ESRC Teaching and Learning Research Programme
Research Capacity Building Network

Occasional Paper Series
Paper 47

How do we overcome the methodological schism (or can there be a 'compleat' researcher)?

Stephen Gorard

2002

Cardiff University School of Social Sciences

The Research Team

Executive Group:

Professor Stephen Gorard (Director)
Professor Gareth Rees
Professor John Furlong
Dr Laurence Moore
Professor Ken Prandy
Dr Ray Crozier

Research Staff:

Dr Chris Taylor
Karen Roberts
Patrick White
Helen Taylor (Project Administrator)

Our Contact Details

ESRC TLRP Research Capacity Building Network
Cardiff University School of Social Sciences
Glamorgan Building
King Edward VII Avenue
Cardiff
CF10 3WT

Tel. 029 2087 5345
Fax. 029 2087 4678
Email. TaylorH1@Cardiff.ac.uk

www.cardiff.ac.uk/socsi/capacity

How do we overcome the methodological schism (or can there be a 'compleat' researcher)?

Abstract

This paper considers a variety of approaches to combining research findings drawn from what are traditionally deemed 'quantitative' and 'qualitative' methods. These include models for Bayesian syntheses, new political arithmetic, complex interventions, and design experiments, as well as the more usual literature review and 'new' realism. I argue here that none of these approaches pose insurmountable epistemological or technical problems. Rather, opposition to the use of such models stems from wasteful 'paradigm' wars fed perhaps by fear of the unknown, and leading to pointless methodological schism. The 'compleat' researcher should presumably be prepared to find, use and critique *all* evidence relevant to their quest, regardless of its form.

Do we need a 'compleat researcher'?

Educational research capacity in the UK has come under attack from a variety of sources over the last decade (e.g. Millett 1997, Hillage et al. 1998, Tooley and Darby 1998, Woodhead 1998). The various complainants have different reasons for their criticisms, and do not always offer robust evidence in support of their claims (see Gorard 2001a). However, sufficient evidence has been presented to suggest that something is amiss. According to a recent study of educational research capacity in the UK, there is a lack of developed research expertise among many of the people involved, perhaps especially in comparison to other disciplines (McIntyre and McIntyre 2000). The outcome is that a 'large proportion' of the research done (in teaching and learning in this instance) is of poor quality. The skills that do exist are in specific regions and institutions, sometimes represented by one individual (Furlong and White 2001). These skills tend to be more common in 'qualitative' approaches, and many areas of educational research are now 'dominated by reports of small-scale local studies' (Dyson and Robson 1999, p.vi).

There is an apparent system-wide gap in expertise in large-scale numeric-based studies, especially field trials derived from laboratory experimental designs. This does not only apply to education (Marshall 2001), nor only to the UK (NRC 1999, Resnick 2000). A large proportion of existing statisticians in social science are apparently facing retirement, and potential replacements are being lost due to a widening pay gap with a private sector anxious for their expertise (Research Fortnight 2001). Perhaps partly as a response, over the last twenty years there has undoubtedly been a move towards much greater use of 'qualitative' approaches (Hayes 1992), even in traditionally numerate areas of educational research (Ellmore and Woehilke 1998). In addition, acceptance rates for 'qualitative' publications are higher than for 'quantitative' pieces, by a ratio of around two to one in one US journal (Taylor 2001). In an analysis of submissions to the annual American Educational Research Association conference it was found that 'qualitative' pieces out-numbered 'quantitative' pieces by three to one, and that pieces using mixed methods were by far the least common (Hausman 2000). Yet one of the key components of increasing the quality of educational research, as suggested by the newly formed National Educational Research Forum, is an increase in our capacity to conduct multi-method work.

This paper considers several of combining methodological approaches. It starts by considering the supposed schism between 'qualitative' and 'quantitative' work, and then rehearses some of the common ways in which this schism is already healed. The main sections of the paper consider in turn four more formal ways of combining approaches, namely: Bayesian synthesis; the new political arithmetic; the MRC model for complex interventions, and design experiments. The paper concludes with a discussion of some of the problems experienced and foreseen in working with combined approaches.

Overcoming the schism

It is particularly important for the well-being of educational research that we do not waste time in methodological ‘paradigm wars’ instead of concentrating on the development of all methods (Mahoney 2000). One practical advantage would be that we could cease wasting time and energy in pointless debates about the virtues of one or other approach. In particular we need to overcome the false dualism of ‘quantitative’ and ‘qualitative’ approaches (Pring 2000a). The supposed distinction between qualitative and quantitative evidence (Popkewitz 1984) is essentially a distinction between the traditional methods for their analysis rather than between a philosophy, paradigm, or method of data collection (Frazer 1995). To some extent all methods of educational research deal with qualities, even when the observed qualities are counted. Similarly, most methods of analysis use some form of number, such as ‘tend, most, some, all, none, few’ and so on. This is what the patterns in qualitative analysis are based on (even where the claim is made that a case is ‘unique’ since *un*iqueness is, of course, a numeric description). Words can be counted, and numbers can be descriptive. Patterns are, by definition, numbers, and the things that are numbered are qualities. Most crucially, it is important to realise that ‘qualitative’ and ‘quantitative’ are not differing research paradigms (at least not in the sense that Kuhn, 1970, uses the term). As Heraclitus has written, 'logic is universal even if most people behave differently' (for if logic were not universal we could not debate from common ground, so making research pointless). It is difficult to sustain an argument that all methods, including data collection, carry epistemological or ontological commitments (Bryman 2001).

