environment. These are the circumstances in which social research now operates and it may be that larger scale collaborations are indeed required and inevitable. But equally the values and processes of qualitative research must be asserted as important in and of themselves if such collaborations are to connect with communities in such as way as to empower them in their own search for solutions, and not simply render them subject to 'proven' interventions. As Lather (2010) asks:

> How do we break the hold of the natural science imaginary and, in Foucault's terms (1982) "refuse what we are" in colluding in the reduction of qualitative research to an instrumentalism that meets the demands of audit cultures? (p.64).

The scholarly retreat into trying to define the "scientific" merit of qualitative research simply in terms of theoretical and methodological standards, rather than in wider terms of social robustness and responsiveness to practice, seems to betray a

defensiveness and loss of nerve on the part of the scholarly community. We need to acknowledge and discuss the imperfections of what we do, rather than attempt to legislate them out of existence. We need to embody and enact the deliberative process of academic quality assurance, in collaboration with research participants, not subcontract it to a committee. Assuring the quality of research, and particularly the quality of qualitative research, must be conceptualized as a vital and dynamic process that is always subject to further scrutiny and debate. The process cannot be ensconced in a single research method or a once-and-for-all set of standards. Furthermore, it should be oriented toward risk taking and the production of new knowledge, including the generation of new questions (some of which may derive from active engagement with research respondents and policy makers) rather than supplication, risk aversion, and the production of limited data on effectiveness for a center-periphery model of system maintenance ("what works").

What this means for the actual conduct of social research, particularly qualitative research, over the medium to long term is still difficult to say, but various examples are emerging. These involve designing studies with collaborating sponsors and participants, including policy makers and those "on the receiving end" of policy, and talking through issues of validity, warrant, appropriate focus, and trustworthiness of the results, rather than trying to establish all of the parameters in advance (see, e.g., Fine et. al. 2012; James, 2006; Pollard, 2005, Somekh et al., 2007; Somekh & Saunders, 2007; Torrance et al., 2005). As Schorr and Farrow (2011) observe,

> We need to be able to draw on…evaluation combined with…other research, theory, practice and…local wisdom (p.7).

More radically, Fine et. al. (2012) argue that:

> Public science…is…strategic research for organizing campaigns on the ground and…[ ]…critical, participatory forms of inquiry, organizing and policy development (p. 687).

Such work can also involve new forms of dissemination and intellectual engagement with participants, rather than the simple reporting of "research findings". Thus for example Holmes, MacLure & Jones (2010) have produced a film based on a recent research project that seeks to disturb rather than settle the nature of the issues under investigation and through such provocation produce new ideas and practices on how

to address the issues from those involved (see also MacLure, Holmes, Macrae and Jones 2010).

Such processes are not without their problems or critics of course, especially with respect to issues of co-option into a too closely defined "bureaucratic" agenda—policy makers and sponsors usually being rather more powerful than research participants. But in essence, my argument is that if research is to engage critically with policy and practice, then research and policy making must progress, both theoretically and chronologically, in tandem. Neither can claim precedence in the relationship. Research should not simply "serve" policy, far less seek to determine it; equally, policy cannot simply "wait" for the results of research. And just as participant and practitioner perspectives (often called research "end-users" by policy makers) may be used by policy to attempt to discipline the research agenda pursued by researchers, equally, such perspectives can be used to critically interrogate policy. Research will encompass far more than simply producing policy-relevant findings; policy making will include far more than simply disseminating and acting upon research results. Where research and policy do cohere, the relationship should be pursued as an iterative one, with gains on both sides.

Might this be accomplished in current circumstances? It would likely require a dialogue to develop between proponents of experimental design and the field of qualitative research, and sharing that dialogue with policymakers. Although proponents of RCTs seem to see no need for such a dialogue at present, a further wave of RCT disillusionment, following the lack of definitive evidence produced by expensive trials, could well see the tide turn again. But rather than simply surfing the next wave of enthusiasm for qualitative research, perhaps the opportunity might be taken to acknowledge overlapping concerns and interests. Proponents of RCTs certainly need to acknowledge that there is more to scientific method than designing and evaluating interventions. Equally qualitative researchers might think more openly about what experiments in educational and social research should involve. Experimentation is about interrupting the taken-for-granted, doing something different, trying something out to see what happens, creating the new. It actually has a long and distinguished history in qualitative research - through Garfinkel's

experiments in ethnomethodology (Garfinkel 1967) to the origins of action research (Lewin 1946, 1947). If we can open up a wider definition of experimental work, involving the exploration of new ideas *in situ*, and combine this with a better understanding of scientific method grounded in curiosity, observation, interpretation and judgement, then a dialogue about the purpose and methods of research, and the role of research participants in producing knowledge for their own benefit and that of their communities, might develop.

