ESRC Teaching and Learning Research Programme
Research Capacity Building Network

Occasional Paper Series
Paper 48

Warranting research claims from non-experimental evidence

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

Warranting research claims from non-experimental evidence

Abstract

This paper contains a preliminary consideration of the nature and role of warrants for research conclusions in educational research. It focuses on warrants for conclusions drawn from findings based on non-experimental evidence. The paper describes some of the standard non-experimental designs, and their need for a warrant in the form of a logical and persuasive link between the evidence produced and the conclusions drawn (with appropriate qualifications and caveats). It describes social scientific warrants, and suggests that the elimination of plausible rival alternative explanations for the evidence is a key approach (and one that is independent of the method used). The paper then briefly discusses the nature of warrants as used by practitioners and policy-makers, and examines some objections to the 'scientific' basis of warranted practice. It concludes by suggesting that the much publicised criticisms of the quality and relevance of educational research may be, at least partly, misdirected. What critics may have objected to is not so much poor research, but inappropriately warranted conclusions (i.e. overclaiming).

Introduction

This paper considers ways in which non-experimental research designs can produce warrantable claims to knowledge, particularly of the causal mechanisms underlying observed phenomena. It therefore builds directly upon previous work considering the nature of cause:effect models (Gorard 2002a), and the role of experimental designs in educational research (Gorard 2001a). The paper starts from the premises that: causal models (in the widest sense) are important; that they underlie (at least implicitly) most significant research findings; that they are not open to direct inspection and therefore provide the key role for data-driven theory, and finally that (despite their clear difficulties) experimental approaches are especially valuable in testing cause and effect models.

The paper starts by summarising the nature and role of experiments in educational research, before outlining a variety of other approaches to providing relatively robust research claims. These include structural equation modelling, subjective expected utility, historical approaches, mental experiments and, above all, combining methodological approaches (see Gorard 2002b). The paper also examines some objections to the above suggestions. The conclusion is that the methods used, and even the quality of the findings generated are largely irrelevant to their perceived warrant. This is so for two reasons. First: many of the high-profile criticisms of educational research are not, on reflection, about the nature of the evidence produced but about the way in which it is presented as unwarranted conclusions (generalising from non-representative samples and so on). A piece of evidence cannot be either good or bad as long as it is presented with its appropriate caveats. It is when the researcher, or others, seek to go beyond what that evidence entails that problems occur (e.g. when there is overclaiming). Second: perhaps because of this, users and consumers of research largely do not worry about the relationship between evidence and conclusions, but use warrants of their own devising. Therefore the actual impact of research findings is weak, and apparently not related to research quality (e.g.

findings may be used because they are convenient, or as rhetorical justification for an existing position, rather than warranted).

Experimental evidence

Writers have dealt elsewhere with the relative merits of experimental evidence in educational research (e.g. Moore 2002, Gorard 2001b). Only a very brief summary of the position is presented here, for the chief purpose of this paper is to consider work that, for whatever reason, is not experimentally based.

In many ways the experiment is seen as the 'flagship' or gold standard of research designs. The basic advantage of this approach over any other is its more convincing claim to be testing for cause and effect, via the manipulation of otherwise identical groups, rather than simply observing an unspecified relationship between two variables. In addition, some experiments will allow the size of any effect to be measured. It has been argued that only experiments are thus able to produce secure and uncontested knowledge about the truth of propositions. Their design is flexible, allowing for any number of different groups and variables, and the outcome measures taken can be of any kind although they are normally converted to a coded numeric form. The design is actually so powerful that it requires smaller numbers of participants as a minimum than would be normal in a survey for example. The analysis of the results is also generally easier than when using other designs. Educational research has, for too long, relied on fancy statistical manipulation of poor datasets, rather than such well designed studies (FitzGibbon 1996, 2001). When subjected to a definitive trial, many interventions and treatments actually show no effect, identifying wasted resources on policies and practices. Perhaps that is also partly why there is considerable resistance to the idea of increased use of experimental evidence in education. Social work was one of the areas where natural experiments were pioneered but, when these seldom showed any positive impact from social work policies, social workers rejected the test as a method rather than the ineffective practices (Torgerson and Torgerson 2001). Those with vested interests in current educational practice may, similarly, have little to gain from definitive trials.