Since the preponderance of work in UK educational research is qualitative in nature, it is usually an increase in awareness of quantitative skills that is demanded. This is not to prioritise one over the other, but is simply a recognition of current skill needs. There are several reasons why *all* researchers should learn something about quantitative techniques (Gorard 2001b). The first and most obvious point is that the process of research involves some consideration of previous work in the same field. All researchers read and use the research of others. So they need to develop what Brown and Dowling (1998) refer to as a 'mode of interrogation' for reading and using research results. It is not good for the health of the field if readers routinely either ignore or take on trust results involving statistics. Some numeric techniques are anyway common to all research - the choice and use of a sample, for example, is a very common phenomenon in all kinds of research using many different approaches to data collection and analysis. Similarly, all studies gain from a prior consideration of the available secondary sources of relevance. Existing statistics, whatever their limitations, provide a context for any new study which is as important as the 'literature review' and the 'theoretical background'. Above all, it is important to realise that quantitative research is generally very easy. Much analysis in social science involves calculations with nothing more complex than addition or multiplication - primary school arithmetic in fact. Even these, and any more complex calculations, are conducted by a computer.

I would not wish readers to infer from the above that I am advocating numeric techniques above all others, or defending in particular the record of quantitative researchers in education. I intend neither (Gorard 2001a, Gorard 2002a). In fact, another clear reason for a wider spread of quantitative skills is the need for wider critical review of such work. What is needed is more researchers able and motivated to use, read, and critique work based on all methods. This probably involves a increase in the number *and* quality of researchers with quantitative skills, and an increase in the quality of researchers with qualitative skills. UK educational research retains a generally mono-method culture, and will continue to do so while senior researchers in the field approve of it, and so inadvertently reinforce the methodological schism.

In some relatively established disciplines, such as psychology, there has been a tradition that only numeric data is of relevance. Students are therefore encouraged to count or measure everything, even where this is not necessarily appropriate (as with many attitude scales for example), and one outcome is that statistical analysis is done badly and so gets a bad press. Allied to this approach is a cultural phenomenon I have observed with some international students and their sponsors which again approves only research involving numbers. A corollary for both groups appears to be that forms of evidence not based on numbers are despised, while evidence based on numbers is accepted somewhat uncritically. This last is clearly a particular problem, as I quite regularly come across findings which when reanalysed show the opposite to what is being claimed (Gorard 1998).

On the other hand, within some disciplines including perhaps sociology there is now a culture that derides and rejects all numeric evidence. Having realised that numbers can be used erroneously, sometimes even unscrupulously, some researchers simply reject all numeric evidence. This is as ludicrous a position as its opposite. As Clegg (1992) points out, we know that people sometime lie to us but we do not therefore reject all future conversation. If we reject numeric evidence, and usually all of the associated concerns about validity, generalisability and so on, as the basis for research, then we are left with only subjective judgements. The danger therefore for 'qualitative' research conducted in isolation from numeric approaches is that it can be used simply as a rhetorical basis for retaining an existing prejudice. Without a combination of approaches we are left with no clear way of deciding between competing conclusions. Happily a philosophical background for such research is beginning to emerge (in Gorard 1997). The 'new realists' accept the imprecision of measurement, the impact of subjectivity, and the dangers of reductivism, and so strive for even greater rigour in their studies, in the form of "triangulation" between the methods within one investigation (Frazer 1995).

How are findings combined?

As the introduction has suggested, I do not believe that there are insuperable philosophical or technical difficulties involved in combining different forms of data. In fact, I know that such combination goes on in social science all of the time in many relatively uncomplicated ways. One of the purposes of this paper is to remind readers how combining different findings is an 'everyday' occurrence, before considering briefly some more developed models for formal combination.

When I conduct a literature review, as is normal at the start of any new project, I use all and any forms of evidence relevant to my topic. I use peer-reviewed papers, books, 'grey' literature such as websites, previous reviews, and personal communication with experts. I read sources involving theory, practice, method, and evidence of all sorts. I do not ignore or reject any source simply because of the *form* of evidence it provides, and this is reflected in the final result (e.g. Gorard 1999). I have no reason to believe that other reviewers do, or should, behave any differently. The balanced literature review is a very common example of combining data from different methods, requiring, of course, a working knowledge of both qualitative and quantitative techniques to allow the reviewer to be appropriately critical.

Similarly, when I conduct primary research I do not ignore or avoid evidence I encounter because it is of the wrong sort. I have had long letters attached to uncompleted questionnaires from respondents in a survey design (e.g. Gorard 1997). These are generally fascinating and useful. I would not dream of claiming that these, or the pencilled comments on the margins of a form, were of no use since they were not foreseen as part of the survey. Similarly, in conducting interviews with headteachers I would not refuse to accept a school brochure proffered during an interview, or some school statistics sent on after an interview (e.g. Gorard and Taylor 2002). When conducting a series of household structured interviews, the notes I take on the appearance of the house and its occupants can be very valuable (e.g. Gorard et al. 1999a). Once on the road to conduct research, everything is potentially informative and I become a 'hoover' of data, as far as possible. I start with draft research questions, then I attempt to answer them by whatever means it takes. I cannot imagine anyone doing anything very different. Why would a researcher spurn evidence relevant to the research because it was of the 'wrong' sort? Practical fieldwork is therefore another common example of 'combining' data relatively unproblematically.

Once you set your mind to it, examples of combining approaches abound. The methods of history have to cope with primary and documentary evidence, with information from genetics, archaeology, and linguistics, for example. While this is a skilled task, it is not usually considered particularly complex. In psychology, approaches such as personal construct theory explicitly advocate the mixing of numeric and 'qualitative' data. In designing a questionnaire, standard textbooks advocate the prior use of in-depth approaches, such as focus groups, to help identify and modify the relevant questions (e.g. Oppenheim 1992). In analysing interview data, standard textbooks describe a quasi-statistical approach to counting responses, in order to establish patterns. Even 'pure' statistical analysis is misunderstood by observers if they do not consider also the social settings in which it takes place, and the role of 'qualitative' factors in reaching a conclusion (Gephart 1988).

The next sections of the paper consider four more formal approaches to combining the results emerging from different methodological approaches.