Ultimately, the issue revolves around whether or not quality is protected and advanced by compliance with a particular set of standards and procedures, or by the process of open democratic engagement and debate. Governments, and some within the scholarly community itself, seem to be seeking to turn educational and social research into a technology that can be applied to solving short-term problems, thereby also entrenching the power of the expert in tandem with the state. With respect to qualitative research in particular, governments, as noted above, will always require detailed knowledge of social issues and how policy is understood and operationalised in action. The issue is how to make this engagement a two-way process and produce research as a democratic resource in all senses of the word – i.e. involve respondents in its production, with all the attendant ethical and local capacity-building issues; and have the resultant reports and policies accessible to public debate.

The current debate is being conducted in a research environment influenced by the uncertain status and legitimacy of both science *and* government. Government, and the process of mainstream electoral politics, is itself generally unpopular and under pressure 'to deliver', especially with respect to economic competence and with regard to the provision of public services. In such circumstances how appropriate is it for social research simply to serve government attempts to influence behavior and develop interventions? It is at least arguable that government intervention can dis-empower communities, and it certainly locates agency in government and professional bureaucracies, including those of social science, rather than local communities. A different approach would involve social research helping to build communities' capacities to develop themselves, rather than simply providing evidence for central policymaking and the development and evaluation of government

intervention programs. In this respect it may be the case that deriving legitimacy for social research from proximity to government is self-defeating. Such a strategy links social research to an inherently unpopular institution and at one and the same compromises the basic claim for the legitimacy of science – that of disinterested inquiry. Of course many qualitative researchers also want to pursue a social justice agenda, not just a scientific agenda, and in so doing engage with government and policy. Similarly definitions of what counts as science are changing in any case, as reviewed above. But here too, collaborating with local organizations, institutions and communities, rather than, or at least in addition to, government, would seem to hold much more promise with respect to both the quality of the research and its potential 'impact' on social and economic life.

An alternative vision proposes research as a system of reflective and engaged enquiry that might help practitioners and policy makers think more productively about the nature of the problems they face and how they might be better addressed. And in fact, the latter process will be as beneficial to policy as to research. Producing research results takes time, and, as we have seen above, such results are unlikely to be unequivocal. Drawing policy makers and practitioners into a discussion of these issues will improve the nature of research questions and research design, while also signaling to them that the best evidence available is unlikely ever to be definitive—it should inform and educate judgment, but it cannot supplant judgment, nor should it. Both the concept and the practice of science and government are under severe pressure at present, and ironically, despite all the recent criticisms of qualitative research, it is qualitative research that is best placed to recover and advance new forms of science and government, precisely because it rests on direct engagement with research participants. Many discussions of quality in qualitative research revolve around issues of engagement, deliberation, ethical process, and responsiveness to participant agendas, along with the need to maintain a critical perspective on both the topic at hand and the power of particular forms of knowledge (Denzin 2010; Lincoln, 1995; Schwandt, 1996; Lather, 2004, 2010; Smith, 2005, Tracy 2010). It is these strengths of a qualitative approach that are needed to reinvigorate the research enterprise and reconnect it with democratic processes.

Notes

1. See Grissmer, Subotnik, and Orland (2009) for another illustration of the significance of qualitative data in focusing research questions and modifying the analyses of an experimental study of housing provision

References

Alcoff, L. (1991, Winter). The problem of speaking for others. *Cultural Critique,* 5–32.

American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. *Educational Researcher, 35*(6), 33–40.

Attree, P., & Milton, B. (2006). Critically appraising qualitative research for systematic reviews: Defusing the methodological cluster bombs. *Evidence and Policy, 2*(1), 109–126.

Baez, B., & Boyles, D. (2009). *The politics of inquiry: Education research and the "culture of science."* Albany: State University of New York Press.