However, without wishing the exaggerate the difficulties, it is the case that experimental designs are not always possible - for practical and ethical reasons, because of the nature of the research questions, or because the dataset to be considered already exists. Much of what we know is not experimental, and what we want to know cannot be (Glymour et al. 1987). Trials by themselves (although never intended to be used in isolation, see Gorard 2002b) are also unlikely to lead to an understanding of detailed causal mechanisms (Morrison 2001). Their simplicity, which is part of their appeal, might lead to concentration on one 'effect' but not pick up multiple side-effects (including possibly deleterious ones). What are the alternatives?

Alternative forms of evidence

If we assume for the purpose of this paper that the description of experimental approaches above includes the rigorous evaluation of any engineered intervention whether in a laboratory or in a real-life setting, then the alternatives are based on what Fitz-Gibbon (2000) calls essentially 'passive' approaches to research. The key difference is that in the former design the researcher introduces a change into an environment and monitors subsequent events for the potential impact

of that change. In the latter design the researcher simply monitors events and attempts to track back to a 'cause' *post hoc* - a much more difficult task both conceptually and technically (in terms of compensatory statistical analysis, for example). Ironically, it is not always easier to judge with the benefit of hindsight.

Nevertheless the model of an experiment can still be of assistance, even in passive designs. First: we can use a natural experiment model (Gorard 2001a). A passive approach can include the recognition that there are naturally occurring interventions going on around us all of the time. The subsequent monitoring phase can be attached to an intervention that is not controlled by the researcher, but using the same analytical methods as for an experiment, sometimes long after the intervention (e.g. Gorard et al. 2001). Natural experiments are possible, when interventions occur as part of the normal policy process. If one local education authority changes its practice in some way then it can be construed as an experimental group and the remaining authorities as controls in a natural experiment. In fact, much social science research is of this type - retrospectively trying to explain differences between two groups. This is inferior in terms of validity to a true experimental design but much more practical. Knowing how an experiment works is important because it enables us to see how far a natural experiment is from that 'ideal'. But it also alerts us to the need for things like comparison or control groups, and it is alarming how often passive researchers attempt to make comparisons over time and place on the basis of one observation (and even more alarmingly are believed and cited favourably by others, see Budge 2000 for example). It also alerts us to the need for a transparent written protocol, so that our findings can be replicated just like those of a real experiment (Moses 2001).

Another way in which experimental models retain relevance in passive research is via the invaluable 'thought experiment'. Knowing the format and power of experiments gives us a template against which to measure what we do instead, and even helps us to design what we do better. In a thought experiment, we can freely consider how to gain secure and uncontested knowledge about the truth of our propositions, without any concern about practical or ethical considerations. This becomes an ideal, and it helps us to recognise the limitations of our actual approach. It therefore provides qualifications for our warrant (see below).

A third way in which experimental designs can contribute is via 'mental experiments' (Miles and Shevlin 2001). These are used in statistical modelling (often erroneously referred to as causal modelling) to help determine plausible directions of causation in our explanations. Structural equation modelling allows models based on non-experimental evidence to be rejected as inconsistent or to be tentatively retained as an explanation (Maruyama 1998). Such approaches cannot make up for poor design/data, and the models they generate should not be based on statistical criteria alone (Kline 1998). This is why the mental experiment is valuable, allowing the researcher to specify a model or models in advance of analysis (and preferably before collecting the data). In the model the 'causal' paths are specified in advance in order to avoid the charge of dredging (simply looking for anything 'significant' post hoc) or of both building and testing a model on the same dataset, or adopting statistics as a pseudo-science. (Glymour et al. 1987). Of course the dataset can never confirm the model, so the key issue here is to consider (and presumably eliminate) all plausible rival explanations and to work towards the most parsimonious version of the model. The same applies to all other forms of statistical modelling, whatever they are called, since despite differences in classification there is no real difference between all forms of analysis of non-experimental work (Johnson 2001, Gorard 2002c).

The most common passive alternative to all of the above is based on narrative (whether textual, audio, or video-based), and derived from the well-established traditions of historical research (note that where narrative approaches are used in intervention studies they are not passive, but perhaps action research or design experiments, see Gorard 2002b). In the passive approach, processes and outcomes are described, often in complex ways such as 'braided' accounts. The essential problem with this approach is that it is difficult to convince a sceptical audience that the account given is the best or 'truest'. In other words, it is more difficult to warrant the conclusions. Of course, this problem is shared with the other passive approaches, above, but to a lesser extent since they are (con)testable. Structural equation modelling, for example, can easily produce a finding which leads to the rejection of the prior model. Unpicking and understanding a complex narrative, on the other hand, does not lead to the same situation so easily. The problem is also shared with disciplines like history (archaeology, palaeontology, astronomy etc.), but the difference here is that history (like the others) is *constrained* to be non-experimental and is, in effect, making the best of what is possible. Historians also try to integrate other material with their narratives, including genetic, linguistic and archaeological evidence. Educational research has no such constraint about experiments (although it applies to *some* educational research questions).