Research syntheses: a Bayesian alternative?

The Cochrane/Campbell collaboration and the setting up of evidence-based centres for educational research are based on the notion of research syntheses (an idea with many merits and some problems, Gorard 2001a). Research can then be used as a basis for establishing 'evidence-based' practice where pedagogical and other decisions are guided by nationally agreed 'protocols' (as in the field of medicine, Department of Health 1997). Syntheses of high quality studies are used to produce the findings, which are then ‘engineered’ into practice. The assumption is therefore not that good evidence has not been provided by previous work, but that it is difficult to see its pattern without systematic evaluation (the work is fragmented, Pring 2000b), and impossible for it to have an impact on policy and practice without re-engineering. Simply publishing results is not enough. The beauty of this solution is that it apparently addresses issues of both relevance and quality, and it can be justified on solid practical grounds. For example, in a review of administering albumin to humans, Roberts (2000) concludes that it ‘provides a strong argument for preparing scientifically defensible syntheses of the evidence from randomised controlled trials in medicine, as well as in other important areas of social policy, such as education' (p.235). This approach sees large-scale randomised controlled trials as the ideal form of evidence, which a systematic review further improves by minimising bias through selection and omission, leading to safe and reliable results (Badger et al. 2000).

However, while plausible, this approach does face technical difficulties that are not always highlighted by its advocates. Steering research in the direction of experimental trials (Evans and Benefield 2001) means that ‘qualitative’ evidence is largely ignored, which is particularly wasteful (Levacic and Glatter 2001). Systematic reviews can therefore be misleading by hiding details, and privileging trials even where considerable evidence of other forms contradicts them. This has led to false conclusions that are just as important, in reverse, as those claimed for the evidence-based approach (Speller et al. 1997). Even in medicine, which receives a lot more funding than educational research, the approach is therefore being criticised (Hammersley 1997). Meta-analysis, or synthesis, of experimental evidence may show what works but it cannot uncover detailed causal mechanisms (Morrison 2001). 'It is unclear how an RCT can untangle this' (p.74), nor how it can pick up multiple (side) effects. More detailed data collected in conjunction with the trials may, however, be able to remedy these deficits. How can we combine these two forms of data within a research synthesis?

In medicine, qualitative evidence has been a traditional precursor to other research, such aiding the design of questionnaires or the selection of an intervention or outcome measure (Dixon-Woods et al. 1999). It has been particularly valuable in helping to challenge the tendency for research to reflect the clinicians' and not the patients' perspective. It has also been used to help explain quantitative results (as in new political arithmetic, see below), especially in explaining why an experimental trial does not work or will not generalise. It has not until recently been used in syntheses for a variety of reasons. Researchers are concerned that it may signal a return to haphazard reviews; qualitative work has less clear criteria for judging suitability for admission to the synthesis, and discussion of the issue tends to flounder on philosophical and epistemological problems rather than moving on to practicalities. One solution is clearly to treat qualitative work as small-scale quantitative work, and convert it to numeric form by frequency counting. Another possibility is meta-ethnography, but no actual examples of this have emerged yet (Dixon-Woods et al. 1999).

Another suggested formal solution to the problem of combining different forms of evidence in syntheses is based on Bayesian analyses (Roberts et al. 1999). This starts with the very credible assumptions that evidence about a phenomenon does not exist in a vacuum and that its likely impact on an observer will depend to some extent on that observers' prior beliefs about the topic (West and Harrison 1997). So, unlike in standard 'frequentist' statistics, judgements about evidence are subjective. Put another way, any observer will have a prior knowledge of the probability/uncertainty about any phenomenon. New evidence about a phenomenon provides a new likelihood that will modify, rather than completely override, that prior probability. Therefore, the same evidence does not lead to precisely the same posterior probability/uncertainty for all observers. When all observers then agree, whatever their prior position, this shows the convincing power of the new evidence. What Bayes and others have produced, and technological advances have now made feasible, is a method for calculating the posterior distribution, making it proportional to the new likelihood multiplied by the prior distribution (French and Smith 1997). Bayes theorem offers us a prescription of how to learn, collectively, from evidence (Bernado and Smith 1994).

One, of many, possible ways forward is to use qualitative evidence to help create the prior probability for a Bayesian model, which can then be adjusted via Bayes' theorem using the quantitative evidence in an otherwise normal synthesis (Roberts et al. 1999). In fact the prior probability can also be based on expert knowledge (the notion that some people intuitively know 'what works' under certain conditions) and on the results of prior meta-analyses. Successful syntheses *have* been conducted using this approach.

However, the Subjective Expected Utility (SEU) theory that underlies the approach is not without problems. It can really only be applied to genuine subjective uncertainties, yet its calculations are based on the assumption that people are coherent and rational (which they may not be). Bayesian probability is about how a person should decide, or bet (Hartigan 1983). It is subjective - the value of a bet is subject to market forces for example - but it is not 'anything goes'. There are tensions between SEU as a description of actual behaviour, and a prescription for best results (see Gorard 1997). 'People are usually seen as trying vaguely to be rational but failing frequently to appreciate normatively appropriate strategies' (Eiser and van der Plight 1988, p.76), and in practice they tend to ignore prior probabilities in favour of 'convincing' new evidence (Glymour et al. 1987).[1]

A new political arithmetic

Other than literature reviews or syntheses, perhaps the most frequently used way of combining research findings is what is termed here the 'new political arithmetic' model (NPA). In its simplest form it involves a two-stage research design. In the first stage, a problem (trend, pattern, or situation) is defined by a relatively large-scale analysis of relevant numeric data. In the second stage, this problem (trend, pattern, or situation) is examined in more depth using recognised 'qualitative' techniques with a sub-set of cases selected from the first stage. This approach clearly differs from attempting to both describe and explain phenomena using complex statistical techniques with the same dataset (Gorard 2002a). In NPA the numeric techniques are simple, and largely descriptive. The approach also clearly differs from attempting to both describe and explain phenomena using only in-depth approaches (which are, of necessity, smaller in scale). In NPA the explanatory phase collects new data, but in a focused attempt to elucidate findings from the descriptive phase. Each type of data has a different purpose for which it may be best suited.