Balas E. and Boren S. (2000) 'Managing clinical knowledge for health care improvements' in Bemmel J. and McCray A. (2000 Eds) *Yearbook of Medical Informatics 2000: Patient-Centred Systems* Stuttgart, Schattauer Verlagsgesellschaft mbH

Barbour, R., & Barbour, M. (2003). Evaluating and synthesizing qualitative research: The need to develop a distinctive approach. *Journal of Evaluation in Clinical Practice, 9*(2), 179–185.

Besley, T. (Ed.). (2009). *Assessing the quality of educational research in higher education: International perspectives.* Rotterdam, The Netherlands: Sense.

Blunkett, D. (2000). Influence or irrelevance: Can social science improve government? Speech to the Economic and Social Research Council (ESRC), reprinted in *Research Intelligence,* British Educational Research Association, No. 71 and Times Higher 4 February 2000 accessable at: http://www.timeshighereducation.co.uk/story.asp?storyCode=150012&sectioncode=26

Boaz, A., Solesbury, W., & Sullivan, F. (2004). *The practice of research reviewing 1: An assessment of 28 review reports.* London: UK Centre for Evidence-Based Policy and Practice, Queen Mary College.

Boaz, A., Solesbury, W., & Sullivan, F. (2007). *The practice of research reviewing 2: Ten case studies of reviews.* London: UK Centre for Evidence-Based Policy and Practice, Queen Mary College.

Boden, R., Cox, D., Nedeva, M., & Barker, K. (2004) *Scrutinising science: The Changing UK government of science.* London: Palgrave.

Bridges, D. (2005, December 16). *The international and the excellent in educational research.* Paper prepared for the Challenges of the Knowledge Society for Higher Education conference, Kaunas, Lithuania.

Bridges, D. (2009). Research quality assessment in education: Impossible science, possible art? *British Educational Research Journal, 35*(4), 497–517.

Brown A. (2010) 'Qualitative method and compromise in applied social research' *Qualitative Research* 10, 2, 229-248

Brown, S. (2003, September 17). *Assessment of research quality: What hope of success?* Keynote address to European Educational Research Association annual conference, Hamburg, Germany.

Bryman A. (1988) *Quantity and Quality in Social Research*, London, Unwin Hyman

Bush, V. (1945, July). *Science: The endless frontier.* A report to the president by Vannevar Bush, Director of the Office of Scientific Research and Development. Washington, DC: U.S. Government Printing Office. Available at http://www.nsf.gov/od/lpa/nsf50/vbush1945.htm

Cabinet Office. (2003a). *Quality in qualitative evaluation: A framework for assessing research evidence* [Summary]. London: Author.

Cabinet Office. (2003b). *Quality in qualitative evaluation: A framework for assessing research evidence* [Full report]. London: Author.

Campbell D. and Stanley J. (1963) *Experimental and Quasi-Experimental Design for Research* Cengage Learning

Cantor and Smith (2013) 'Rethinking Science Funding' *USA Today* 30/09/13, accessed 02/01/2014

Caswell C. and Wensley R. (2007) 'Doors and Boundaries: a recent history of the relationship between research and practice in UK Organisation and Management Research' *Business History* 49, 3, 293-320

Caracelli V. (2006) 'Enhancing the policy process through the use of ethnography and other study frameworks: a mixed methods strategy' *Research in the Schools* 13, 1, 84-92

Ceglowski D, Bacibalupa C. and Peck E. (2011) 'Aced Out: Censorship of Qualitative Research in the Age of 'Scientifically Based Research'' *Qualitative Inquiry* 17, 8, 679-686

Cheek J. (2007). *Qualitative inquiry, ethics, and the politics of evidence. Qualitative Inquiry, 13*(8), 1051–1059

Christ T. (2009) 'Designing, teaching and evaluating two complementary mixed methods research courses' *Journal of Mixed Methods Research* 3, 4, 292-325

Christ T. (2014)'Scientific-Based Research and Randomised Controlled Trials, the 'Gold' Standard? Alternative paradigms and mixed methodologies' *Qualitative Inquiry* 20, 1, 72-80

Coburn Report (2011) *The National Science Foundation: under the microscope* http://www.coburn.senate.gov/public/index.cfm?a=Files.serve&File_id=2dccf06d-65fe-4087-b58d-b43ff68987fa accessed 30/10/14

Cook, T., & Payne, M. (2002). Objecting to the objections to using random assignment in educational research. In F. Mosteller & R. Boruch (Eds.), *Evidence matters: Randomized trials in education research* (pp. 150–178). Washington, DC: Brookings Institution Press.