One possible solution is to treat all such narrative approaches as working towards an explanation that can then be (con)tested (by converting to an intervention study on the next project iteration). But this is seldom done, perhaps because of the dominant mono-method culture in UK educational research, or perhaps because it is cosier not to have to test our 'knowledge' in a relatively brutal way as we would have to if we were designing a helicopter. Another solution would be to base an explanation on several studies, all using different passive approaches, and each of which may show part of the 'signature' of the same 'cause'. For these kinds of studies, perhaps, the need for an explicit warrant is greater than those based around the logic of an intervention (whether researcher-generated or not). Conclusions may also be in need of particular warranting when they are counter-intuitive, when their warrant is unusual, or where the warrant itself challenges the accepted knowledge within a field.

What is a warrant?

Humes and Bryce (2001) cite the Scottish Minister for Education in 2000 who feels that the difference between social and natural science is important, but that it is too often used as an excuse for lack of simple rigour and an over-emphasis on value-judgements. The minister is generally unimpressed by research papers, largely because, in his opinion, the conclusions are often not based on the findings (i.e. the research simply provides a rhetorical backdrop for the description of previously held opinions). If this were so, what would be missing in such accounts is a warrant - the crucial link between the findings and the conclusions ostensibly drawn from them.

Consider as a very simple example a study by Waslander and Thrupp (1995) which presents, among others, the following table. It uses a relatively standard measure of occupational class ranging from 1 for highly prestigious jobs to 5 for less prestigious jobs. Therefore a 'low' number on this scale represents a 'high' score for socio-economic status (SES). Using these figures, Waslander and Thrupp conclude that the intake to 'adjacent' schools (second row in Table 1) is of higher SES after 1990 than it is in 1990 - and from this they argue that dezoning in New Zealand

has disadvantaged working-class families (for more on this and other errors in their findings see Gorard 2000). I use this example here as an extreme case to introduce a point about warrants from evidence. There is no suggestion, and I have never encountered any suggestion, that the figures or analysis presented by Waslander and Thrupp are not accurate. The research itself appears to be of high quality (and imaginatively original to boot). The problem lies in the relationship between the findings (as exemplified by Table 1) and the authors' published conclusions. Put simply, the evidence they present does not support the conclusions they draw. Their conclusions are *not* warranted.

Table 1 - Mean SES of students by locality of school (Table 4 in Waslander and Thrupp 1995)

Locality\Year	1990	1991	1992	1993
Local school	3.20	3.22	3.26	3.19
Adjacent school	3.02	3.27	3.40	3.36

Of course, this is a very simple, and obvious, example. More usually, the warrant of an argument can be considered to be its general principle - an assumption that links the evidence to the claim made from it (Booth et al. 1995). In logic, the simplest standard syllogism takes the form:

All B is C.
This A is B.
Therefore A is C.

While the second part may be likened to the evidence in a research study, and the third is the conclusion, the first is the like the warrant. In research this step is often missed as it is tacitly assumed by the author and the reader. However, where the research is intended to change the views of others it may be necessary to make it explicit. The warrant has a syntax of: when we have evidence like X, we can make a claim like Y. It can be challenged, but unlike a challenge to the evidence it is not about quality but rather about the *relevance* of the evidence to the conclusion.

Only a clear and robust warrant, along with high-quality and relevant research, provides the necessary foundation for changes in evidence-informed policy (or practice), and then ensuring that the proclaimed benefits of change actually arise. At heart a warrant for change contains a causal claim (Gorard 2002a), which states that if the practitioner (policy-maker) does one thing then another will ensue. The warrant may be part of the research design (see above) but is independent of any particular method of data collection (de Vaus 2001). Methods cannot be judged in isolation from the questions they are intended to illuminate (Shavelson and Towne 2001). The results should be disclosed to critique, and the conclusions drawn based on an explicit coherent chain of reasoning which rules out all plausible counter-explanations, and is intended to be persuasive to sceptical reader (rather than playing to a gallery of existing 'converts', for example).