In practice, the model is likely to be more complex than this. There may be different forms of data within each phase, more than two phases, or iteration between phases. Perhaps the most complex NPA study I have been involved in started with an analysis of the social composition of the student body in all secondary schools in England and Wales over 12 years (Gorard 2000). The results were used to select a sub-set of around half of all local education authorities (LEAs) for documentary analysis of the admission procedures to their schools (White et al. 2001). The results of this analysis were used to select a smaller number of adjacent LEAs in which to conduct interviews with admission officers (Fitz et al. 2002). On the basis of these interviews we selected a subset of schools in which to conduct interviews with headteachers within these LEAs (Gorard and Taylor 2002). Our use of large-scale secondary and documentary evidence supports the belief that the patterns we tried to explain were genuine ones, but also enables us to draw reasonably general conclusions from our interview data.

In teaching and learning research, this same model has been used quite widely with some success although it is rather limited to the simple and formulaic survey then case study combination (Suter and Frechtling 2000). For example, Spillane and Zeuli (1999) used a large-scale survey of teachers from the TIMSS study, and used the results to select cases of interest for further exploration. Schorr and Firestone (2001) conducted a survey of 245 teachers, from which they selected extreme cases for further case study.

In post-compulsory teaching and learning research, I have used the NPA model as follows. We used the national population census, Labour Force Survey, and local economic histories, to select research sites. We conducted a large-scale household survey, collecting learning and training biographies from families in the three selected sites (Gorard et al. 1999a). We conducted a second sweep of the survey to contact children of families from the first sweep. We conducted open-ended interviews with 10% of the families, selected to represent the variation from the survey, and supplemented these with interviews with teachers and trainers in each site, and with archived oral family histories from the same areas (Gorard and Rees 2002). Again, the advantage was that the in-depth datasets were easily examined for ways in which they could, and could not, explain the patterns derived from the secondary sources and the survey biographies (Gorard et al. 2001).

The method is one of progressive focusing, starting from a consideration of the importance of 'truth' (Bridges 1999), returning from complex statistical modelling to a prior political arithmetic tradition (Mortimore 2000), and adding other appropriate methods of data collection and analysis in subsequent phases. In this way, researchers tend to avoid what Brown (1992) calls the 'Bartlett' effect of producing false results when basing an analysis solely on qualitative data.

A model of complex interventions

As outlined above, the role of experimental trials (of many sorts, see Gorard 2001b) has been presented as crucial to progress in educational research as it has been valuable in medical research (e.g. Fitz-Gibbon 2001, Torgerson and Torgerson 2001). This approach to research is almost certainly under-used, and therefore valuable. However, the application of the design remains problematic, both in the evaluation of complex health services interventions (MRC 2000) and in the field of education (Hakuta 2000). There are important differences between education and the models of research and development used in industry and biomedicine (the inability to legislate for teacher classroom behaviour being one) and it has been argued that outcome measures in medical matters are generally less problematic than in education (Hammersley 1997). The measurement of ‘lifespan’ may be less ambiguous than that of ‘standards’, for example. Experimental designs should therefore be additional to, not replacements for, other recognised modes such as secondary analysis and detailed case studies in the way that multiple approaches are used in natural sciences (Chalmers 1990).

Within health promotion, the value of the Randomised Controlled Trial (RCT) research design has been hotly disputed (Moore 2002). The established wisdom is that complex health promotion interventions are most likely to be successful if they are based on sound theoretical concepts. While good quality RCTs have been undertaken, these have generally evaluated simple, naïve interventions, delivered in homeopathic doses, and have inevitably produced disappointing results. On the other hand, some good quality, complex interventions have not been *rigorously* evaluated, and their effectiveness has not been unequivocally demonstrated.

The MRC model of complex interventions is therefore a welcome and useful one (Campbell et al. 2000). In this model, the first phase would involve the initial design of an intervention based on current theoretical understanding, with an explicit underlying causal mechanism of its proposed effect. The second phase would involve the formative evaluation of that intervention, using qualitative approaches such as interviews, focus groups, observation and case studies to identify how the intervention is working, the barriers to its implementation, and how it may be improved (compare design studies, see below). The third phase is a feasibility study of the intervention, or full pilot study, involving both measurement and in-depth feedback. This phase also sees the main development of the alternative treatments or controls.

Traditionally, RCTs require that the interventions being tested are standardised and uniformly delivered to all participants. However, since health promotion and educational interventions are so dependent on the quality of delivery, the value of trials predicated on 'ideal' conditions is limited. For example, smoking education interventions have been found to work well in efficacy trials, when delivered by enthusiastic teachers with ample curriculum time, yet when implemented in the real world they have not been found to be effective (Nutbeam et al. 1993). Instead, a pragmatic approach is taken, with the intervention delivered in the trial in a lifelike way. This approach sacrifices standardisation for realism, and means that the natural variability in delivery that occurs between practitioners must be recorded and monitored by in-depth means (perhaps video recording) as well as the more traditional outcome measures.

It is crucially important to conduct a comprehensive qualitative investigation within the trial, so that the quality and variability of intervention delivery is monitored. If the intervention is not found to work, then the qualitative research may identify why this was the case, and if there is variation in the impact of the intervention, this may be related to the variability in implementation. Furthermore, by building into the RCT design a strong qualitative investigation, variability in context can be documented, and its impact on the intervention’s delivery and effectiveness can be identified. In summary, 'the RCT design ensures that an unbiased estimate of the average effect of the intervention is obtained, while the qualitative research provides useful further information on the external factors that support or attenuate this effect' (Moore 2002, p.5).