Creswell J, Klassen A, Plano Clark V. and Smith K. (2011) *Best Practice for Mixed Methods Research in the Health Sciences* Office of Behavioral and Social Sciences Research (OBSSR), National Institutes of Health (NIH)

Cronbach, L. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist, 30,* 116–127.

Cronbach, L., & Associates. (1980). *Toward reform of program evaluation*. San Francisco: Jossey-Bass.

Davies, P. (2004). Systematic reviews and the Campbell Collaboration. In G. Thomas & R. Pring (Eds.), *Evidence-based practice in education* (pp. 21–33). Maidenhead, UK: Open University Press.

Davies, P., & Boruch, R. (2001). The Campbell Collaboration. *British Medical Journal, 323,* 294–295.

Denzin N. (1970, 1978 & 1989: 1st, 2nd and 3rd editions) *The Research Act* Prentice Hall, New Jersey

Denzin N. (2010) *The Qualitative Manifesto: a call to arms* Left Coast Press

Depaepe M. (2002) 'A comparative history of educational sciences: the comparability of the incomparable' *European Educational Research Journal,* 1, 1, 118-122

Department for Business, Innovation, and Skills. (2009). *Higher ambitions: The future of universities in a knowledge economy.* Available at http://www.bis.gov.uk/policies/higher-ambitions

Department for Business Innovation and Skills (2011) *Innovation and research strategy for growth,* Cm 8239. Retrieved from http://www.bis.gov.uk/assets/biscore/innovation/docs/i/11-1387-innovation-and-research-strategy-for-growth.pdf

Department of Health. (2005). *Research governance framework for health and social care* (2nd ed.). London: Author.

Dixon-Woods, M., Booth, A., & Sutton, A. (2007). Synthesizing qualitative research: A review of published reports. *Qualitative Research, 7*(3), 375–422.

Dixon-Woods, M., Fitzpatrick, R., & Roberts, K. (2001). Including qualitative research in systematic reviews: Opportunities and problems. *Journal of Evaluation in Clinical Practice, 7*(2), 125–133.

Donmoyer R. (2012) 'Two (Very) Different Worlds: the cultures of policymaking and qualitative research' *Qualitative Inquiry* 18, 9, 798-807

Economic and Social Research Council. (2009a). *Postgraduate training guidelines.* Available at http://www.esrc.ac.uk/_images/Postgraduate_Training_and_Development_Guidelines_tcm8-2660.pdf

Economic and Social Research Council. (2009b). *Capacity building clusters.* Available at http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/research/CapacityBuildingClusters/index.aspx

Economic and Social Research Council (2015) *Strategic Plan – 2015: Social science shaping society* Available at: http://www.esrc.ac.uk/_images/Postgraduate_Training_and_Development_Guidelines_tcm8-2660.pdf

Eisenhart, M., & DeHaan, R. (2005). Doctoral preparation of scientifically based education researchers. *Educational Researcher, 34*(4), 3–13.

Erickson, F., & Gutierrez, K. (2002). Culture, rigor, and science in educational research. *Educational Researcher, 31*(8), 21–24.

Fetterman, D. (2001). *Foundations of empowerment evaluation.* Thousand Oaks, CA: Sage.

Fielding, M. (2004). Transformative approaches to student voice: Theoretical underpinnings, recalcitrant realities. *British Educational Research Journal, 30*(2), 295–311.

Fine M, Ayala J, and Zaal M. (2012) 'Public science and participatory policy development: reclaiming policy as a democratic project' *Journal of Education Policy* 27, 5, 685-692

Frederiksen L. and Beck S. (2010) 'Caught in the crossfire: educational research in context' *International Journal of Research and Method in Education* 33, 2, 135-149

Garfinkel H. (1967) *Studies in Ethnomethodology* Englewood Cliffs NJ, Prentice Hall

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). *The new production of knowledge.* Thousand Oaks, CA: Sage.