The boxing off of plausible rival explanations is generally at the heart of effective warrants. For any real system of variables there are nearly infinite models that could explain them (Glymour et al. 1987), and therefore no one can consider them all - so that in social science, as in natural science, every law that is ever proposed is quite literally false (but can still be useful, Gorard 2002c). In the same way an infinite number of equations can join any two points on a graph. The purpose of the warrant is show readers that the proposed explanation is the best at this point in

time. A useful short-cut is to employ parsimony to eliminate many of the potential alternatives (cf. Morgan 1903). It is, for example, simpler, and usually safer for a doctor to diagnose a complaint of headache, neck stiffness, fever and confusion as meningitis, rather than as a combination of brain tumour, whiplash, tuberculosis and acute poryphyria. Of course the latter could be correct, but parsimony encourages us to eliminate the more mundane and simplest explanations first. We therefore limit our potential explanations to those that employ the fewest (ideally none) assumptions for which we have no direct evidence.

Alternative warrants for practice

Of course, the 'persuasiveness [of findings] may require more than simply strong research design... the potential for research to contribute to practice depends on its ability to influence teachers' thinking' (Kennedy 1997, p.7). In order to be effective, social science knowledge must be appropriately packaged and mediated by practitioners so that they can 'make it their own'. A key question is 'How can the use of research knowledge be increased in schools and school districts?' (NRC 1999, p. 2). As an academic community we may have several excuses for the difficulties and complexities encountered in our research but we have fewer for our weaknesses in converting our findings into usable formats. Probably no other public sphere rests on such a slight research base, with personal experience and ideology so commonly used in policy formation (NRC 1999). Research impact stems partly from the quality (and therefore the believability) of the findings and partly from the desire and willingness of practitioners to use research as a basis for professional change. The practical use of evidence is never likely to be based solely on quality (Hodkinson 2001).

The kinds of warrants used by teachers and policy-makers may differ both from each other (Lewis 2001), and from those used by researchers as outlined above (McNamara and Corbin 2001). For example, teachers largely ignore 'evidence' in forming practice, but use a variety of other warrants (such as student reaction in the classroom). 'The issue of teachers engaging *with* research as opposed to in research has been widely neglected' (p.264). Teachers are influenced by a range of kinds of evidence from specific context knowledge (personal and informal), through established practices and resources that embody knowledge (tacit and invisible), and case studies of others' accounts, to general research knowledge. The latter often appears to add little to their practice, due to its apparent lack of immediate relevance (Ratcliffe et al. 2001). Perhaps one way to overcome this lack of engagement is for potential users to be involved from beginning, to encourage their 'ownership' of any project, and for them to play a role in generalising the findings (Lewis 2001).

Interactive Social Science (ISS) attempts to involve users of research throughout the project life-cycle. It highlights the significance of developing a research project/programme for a particular user group. It is intended to be pragmatic and user-oriented, and just-for-you rather than just-in-case. In health circles people are talking of researching towards policy (or practice) based evidence, rather than simply calling for evidence based policy (or practice). It has also been called 'context-sensitive' science (Gibbons 2000), or working in Pasteur's Quadrant (Stokes 1997). It is directed more towards the relevance, impact and application of findings, rather than their quality or rigour *per se*. In education this may be less of a problem, since many researchers traditionally are, or recently have been, practitioners as well. In fact it can claim to be an obvious field in which academics, as lecturers, are also routinely practitioners in their area of study. The

ISS model may therefore be more important for disciplines like economics, geography and above all sociology where the issue of everyday relevance is often far less clear. Nevertheless it provides an interesting way forward for consideration.

These considerations might include what the evidence-base for ISS is. For, like evidence-based policy itself, it has been proposed by advocates as a generally good idea but without either the theoretical or empirical basis that would normally be required for such a general proposition in social science. Another consideration should be the relationship between quality and relevance. The ISS model assumes that the problems we face in educational research are largely ones of engineering. We have, it would argue, safe knowledge about teaching and learning for example, and the capacity to generate more if we require (i.e. the research is good enough). What we need to start doing better is making it count in the 'real world'. However, others could argue that the ISS model leads to reduced quality in social science (in Rappert 1997), perhaps by passing too much control to 'client' groups. What little analysis that has been done actually points away from ISS (e.g. Tooley and Darby 1998). Educational research has been heavily criticised for the poor quality of its research, and not for its relevance, and there is no reason suppose that ISS, of itself, can improve quality, while it could lead to the charge of creating evidence on demand (Gorard 2002d).