A role for design studies

A further 'realistic' alternative to rigidly controlled trials is the design experiment, or design study. Instead of conceiving education as a social science (like sociology) to be compared to the natural sciences (like physics), this approach sees education as an artificial or design science, like aeronautics, artificial intelligence, architecture, engineering, or medicine (Brown 1992). From this perspective the object of educational research is to produce or improve an artefact, such as a lesson plan, curriculum document or form of assessment. It is therefore an essentially formative process. Whereas the natural sciences are concerned more with how things work and how they may be explained, design sciences are concerned more with how artefacts behave under different conditions (Collins et al. 2001).

This view of education requires a different form of research. Traditional work on the psychology of learning and memory has been based on made-up situations in laboratory conditions, isolating the subject from its context. Design studies, on the other hand, take place in realistic settings. Collins et al. (2001) suggest seven major differences between the two forms of research. Experiments take place in artificial settings, focusing on a single outcome, controlling other contaminating variables, using fixed procedures, isolating the experimental subjects, and testing hypotheses, to a design created by the experimenter. Design experiments are messier, taking place in real settings, monitoring many dependent variables, characterising the situation ethnographically, revising the procedures at will, allowing participants to interact, developing profiles rather than hypotheses, and involving users and practitioners in the design.

As suggested by this summary, the model explicitly combines different approaches, and these approaches are differently appropriate to different phases. For example, imagine that the study starts with identification of an artefact to be progressively refined in practice. The initial idea and design for the artefact may come from prior knowledge, theory, serendipity, or inspiration. In this way it is very like the pre-clinical phase to identify an intervention in the MRC model (above), but expressed in more traditionally 'creative' terminology. Once the artefact is being used, this use is monitored in much the same way as 'gadgets' are monitored for consumer reports. The observations will be detailed, and will naturally contain numeric and non-numeric elements. This prolonged phase is like the modelling and exploration components of the MRC model. The qualitative element is like an ethnography, but is based on an intervention rather than a naturally occurring situation. The quantitative monitoring element is like a field experiment, except that the results can be used to guide live revision of the design. A close analogy from another design field is that of an aeroplane model in a wind tunnel. In fact many models would be tested in tandem, revised and re-tested. The data generated would be both observational and measurement. This analogy helps us because it is often argued that education is a complex phenomenon to research. But the same is also true of aeronautics with its competing claims for the speed, safety, comfort, efficiency and so on, for any design.

There is a debate among users of design studies in education about the need for a separate assessment of the final artefact. Some users claim to be doing design experiments by simply modifying multiple classroom interventions continuously and documenting the process (often via video recordings). It is not clear that this behaviour is any different to action research. In the original formulation, Brown (1992) assumed that work would iterate between laboratory and classroom - capturing the advantages of both. Clearly it was never the intention that the same dataset would be used to both modify and test the artefact/intervention. Similarly the outcome(s) of interest must be fixed first, else, if it is modified along with the intervention during the study, there is no fixed point to the research. The approach simply becomes a 'trawl' which will eventually find *something*. To retain rigour it is important that the design phase leads to a testable output. If this is tested, whether in the laboratory or no, it leads to a stage very similar to the 'definitive' trial of the MRC model. Put another way, design science gives us a well-tested approach to the model and hypothesis-generating stages of research - working towards a trial.

Nevertheless, in the design experiment, drawing on scientific and design literatures, we may have the best attempt yet to overcome those objections to the social science model of education. We can imagine, and I am working on as part of a study funded by the National Science Foundation, a composite model of the research process using elements of all of the approaches above. The result would be more complex than any of the above, allowing iteration between stages and even permitting 'failure' as a research outcome, for example. It would, however, retain one important characteristic for combining methods. At each phase, different methods would be more or less appropriate and complementary. Some phases would appear like art or craft, and some more like science. To me, it does not matter what they are called or what analogies and metaphors people find 'cosy' in describing the stages. Fitness for purpose is the key.

Potential problems

I realise that this paper has only begun a consideration of the problems of, and potential solutions to, combining different methods and the data they derive. In describing common ways in which we all routinely combine approaches it has not considered whether these are done well or not. It is, for example, possible to 'hoover' up fieldwork data of all types, but still have little idea what to do with it either separately or together. In dealing with more formal approaches to combining, the paper has not addressed the practical issue of what to do if two or more datasets disagree - where they do not 'triangulate' in traditional terms. These are both important issues, and ones that I have written about elsewhere (e.g. Gorard et al. 1999b). I prefer here to outline some of the wider objections readers may have to *any* approach to combining methods or data.

Perhaps the most basic objection is the near tautological one that we cannot combine methods because they are not combinable. In various guises, this is probably the most common problem. In essence, it has been suggested that there are two or more different fields of educational research which cannot interact. One cannot gainsay qualitative evidence with quantitative data, and vice versa, the argument goes. 'Some educational researchers evidently believe that the choice of a research method represents commitment to a certain kind of truth and the concomitant rejection of other kinds of truth' (Snow 2001, p.3). Such researchers think that the value of their methods can be judged apart from the questions they are used for. They are in evidence among student researchers who state that they want to 'do an ethnography', for example, without being clear on their research questions or even a topic. They are seeking a problem that fits the method they admire, and ignoring the fact that good and bad examples of the use of all methods abound. However, this attitude is also in evidence among senior researchers, who are therefore the trainers and reviewers of newer researchers (Gorard 2001a). As I have written elsewhere: 'This leads to mono-methodic researchers (and even entire fields), who use one technique like multi-level modelling again and again. Presumably the only logical way that this behaviour could be explained is that these researchers deliberately seek out problems suitable for their one technique, and deliberately ignore problems that would require them having to learn a new one' (Gorard 2002a).