Goldacre B. (2013) *Building Evidence into Education Department for Education,* London https://www.gov.uk/government/news/building-evidence-into-education

Goodley, D. (1999). Disability research and the "researcher template": Reflections on grounded subjectivity in ethnographic research. *Qualitative Inquiry, 5*(1), 24–46.

Gorard, S., & Taylor, C. (2004). *Combining methods in educational and social research.* Maidenhead, UK: Open University Press.

Gough, D., & Elbourne, D. (2002) Systematic research synthesis to inform policy, practice, and democratic debate. *Social Policy and Society, 1*(3), 225–236.

Greenberg, D. (2001). *Science, money, and politics* Chicago: University of Chicago Press.

Greene J. (2008) 'Is Mixed Methods Social Inquiry a Distinctive Methodology?' *Journal of Mixed Methods Research*, 2, 1, 7-22

Greene J. et. al (1989) 'Toward a Conceptual Framework for Mixed Methods Evaluation Designs' *Educational Evaluation and Policy Analysis* 11, 3, 255-274

Grissmer, D., Subotnik, R., & Orland, M. (2009). *A Guide to incorporating multiple methods in randomized controlled trials to assess intervention effects.* Available at http://www.apa.org/ed/schools/cpse/activities/mixed-methods.aspx

Guba, E., & Lincoln, Y. (1981). *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches.* San Francisco: Jossey-Bass.

Guba, E., & Lincoln, Y. (1989). *Fourth generation evaluation.* Newbury Park, CA: Sage.

Gueron, J. (2002). The politics of random assignment: Implementing studies and affecting policy. In F. Mosteller & R. Boruch (Eds.), *Evidence matters: randomized trials in education research* (pp. 15–49). Washington, DC: Brookings Institution Press.

Hamilton, D., Jenkins, D., King, C., MacDonald, B., & Parlett, M. (1976). *Beyond the numbers game.* London: Macmillan.

Hammersley, M. (1997). Educational research and teaching: A response to David Hargreaves' TTA lecture. *British Educational Research Journal, 23*(2), 141–161.

Hammersley, M. (2001). On systematic reviews of research literature: A narrative response. *British Educational Research Journal 27*(4), 543–554.

Hammersley, M. (2005). The myth of research-based practice: The critical case of educational inquiry. *International Journal of Social Research Methodology, 8*(4), 317–330.

Hammersley, M. (2008). Paradigm war revived? On the diagnosis of resistance to randomized controlled trials and systematic review in education. *International Journal of Research and Method in Education, 31*(1), 3–10.

Hargreaves, D. (1996). *Teaching as a research-based profession.* Teacher Training Agency 1996 Annual Lecture. London: Teacher Training Agency.

Harrits G. (2011) 'More Than Method? A discussion of paradigm differences within mixed methods research' *Journal of Mixed Methods Research* 5, 2, 150-166

Haynes L, Service O, Goldacre B. and Torgerson D. (2012) *Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials* Cabinet Office Behavioural Insights Team, London, https://www.gov.uk/government/publications/test-learn-adapt-developing-public-policy-with-randomised-controlled-trials accessed 15/02/15

Hesse-Biber S. (2010a) 'Qualitative approaches to mixed methods practice' *Qualitative Inquiry* 16, 6, 455-468

Hesse-Biber S. (2010b) 'Feminist Approaches to Mixed Methods Research: linking theory and praxis' in Tashakkori A. and Teddlie C. (Eds. 2nd Edition) *Mixed Methods in Social and Behavioral Research* Sage

Hesse-Biber S. (2012) 'Weaving a Multimethodology and Mixed Methods Praxis into Randomized Control Trials to Enhance Credibility' *Qualitative Inquiry* 18, 10, 876-889

Higher Education Funding Council for England. (2009). *Research Excellence Framework.* Bristol, http://www.hefce.ac.uk/media/hefce1/pubs/hefce/2009/0938/09_38.pdf accessed 15/02/15

Higher Education Funding Council for England (2014) *Research Excellence Framework: results* Bristol, http://www.ref.ac.uk/ accessed 15/02/15

Hillage, J., Pearson, R., Anderson, A., & Tamkin, P. (1998). *Excellence in research on schools* (DfEE Research Report 74). London, Department for Education and Employment.