Objections to scientific approaches

Although the relevance of educational research has been called into question (Hillage et al. 1998) it is generally issues of quality that have attracted greater attention (Tooley and Darby 1998) and this have been used to provide pressure for greater political influence. Strategies for 'packaging' results and for dissemination, to aid the successful use of research findings, are bound to fail if those findings are deemed somehow not trustworthy. It can be argued that genuine improvements in practice and policy are more likely to be based on good social science than on 'craft principles'. Good social science will generally reflect scientific principles and rigorous standards and share scientific norms such as explicit hypotheses, sound designs, appropriate measures, quality data, and logical analyses (NERPP 2000). In the long term these are also likely to be the criteria for believable and usable results (Bridges 1999).

At heart any kind of science is the same in all fields (Shavelson and Towne 2001). It requires rigorous reasoning supported by mixed methods, and findings leading to testable models or theories. Despite the fact that it is portrayed by outsiders and opponents as the mechanistic application of pre-determined procedures, progress is actually achieved via the self-regulating norms of the scientific community. The key point is that a specific design or method does not make a study scientific (but only if it allows *direct* investigation of the question being asked). Shavelson and Towne (2001) suggest that for success in education as a field of scientific enquiry we need a scientific culture, skilled staff, ample peers, collaborative creation of the research agenda, insulation from political short-term interference, a mix of short-term and long-term objectives, adequate funding, and high levels of investment in training.

Of course, there will be commentators who oppose the call for greater rigour because of the term 'science' (Gorard 2001b). But the call for more scientific approaches is simply for more empirical evidence and reasoned argument (i.e. better warrants) versus opinion and ideology (Mayer 2001). However, for some the whole enterprise is likely to be condemned in a single word - 'positivist'.

'Nowadays the term "positivist" is widely used as a generalized term of abuse' (Phillips 1992, p.95), but other than that it signifies very little. One key difference between scientific research and other endeavours is that the results of the former can be 'accepted as true'. It would be a category mistake to say that some research descriptions are not meant to be true, else why should they believed to have warranted assertability (Gorard 2000). Multiple perspectives do not mean the end of truth as an ideal. We could, for example, view one classroom in terms of its efficiency, economy, heating, and lighting etc. Each account so generated may be true, but also orthogonal. We cannot, because of this, seriously assert that *anything* must be true. In fact the researcher who claims not to be scientific may merely be insufficiently aware of the basis of their own approach - there are many examples of social scientists who claim to be relativists, for example, while behaving with respect to the ideas of others as nothing of the sort. 'So unconscious is the average social scientist... of the gnoseological presuppositions of his [*sic*] study that he finds it only to easy to avow allegiance to doctrines wholly at variance with the philosophical pre-requisites of his own researches... intellectual fashions are made up of avowed philosophies and not assumed ones' (Postan 1971, p.ix).

Conclusion

Research itself is easy. Everyone (even an infant) does it every day by gathering information to answer a question and so solve a problem (e.g. to plan a rail journey, Booth et al. 1995). In fact most of what we 'know' is research-based, but reliant on the research of others (such as the existence of Antarctica). Where we have no other choice we may rely on our judgement of the *source* of that information (an atlas may be more reliable than memory, the rail enquiries desk may be more reliable than last year's timetable). Where we have access to the research findings on which any conclusions are based we can also examine their quality and the warrant that connects the two. Similarly when we present our own research findings, we need to give some indication, via caveats, of the extent to which we would be prepared to bet on them being true, or the extent to which we would wish others to rely on them being true. This is part of our 'warrant'. Obviously, high quality research is important but even high quality work can lead to inappropriate conclusions (as in the example of Waslander and Thrupp 1995, above). Conversely, poorer quality work (in an un-researched area perhaps) need not be dismissed as long as its conclusions are appropriate to the level of security of the findings. There are therefore no real rules about research methods (and certainly no best methods), and being a little cavalier in producing results may be fine as long as one is then a little conservative when presenting them. Perhaps the much-vaunted problems in educational research have arisen not so much from poor use of methods as from inappropriately (or perhaps only implicitly) warranted conclusions.

Acknowledgements

This paper is based on a presentation made to team leaders of the ESRC Teaching and Learning Programme, London, January 2002 (grant number L139251106).