Sometimes the methodological schism is supported by a philosophical 'argument' along the lines of 'but that is just positivist' (in Oakley 2001). Terms like 'positivist' referring to all numeric, or even all reasoned, approaches to research have become insults, used to dismiss opponents but never by the opponents themselves (in Fitz-Gibbon 2000). I suspect, and this can only be an opinion, that this lazy attack on scientific approaches to educational research stems more from fear and incompetence than a reasoned philosophical argument about the incompatibility of data derived by investigators with different world views. The call for a better science of education is simply for empirical evidence and reasoned argument, as opposed to prejudice and untestable opinion, and the call is in response to those people who seriously propose *unscientific* research that allows fictional drama, for example, to be treated as evidence (in Mayer 2001). As I have suggested above, combining methods does not necessarily involve a commitment to education as a natural or even a social science. Design science includes creation, artisanship, craft principles, inspiration and fuzzy science. However, at the end of the day the artefact has to work for the design to be successful. There is no room for relativism here (Bailey 2001). Either the aeroplane flies, or it does not.

The supposed incompatibility of science and social science is anyway based on a public misconception (Hammersley 2001), usually of the nature of science. Science is not really what many non-scientists tend to think it is (Collins and Pinch 1993), but an essentially anarchic, rather than rule-based, activity (Feyerabend 1993). It has no universally agreed account of its nature or method (Chalmers 1999). It offers no certainties. Yet people still booked holidays to coincide successfully with the return of Halley's comet, or the eclipse of the sun. And this is the whole point. Research findings, and the models based on them, represent a simplified description of a system that assists us with calculations and predictions. They do not represent complete truth, and are good and useful only so far as they enable us to make good decisions or improve performance (West and Harrison 1997).

A related objection to the apparent intrusion of research capacity building is to argue that it is a disciplinary project stemming from the 1996 Hargreaves lecture, intending to tell researchers what and how to research (Ball 2001). My view is that rather than fearing methodological advance, and perhaps demonising those who seek to take the field forward, we should embrace it as a sign of progress. Lakatos (1998) suggests that it is precisely the kind of methodological debate current in UK educational research that characterises an immature research field, and distinguishes it from a mature one that has already passed through the stage. So, an optimistic view would be that, via this recent crisis, educational research may be coming of age.

A more practical problem, at least in the short term, may be to gain appropriate recognition for combined approaches. Journals tend to display considerable cronyism (Travis and Collins 1991) which distort the review process. Although the two are not wholly distinct, this is more dangerous when it is cognitive (i.e. based on intellectual similarity) rather than institutional (which tends to be more visible). Papers are routinely rejected by their expert referees, perhaps for ideological reasons, or because they dispute the views of current ‘experts’ in the field (the referees). Papers and proposals are generally rejected where the reviewers disagree, since this is seen as the protection of quality. Therefore new work, radical approaches, divergent thinking, and even new techniques are likely to lead to rejection. The mechanisms of review provide short-term inequity and a longer-term encouragement of orthodoxy (Travis and Collins 1991). The worst acceptance rates are for inter-disciplinary work, and I would suspect that the same fate could befall combined methods work. Capacity building risks hostility from established groups with vested interests in maintaining the status quo. Journals, such as those that I have been on the editorial board for, tend to send papers to reviewers who are methodologically in sympathy with them. Ethnographies are more likely reviewed by ethnographers, and complex statistical models by experienced statisticians. Who, therefore, gets the combined approach papers, and if they are the traditional mono-methodic researchers will they be happy to comment on all of the paper? Rather than combining approaches the paper may be judged as falling between two stools.

Conclusion

I see no practical or philosophical objection to trying to combine methods and the datasets they generate. This paper has outlined some of the ways this is already done, and some further ways in which it may be improved. Eliciting causal mechanisms in educational research (indeed any research) is a complex business (Gorard 2002b). We could use all of the help we can get from any data source. The tentative model of combined approaches I have outlined here is based on the notion of fitness for purpose. Methods are seen as complementary rather than direct alternatives, but their appropriateness changes with the phases of a research programme. Designing an intervention, for example, might justifiably use all and any approaches including inspiration and serendipity. Developing an intervention may involve simple measuring, and in-depth monitoring. Testing an intervention may mostly involve precise measurement. Working out how to generalise from a successful intervention, or learn from an unsuccessful one, may mostly involve analysis of the 'braided' narratives of set and setting.

Notes

1. Any reduction in the probability of an outcome from 100% certainty produces a greater loss of its perceived attractiveness than an equivalent drop in probability from an originally lower figure, so that perception of probability is not a straight line function. Paradoxically, low probabilities are often greatly over-weighted in decisions, such as insuring for fire damage, or entering a lottery, but they can also be neglected entirely. One reason for this may be that risks from easily pictured causes, such as plane crashes, are more likely to be in the media than more common risks, such as diabetes, and so are exaggerated in subjective estimates. Decisions also depend on the phrasing of the problem, and the frame of reference of the subject. For example, losses loom much larger than gains, so that most people prefer to make £100 than to have a 50% chance of gaining £200, but the same people would prefer a 50% chance of losing £200 to definitely losing £100 (see also French and Smith 1997).

A further problem arises when probabilities have to be multiplied, as they would in practice in most real life decisions. Even professional mathematicians, who have correctly stated that successive tosses of a coin are independent, have been found to believe that a run of HHHHTTTT is less likely than HTTHTHTH, or that '1 2 3 4 5 6' is less likely to win the British national lottery than '3 12 17 28 44 47' (personal observation), for example. In one experiment, most respondents rated the probability of two independent events happening together as greater than one of them in isolation. For example, when told a story about a strong independent woman, she was rated by participants as being more likely to be a feminist cashier than a cashier (Eiser and van der Plight 1988), which is, of course, absurd.

Acknowledgements

This paper arose from work undertaken as part of a research capacity building exercise for the ESRC Teaching and Learning Research Programme (grant number L139251106). I would like to thank members of that project, especially Laurence Moore and Karen Roberts, for their help in forming the ideas for this paper.