Holmes R, MacLure M and Jones (2010) *Addressing 'problem behaviour' in the early years: an innovative film resource* http://www.esri.mmu.ac.uk/resprojects/project_outline.php?project_id=127

House, E., & Howe, K. (1999). *Values in evaluation and social research.* Thousand Oaks, CA: Sage.

Howe, K. (2004). A critique of experimentalism. *Qualitative Inquiry, 10*(1), 42–61.

Jackson, A., & Mazzei, L. (Eds.). (2009). *Voice in qualitative inquiry.* London: Routledge.

James, M. (2006). Balancing rigor and responsiveness in a shifting context: Meeting the challenges of educational research. *Research Papers in Education, 21*(4), 365–380.

Johnson R. B. et. al. (2007) 'Toward a Definition of Mixed Methods Research' *Journal of Mixed Methods Research*, 1, 2, 112-133

Kemmis S. (2012) 'Research and educational praxis: spectator and participant perspectives' *British Educational Research Journal* 38, 6, 885-1074

Kuhn T. (1962) *The Structure of Scientific Revolutions* University of Chicago Press

Lagemann, E. (2000). *An elusive science: The troubling history of education research.* Chicago: University of Chicago Press.

Lather, P. (2004). This IS your father's paradigm: Government intrusion and the case of qualitative research in education. *Qualitative Inquiry, 10*(1), 15–34.

Lather P. (2010) *Engaging Science Policy from the Side of the Messy* New York, Peter Lang

Leach J. (2013) 'Science Communication' in Simonson P, Peck J, Craig R, and Jackson J. (2013 Eds) *Handbook of Communication History* London, Routledge

Lemke T. (2012) *Foucault, Governmentality and Critique* Paradigm Publishers, Boulder Co.

Lempert R. (2013) *Transformative Research, Silly Studies and the Future of NSF's SBE Directorate* available at: http://www.cossa.org/advocacy/DefundingSBE-RichardLempert.pdf

Lewin K. (1946) 'Action research and minority problems' *Journal of Social Issues* 2, 4, 34-46

Lewin K. (1947) 'Frontiers in group dynamics: concept, method and reality in social science, social equilibria and social change' *Human Relations* 1, 1, 5-41

Lincoln, Y. (1995). Emerging criteria for quality in qualitative and interpretive research. *Qualitative Inquiry, 1*(3), 275–289.

Lincoln, Y., & Tierney, W. (2004). Qualitative research and institutional review boards. *Qualitative Inquiry, 10*(2), 219–234.

MacDonald, B. (1987). Evaluation and the control of education. In R. Murphy & H. Torrance (Eds), *Evaluating education: Issues and methods.* London: Harper & Row. (Reprinted from MacDonald B. & Walker R. (1974 Eds) Innovation, Evaluation, Research and the Problem of Control, pp.9-22, SAFARI Interim Papers, Centre for Applied Research in Education, University of East Anglia, UK 1974)

MacLure, M. (2003). *Discourse in education and social research.* Maidenhead, UK: Open University Press.

MacLure, M. (2005). Clarity bordering on stupidity: Where's the quality in systematic review? *Journal of Education Policy, 20*(4), 393–416.

MacLure, M., Holmes, R., MacRae, C., & Jones, L. (2010). Animating classroom ethnography: Overcoming video-fear. In L. Mazzei & K. McCoy (Eds.), Thinking with Deleuze in qualitative research [Special issue]. *International Journal of Qualitative Studies in Education, 23*(5), 543–556.

Major C. H. and Savin-Baden M. (2011) 'Integration of Qualitative Evidence' *Qualitative Research* 11, 6, 645-663

Mason J. (2006) 'Mixing methods in a qualitatively driven way' *Qualitative Research* 6, 1, 9-25

Maxwell, J. (2004). Causal explanation, qualitative research, and scientific enquiry in education. *Educational Researcher, 33*(2), 3–11.