References

Booth, W., Colomb, G. and Williams, J. (1995) *The craft of research*, Chicago: University of Chicago Press

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

Budge, D. (2000) Pupils shoulder burden of proving their self-worth, *Times Educational Supplement*, 23/6/00, p.28

de Vaus, D. (2001) *Research design in social research*, London: Sage

Fitz-Gibbon, C. (1996) *Monitoring education: indicators, quality and effectiveness*, London: Cassell

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) *Value-added for those in despair: research methods matter*, British Psychological Society

Gibbons, M. (2000) Mode 2 society and the emergence of context-sensitive science, *Science and Public Policy*, 27, 3

Glymour, C., Scheines, R., Spirtes, P. and Kelly, K. (1987) *Discovering causal structure*, Orlando: Academic Press

Gorard, S. (2000) *Education and Social Justice*, Cardiff: University of Wales Press

Gorard, S. (2001a) *Quantitative Methods in Educational Research: The role of numbers made easy*, London: Continuum

Gorard, S. (2001b) *A changing climate for educational research*, *Occasional Paper 45*, Cardiff: School of Social Sciences, ISBN 187233072X, 43 pages

Gorard, S. (2002a) The role of causal models in education as a social science?, *Evaluation and Research in Education* (forthcoming)

Gorard, S. (2002b) Can we overcome the methodological schism?: combining qualitative and quantitative methods, *Research Papers in Education*, (forthcoming)

Gorard, S. (2002c) When should we use multi-level modelling?, *British Educational Research Journal*, (submitted)

Gorard, S. (2002d) Political control: A way forward for educational research?, *British Journal of Educational Studies* , (forthcoming)

Gorard, S., Fitz, J. and Taylor, C. (2001) School choice impacts: what do we know?, *Educational Researcher*, 30, 7, 18-23

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

Hodkinson, P. (2001) How can research influence practice?, *Learning and Skills Research*, 5, 1, 15-16

Humes, W. and Bryce, T. (2001) Scholarship, research and the evidential basis of policy development in education, *British Journal of Educational Studies*, 49, 3, 329-352

Johnson, B. (2001) Towards a new classification of nonexperimental quantitative research, *Educational Researcher*, 30, 2, 3-14

Kennedy, M. (1997) The Connection Between Research and Practice, *Educational Researcher*, October 1997, pp. 4-12

Kline, R. (1998) *Principles and practice of structural equation modelling*, New York: Guilford Press

Lewis, J. (2001) Getting social science research into praactice, *Learning and Skills Research*, 5, 1, 17-18

Maruyama, G. (1998) *Basics of structural equation modelling*, London: Sage

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

McNamara, O. and Corbin, B. (2001) Warranting practices: teachers embedding the national numeracy strategy, *British Journal of Educational Studies*, 49, 3, 260-284

Miles, J. and Shevlin, M. (2001) *Applying regression and correlation*, London: Sage

Moore, L. (2002) Lessons from using randomised trials in health promotion, *Building Research Capacity*, 1, 2-3

Morgan, C. L. (1903) *Introduction to comparative psychology*, London: Walter Scott

Morrison, K. (2001) Randomised controlled trials for evidence-based education: some problems in judging 'what works', *Evaluation and Research in Education*, 15, 2, 69-8

Moses, L. (2001) *A larger role for randomized experiments in educational policy research*, presentation at AERA annual conference, Seattle, April 2001

National Educational Research Policy and Priorities Board (2000) *Second policy statement with recommendations on research in education*, Washington DC: NERPP

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

Phillips, D. (1992) *The social scientist's bestiary*, Oxford: Pergamon Press

Postan, M. (1971) *Fact and relevance: essays on historical method*, Cambridge: Cambridge University Press

Ratcliffe, M., Bartholomew, H., Hames, V., Hind, A., Leach, J., Millar, R. and Osborne, J. (2001) *Understanding the researcher-user interface*, presentation at BERA annual conference, Leeds, September 2001

Shavelson, R. and Towne, L. (2001) *Scientific Inquiry in Education*, report of the National Research Council, Washington: National Academy Press

Stokes, D. (1997) *Pasteur's Quadrant: Basic Science and Technological Innovation*, Washington DC: Brookings Institution Press

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

Waslander, S. & Thrupp, M. (1995) Choice, competition, and segregation: An empirical analysis of a New Zealand secondary school market, 1990-93, *Journal of Education Policy,* 10, 1., 1-26