References

Badger, D., Nursten, J., Williams, P. and Woodward, M. (2000) Should all literature reviews be systematic?, *Evaluation and Research in Education*, 14, 3&4, 220-230

Bailey, R. (2001) Overcoming veriphobia - learning to love truth again, *British Journal of Educational Studies*, 49, 2, 159-172

Ball, S. (2001) 'You've been NERFed!' Dumbing down the academy, *Journal of Education Policy*, 16, 3, 265-268

Bernado, J. and Smith, A. (1994) *Bayesian theory*, Chichester: John Wiley

Bridges, D. (1999) Educational research: pursuit of truth or flight into fancy?, *British Educational Research Journal*, 25, 5, 597-616

Brown, A. (1992) Design experiments: theoretical and methodological challenges in creating complex interventions in classroom settings, *The Journal of the Learning Sciences*, 2, 2, 141-178

Brown, A. and Dowling, P. (1998) *Doing research/Reading research: a mode of interogation for education*, London: Falmer

Bryman, A. (2001) *Social research methods*, Oxford: Oxford University Press

Campbell, M, Fitzpatrick, R, Haines, A, et. al. (2000) Framework for design and evaluation of complex interventions to improve health. *British Medical Journal*, 321, 694-6

Chalmers, A. (1990) *Science and its fabrication*, Milton Keynes: Open University Press

Chalmers, A. (1999) *What is this thing called science?*, Milton Keynes: Open University Press

Clegg, F. (1992) *Simple Statistics: a course book for the social sciences*, Cambridge: Cambridge University Press

Collins, A., Bielaczyc, K. and Joseph, D. (2001) *Design experiments: theoretical and methodological issues*, (mimeo)

Collins, H. and Pinch, T. (1993) *The Golem: what you should know about science*, Cambridge: Cambridge University Press

Department of Health (1997) *Research and development: Towards an Evidence-based Health Service*, London: Department of Health

Dixon-Woods, M., Fitzpatrick, R. and Roberts, K. (1999) Including qualitative research in systematic reviews, *Journal of Evaluation in Clinical Practice*, 7, 125-133

Dyson, A. and Robson, E. (1999) *School, family, community: mapping school inclusion in the UK*, Leicester: Youth Work Press

Eiser, J. and van der Plight, J. (1988) *Attitudes and Decisions*, London: Routledge

Ellmore, P. and Woehilke, P. (1998) *Twenty years of research methods employed in American Educational Research Journal, Education Researcher, and Review of Educational Research from 1978 to 1997*, (mimeo) ERIC ED 420701

Evans, J. and Benefield, P. (2001) Systematic reviews of educational research: does the medical model fit?, *British Educational Research Journal*, 27, 5, 527-542
Feyerabend, P. (1993) *Against method*, London: Verso
Fitz, J., Taylor, C., Gorard, S. and White, P. (2002) Local education authorities and the regulation of educational markets: four case studies, *Research Papers in Education* (forthcoming)

Fitz-Gibbon, C. (2000) Education: realising the potential, in Davies, H., Nutley, S. and Smith, P. (Eds.) *What works? Evidence-based policy and practice in public services*, Bristol: Policy Press

Fitz-Gibbon, C. (2001) What's all this about 'evidence'?, *Learning and Skills Research*, 5, 1, 27-29

Frazer, E. (1995) What's new in the philosophy of science?, *Oxford Review of Education*, 21, 3, 267

French, S. and Smith, J. (1997) Bayesian analysis, pp.1-25 in French, S. and Smith, J. (Eds.) *The practice of Bayesian analysis*, London: Arnold
Furlong, J. and White, P. (2001) *Educational research capacity in Wales: UCET/BERA report*, Cardiff University School of Social Sciences
Gephart, R. (1988) *Ethnostatistics: Qualitative foundations for quantitative research*, London: Sage
Glymour, C., Scheines, R., Spirtes, P. and Kelly, K. (1987) *Discovering causal structure*, Orlando: Academic Press
Gorard, S. (1997) *School Choice in an Established Market*, Aldershot: Ashgate
Gorard, S. (1998) *The call for a middle-way in educational research*, BERA Internet Conference, March 1998, www.bera.ac.uk/debate/reply/g1.html
Gorard, S. (1999) 'Well. That about wraps it up for school choice research': a state of the art review, *School Leadership and Management*, 19, 18, 25-47
Gorard, S. (2000) *Education and Social Justice*, Cardiff: University of Wales Press
Gorard, S. (2001a) *A changing climate for educational research? The role of research capability-building*, presentation to BERA September 2001, Leeds, also available on Education-line

Gorard, S. (2001b) *Quantitative Methods in Educational Research: The role of numbers made easy*, London: Continuum
Gorard, S. (2002a) When should we use multi-level modelling?, *British Educational Research Journal*, (submitted)
Gorard, S. (2002b) The role of causal models in education as a social science?, *Evaluation and Research in Education* (forthcoming)
Gorard, S. and Rees, G. (2002) *Creating a learning society*, Bristol: Policy Press
Gorard, S. and Taylor, C. (2002) Diversity or hierarchy? The role of school allocation procedures in maintaining equity, in *Handbook of Educational Management*, London: Pearson
Gorard, S., Rees, G. and Fevre, R. (1999a) Patterns of participation in lifelong learning: do families make a difference?, *British Educational Research Journal*, 25, 4, 517-532
Gorard, S., Salisbury, J. and Rees, G. (1999b) Reappraising the apparent underachievement of boys at school, *Gender and Education*, 11, 4, 441-454
Gorard, S., Rees, G., Fevre, R. and Welland, T. (2001) Lifelong learning trajectories: some voices of those in transit, *International Journal of Lifelong Education* , 20, 3, 169-187
Hakuta, K. (2000) *Perspectives on the state of education research in the US*, presentation at AERA, New Orleans, April 2000