Maxwell J. (2012) 'The Importance of Qualitative Research for Causal Explanation in Education' *Qualitative Inquiry* 18, 8, 655-661

McCall W. (1923) *How to Experiment in Education* London, MacMillan

Mertens D. (2007) 'Transformative Paradigm: Mixed Methods and Social Justice' *Journal of Mixed Methods Research*, 1, 3, 212-225

Mertens D. Bledsoe K, Sullivan M, and Wilson A. (2010) 'Utilisation of Mixed Methods for Transformative Purposes' in Tashakkori A. & Teddlie C. (Eds) *Handbook of Mixed Methods in Social and Behavioural Research, second edition* Sage

Middleton, S. (2009). Becoming PBRF-able: Research assessment and education in New Zealand. In T. Besley (Ed.), *Assessing the quality of educational research in higher education: International perspectives* (pp. 193–208). Rotterdam, The Netherlands: Sense.

Mills, D., & Ratcliffe, R. (2012). 'After method: Ethnography in the knowledge economy'. *Qualitative Research, 12* (2), 147-164.

Morgan D. (2007) 'Paradigms lost and paradigms regained: methodological implications of combining qualitative and quantitative methods' *Journal of Mixed Methods Research* 1, 1, 48-79

Morgan D. (2014) 'Pragmatism as a Paradigm for Social Research' *Qualitative Inquiry* 20, 8, 1045-1053

Moss, P., Phillips, D., Erickson, F., Floden, R., Lather, P., & Schneider, B. (2009). Learning from our differences: A dialogue across perspectives on quality in education research. *Educational Researcher, 38*(7), 501–517.

Mosteller, F., & Boruch, R. (Eds.). (2002). *Evidence matters: Randomized trials in education research.* Washington, DC: Brookings Institution Press.

National Centre for Research Methods. (n.d.). *A Strategic Framework for Capacity Building within the ESRC National Centre for Research Methods (NCRM)* Available at http://www.ncrm.ac.uk/TandE/capacity/documents/NCRMStrategicFrameworkForCapacityBuildingMain.pdf

National Research Council. (2002). *Scientific research in education.* Washington, DC: Author.

National Research Council. (2005). *Advancing scientific research in education.* Washington, DC: Author.

Nisbet, J., & Broadfoot, P. (1980). *The impact of research on policy and practice in education.* Aberdeen, Scotland: Aberdeen University Press.

No Child Left Behind (NCLB 2002) *No child left behind act*. 115 STAT. 1965. Retrieved from http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf.

Oakley, A. (2000). *Experiments in knowing.* Cambridge, UK: Polity Press.

Oakley A. (2003). Research evidence, knowledge management and educational practice: Early lessons from a systematic approach. *London Review of Education, 1*(1), 21–33.

Oakley, A. (2006). Resistances to new technologies of evaluation: Education research in the UK as a case study. *Evidence and Policy, 2*(1), 63–88.

Peters M, Besley A, Olson M, Maurer S. and Weber S. (Eds. 2009) *Governmentality Studies in Education Rotterdam*, Sense Publishers

Pollard, A. (2005). Challenges facing educational research. *Educational Review, 58*(3), 251–267.

Ragin, C., Nagel, J., & White, P. (2004). *Workshop on scientific foundations of qualitative research.* Available at http://www.nsf.gov/pubs/2004/nsf04219/start.htm

Rizvi F. and Lingard B. (2010) *Globalising Education Policy* Routledge, London

Roberts, K., Dixon-Woods, M., Fitzpatrick, R., Abrams, K., & Jones, D. (2002). Factors affecting uptake of childhood immunization: A Bayesian synthesis of qualitative and quantitative evidence. *The Lancet, 360,* 1596–1599.

Schwandt, T. (1996). Farewell to criteriology. *Qualitative Inquiry 2*(1), 58–72.

Schorr L, & Farrow F. (2011) *Expanding the Evidence Universe: doing better by knowing more,* Centre for the Study of Social Policy, Washington DC
http://lisbethschorr.org/doc/ExpandingtheEvidenceUniverseRichmanSymposiumPaper.pdf
Also available at:
http://www.cssp.org/publications/harold-richman-public-policy-symposium/Expanding-Evidence-the-Evidence-Universe_Doing-Better-by-Knowing-More_December-2011.pdf
Both accessed 30/10/14

Shahjahan R. A. (2011) 'Decolonising the evidence-based education and policy movement: revealing the colonial vestiges in educational policy, research and neo-liberal reform' *Journal of Education Policy* 26, 2, 181-206

Shavelson, R., Phillips, D., Towne, L., & Feuer, M. (2003). On the science of education design studies. *Educational Researcher, 32*(1), 25–28.

Slavin, R. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher, 31*(7), 15–21.

Smith, L. (2005). On tricky ground: Researching the native in the age of uncertainty. In N. Denzin & Y. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed., pp. 85–107). Thousand Oaks, CA: Sage.

Somekh, B., & Saunders, L. (2007). Developing knowledge through intervention: Meaning and definition of "Quality" in research into change. *Research Papers in Education, 22*(2), 183–197.

Somekh, B., Underwood, J., Convery, A., Dillon, G., Jarvis, J., Lewin, C., et al. (2007). *Final report of the evaluation of the ICT Test Bed Project.* Coventry, UK: Becta.

Stake, R. (1967). The countenance of educational evaluation. *Teachers' College Record,* 68, 523–540.

Stake R (1973) 'Program evaluation, particularly responsive evaluation', *keynote presentation* at a conference on "New Trends in Evaluation" October 1973, Institute of Education, Göteborg University;
http://education.illinois.edu/circe/Publications/Responsive_eval.pdf
Accessed 30/10/14

Stake, R. (1978). The case study method in social inquiry. *Educational Researcher, 7*(2), 5–8.

Stenhouse, L. (1975). *An introduction to curriculum research and development.* London: Heineman.

Stenhouse, L., Verma, G., Wild, R., & Nixon, J. (1982). *Teaching about race relations: Problems and effects.* London: Routledge.

Tashakkori A. & Teddlie C. (2003, Eds) Handbook of Mixed Methods in Social and Behavioural Research Sage

Tashakkori A. & Teddlie C. (2010, Eds) Handbook of Mixed Methods in Social and Behavioural Research, second edition Sage

Tooley, J., & Darby, D. (1998). *Educational research: A critique.* London: Office for Standards in Education.

Torrance, H. (2006). Research quality and research governance in the United Kingdom: From methodology to management. In N. Denzin & M. Giardina (Eds.), *Qualitative inquiry and the conservative challenge* (pp. 127–148). Walnut Creek, CA: Left Coast Press.

Torrance, H. (2008). *Overview of ESRC research in education: A consultancy commissioned by ESRC: Final report.* Available at http://www.sfre.ac.uk/uk/.

Torrance H. (2014) 'Qualitative Research in the UK: short term problems, long term issues' *Qualitative Inquiry*

Torrance H. (2015) 'Investigating research power: networks, assemblages and the production of 'big' social science' in Cannella G, Perez M. & Pasque P. (Eds) *Critical Qualitative Inquiry,* Left Coast Press

Torrance, H., Colley, H., Ecclestone, K., Garratt, D., James, D., & Piper, H. (2005). *The impact of different modes of assessment on achievement and progress in the learning and skills sector.* London: Learning and Skills Research Centre.

Tracey S. (2010) 'Qualitative Quality: Eight 'Big Tent' Criteria for Excellent Qualitative Research' *Qualitative Inquiry* 16,10, 837-851

Viadero, D. (2009, April 1). "No effects" studies raising eyebrows. *Education Week.* Available at http://www.projectcriss.com/newslinks/Research/MPR_EdWk--NoEffectsArticle.pdf

Wade, C., Turner, H., Rothstein, H., & Lavenberg, J. (2006). Information retrieval and the role of the information specialist in producing high-quality systematic reviews in the social, behavioral, and education sciences. *Evidence and Policy, 2*(1), 89–108.

Walters P, Lareau A. and Ranis S. (2009) *Education on Trial: policy reform and the call for scientific rigour* New York, Routledge

Weiss, C. (1972). *Evaluating action programs.* Boston: Allyn & Bacon.

Weiss, C. (1980). *Social science research and decision-making.* New York: Columbia University Press.

Wilson R. (2013) 'The War on Social Science' *Symposium Magazine* http://www.symposium-magazine.com/the-war-on-social-science-rick-k-wilson/ accessed 02/01/14

Woodhead, C. (1998, March 20). Academia gone to seed. *New Statesman,* pp. 51–52.

Yates, L. (2004). *What is quality in educational research?* Buckingham, UK: Open University Press.