Hammersley, M. (1997) Educational research and teaching: a response to David Hargreaves' TTA lecture, *British Educational Research Journal*, 23, 2, 141-162

Hammersley, M. (2001) *Some questions about Evidence-based Practice in Education*, presentation to BERA, Leeds, 13-15th September 2001

Hartigan, J. (1983) *Bayes theory*, New York: Springer-Verlag

Hausman, C. (2000) *Reviewing the review process: what can we learn?,* presentation at AERA, New Orleans, April 2000

Hayes, E. (1992) The impact of feminism on adult education publications: an analysis of British and American Journals, *International Journal of Lifelong Education*, 11, 2, 125-138

Hillage, J., Pearson, R., Anderson, A. and Tamkin, P. (1998) *Excellence on research in schools*, Sudbury: DfEE

Kuhn, T. (1970) *The structure of scientific revolutions*, Chicago: University of Chicago Press

Lakatos, I. (1978) *The methodology of scientific research programmes*, Cambridge: Cambridge University Press

Levacic, R. and Glatter, R. (2001) Really good ideas? Developing evidence-informed policy and practice in educational leadership and management, *Educational Management and Administration*, 29,1, 5-25

Mahoney, J. (2000) Strategies of causal inference in small-N analysis, *Sociological Methods and Research*, 28, 4, 387-424

Marshall, G. (2001) *Social Sciences*

Mayer, R. (2001) Resisting the assault on science: the case for evidence-based reasoning in educational research, *Educational Researcher*, 30, 7, 29-30

McIntyre, D. and McIntyre, A. (2000) *Capacity for research into teaching and learning*, Report to TLRP

Medical Research Council (2000) *A framework for development and evaluation of RCTs for complex interventions to improve health*, London: MRC.

Millett, A. (1997) *Speech to TTA Research Conference*, 5 December 1997, London: TTA

Moore, L. (2002) Lessons from using randomised trials in health promotion, *Building Research Capacity*, 1, 1, 4-5

Morrison, K. (2001) Randomised controlled trials for evidence-based education: some problems in judging 'what works', *Evaluation and Research in Education*, 15, 2, 69-83

Mortimore, P. (2000) Does educational research matter?, *British Educational Research Journal*, 26, 1, 5-24

National Research Council (1999) *Improving student learning: a strategic plan for educational research and its utilization*, Washington DC: National Academy Press

NERF (2001) *A research and development strategy for education: developing quality and diversity*, Nottingham: National Educational Research Forum

Nutbeam, D, Macaskill, P, Smith, C, et. al. (1993) Evaluation of two school smoking programmes under normal classroom conditions, *British Medical Journal*, 306, 102-107

Oakley, A. (2001) Making evidence-based practice educational: a rejoinder to John Elliott, *British Educational Research Journal*, 27, 5, 575-576

Oppenheim, A. (1992) *Questionnaire design, interviewing and attitude measurement*, London: Continuum

Pedhazur, E. (1982) *Multiple Regression in Behavioural Research*, London: Holt, Rhinehart and Winston

Popkewitz, T. (1984) *Paradigm and Ideology in Educational Research*, Lewes: Falmer Press

Pring, R. (2000a) *Philosophy of educational research*, London: Continuum

Pring, R. (2000b) Editorial conclusion: a philosophical perspective, *Oxford Review of Education*, 26, 3&4, 495-501
Research Fortnight (2001) Crisis in quality for evidence research, *Research Fortnight*, 21 November 2001, p.6
Resnick, L. (2000) *Strengthening the capacity of the research system: a report of the National Academy of Education*, presentation at AERA, New Orleans, April 2000.
Roberts, I. (2000) Randomised trials or the test of time?: The story of human albumin administration, *Evaluation and Research in Education*, 14, 3&4, 231-236
Roberts, K., Dixon-Woods, M., Jones, D., Fitzpatrick, R. and Abrams, K. (1999) *Meta-analysis of qualitative and quantitative evidence: an example based on uptake of childhood immunisation*, presentation at VII Cochrane Colloquium, Rome, October 1999
Schorr, R. and Firestone, W. (2001) *Changing mathematics teaching in response to a state testing program*, presentation at AERA Annual Conference, Seattle WA, April 2001
Snow, C. (2001) Knowing what we know: children, teachers, researchers, *Educational Researcher*, 30, 7, 3-9
Speller, V., Learmonth, A. and Harrison, D. (1997) The search for evidence of effective health promotion, *British Medical Journal*, 315, 361-363
Spillane, J. and Zeuli, J. (1999) Reform and teaching: exploring patterns of practice in the context of national and state mathematics reforms, *Educational Evaluation and Policy Analysis*, 21, 1, 1-28
Suter, L. and Frechtling, J. (2000) *Guiding principles for mathematics and science education research methods*, Arlington VA: National Science Foundation
Taylor, E. (2001) From 1989 to 1999: A content analysis of all submissions, *Adult Education Quarterly*, 51, 4, 322-340
Tooley, J. and Darby, D. (1998) *Educational research: a critique*, London: OFSTED
Torgerson, C. and Torgerson, D. (2001) The need for randomised controlled trials in educational research, *British Journal of Educational Studies*, 49, 3, 316-328
Travis, G. and Collins, H. (1991) New light on old boys: cognitive and institutional particularism in the peer-review system, *Science, Technology and Human Values*, 16, 3, 322-341
West, M. and Harrison, J. (1997) *Bayesian forecasting and dynamic models*, New York: Springer
White, P., Gorard, S., Fitz, J. and Taylor, C. (2001) Regional and local differences in admission arrangements for schools, *Oxford Review of Education*, 27, 3, 317-337
Woodhead, C. (1998) *Academia gone to seed, New Statesman*, 26 March 1998, pp. 51-52