

This article was downloaded by: [University of Warwick]

On: 8 June 2011

Access details: Access Details: [subscription number 773572776]

Publisher Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Evaluation & Research in Education

Publication details, including instructions for authors and subscription information:
<http://www.informaworld.com/smpp/title~content=t794297791>

Fostering Scepticism: The Importance of Warranting Claims

Stephen Gorard

Online publication date: 29 March 2010

To cite this Article Gorard, Stephen(2002) 'Fostering Scepticism: The Importance of Warranting Claims', Evaluation & Research in Education, 16: 3, 136 – 149

To link to this Article: DOI: 10.1080/09500790208667014

URL: <http://dx.doi.org/10.1080/09500790208667014>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Fostering Scepticism: The Importance of Warranting Claims

Stephen Gorard

Cardiff University School of Social Sciences, Cardiff, Wales, UK

This paper contains a consideration of the nature and role of warrants for research conclusions in educational research. The paper argues the need for an explicit 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, some problems arising in real-life research, and the nature of warrants as used by practitioners and policy-makers. It examines some objections to the 'scientific' basis of warranted practice. Overall, the paper argues that greater transparency, complete specifications of the logic, and the elimination of plausible rival alternative explanations for the evidence are key approaches (and ones that are independent of the method used to derive the evidence).

Keywords: warrant, scepticism, argument, evidence-based

What is a Warrant?

Research itself is quite 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). But where we have access to the research evidence 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, producing high quality research is important but even high quality work can lead to inappropriate conclusions (Toulmin, 1958).

Huck and Sandler (1979) remind readers of a humorous example in order to make an important point about warrants. An experimental psychologist trains a flea to jump in response to hearing a noise. Every time the noise is made, the flea jumps. They then cut the legs off the flea, and discover that it no longer jumps when the noise is made. Conclusion: cutting off the legs has affected the flea's hearing. Of course, this is clearly nonsense but it is likely that we have all been persuaded by similar conclusions. If a physiologist cuts out a piece of someone's brain, and the person can no longer tell us about a memory (or perform a skilled action) that they were able to previously, then is this evidence that the specific memory or skill was 'stored' in that section

of brain? Many such claims have been made, and early maps of brain function were based on just this approach. However, the same effect of inability to report recall of memory (or skill) could have been achieved by cutting peoples' tongue out, or removing their heart. All three operations may prevent memory recall for different reasons without showing that the part of the body removed in each case is the site of the memory. What is needed, in addition, is an argument leading from the evidence to the conclusion. This would be the warrant for those findings.

Brignell (2000) provides another example. The chemical industry routinely uses a chemical called 'dihydrogen monoxide'. While tremendously useful, this chemical often leads to spillages, and finds its way into our food supply. It is a major component of acid rain, and a cause of soil erosion. As a vapour, it is a major greenhouse gas. It is often fatal when inhaled, and is a primary cause of death in several UK accidents per year. It has been found in the tumours of terminally ill patients. What should we do about it? In a survey the clear majority of respondents believed that water, for that is what it is, should be either banned or severely regulated. All of those statements about water are basically 'true', yet clearly none of them mean that water should be banned. Replace water with another, less abundant, chemical. How do we feel about banning it now? We have no obvious reason to change our mind. Yet we will all probably have accepted just such evidence as we have about water in order to accept the banning of other chemicals. This shows how difficult, but also how important, the warrants for research conclusions are. In both the flea and the water example, the problem was not principally the research quality (or put another way the problem was separate from any reservations we may have about the quality of the evidence). The problem was that the conclusions drawn were not logically entailed by the research evidence itself.

'Reasoning ... is a way of testing and sifting ideas critically. It is concerned with how people share their ideas and thoughts in situations that raise the question of whether those ideas are worth sharing' (Toulmin *et al.*, 1979: 10). On a weak interpretation (also see below), a warrant is the form in which 'people furnish rationales as to why a certain voice ... is to be granted superiority ... on the grounds of specified criteria' (Gergen, 1989: 74). Perhaps, for the purpose of this paper, a warrant is more simply summarised as that 'which makes the difference between knowledge and mere true belief' (Plantinga, 1993a: 3). 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). Claims must be substantive, specific and contestable. The evidence on which they are based ought to be precise, sufficient, representative, authoritative and clear to the reader (as far as possible). In logical terms, if we imagine that our simplified research evidence is that a specific phenomenon (A) has a certain characteristic (B), then our evidence is that A entails B. If we want to conclude from this that phenomenon A therefore also has the characteristic C, then the third component of our syllogism (the classic form of our argument) is missing or implicit. This third component is that everything with characteristic B also has characteristic C. Thus, our complete syllogism is:

This A is B
 All B are C
 Therefore, this A is also C.

While the first part (A is B) may be likened to the evidence in a research study (e.g. water can be fatal), and the third (A is C) is the conclusion (e.g. water should be banned), then the second (B is C) is like the warrant (e.g. everything that can be fatal should be banned). In research reporting this step is often missed, as it is tacitly assumed by both the author and the reader. However, where the research is intended to change the views of others it is necessary to make the warrant explicit. This warrant can then 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. In the water example the warrant is clearly nonsense. Water can be fatal, but we cannot ban everything that could be fatal. But accepting that this warrant is nonsense also means that no evidence, however good, can be used with this precise format of argument to justify banning anything at all simply because it is fatal.

For Toulmin *et al.* (1979), a warrant is an argument that stands up to criticism, and that moves from a valid dataset to a claim. They present a similar example to the one above. The empirically based claim that Harry is a British citizen can be warranted from the evidence that Harry was born in Bermuda, and the warrant that anyone born in Bermuda will be a British citizen. The chief difference between this chain of reasoning and the classic syllogism is that it may also contain qualifying phrases (such as 'probably'), backing (in the form of relevant statutes about citizenship), and any known conditions for rebuttal (e.g. 'unless both of his parents were aliens').

Definitions of warrant vary between commentators and over time. The history of epistemology has seen clashes between realists and relativists, and empiricists and rationalists. It is not the purpose of this paper to revisit these debates (but see Bonjour, 1998; Musgrave, 1993). The purpose, rather, is to argue that when drawing conclusions from evidence, researchers should draw attention to those parts of their chain of reasoning that could be disputed. It should not matter, for example, whether a researcher believes in the existence of an external reality or not as long as they are clear about this when drawing conclusions. My guess would be that once ideas such as extreme relativism are made explicit in their research claims, then 'financial evolution' will play a large part in deciding whether the taxpayer, charity or funding council wishes to continue funding research by researchers who do not believe in the reality of the world they are researching. Similarly, policy-makers, once genuinely aware of the epistemological positions of researchers with opposing conclusions, will use that knowledge in making a judgement about them one way or the other (see below). Also, despite these earlier debates, most researchers appear to end up working with a mixture of pragmatic rationalism and fallible empirical realism (Platinga, 1993b). And it has been suggested that the remainder may merely be insufficiently aware of the basis of their own approach – there are many examples of researchers 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 pre-suppositions of his [sic] study that he finds it only too 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: ix)

Real-life Examples

Before continuing to further consideration of the nature of warrants, and what may be done to improve them, this section continues with five real-life examples of deficient warrants. This is partly to clarify the concept for, like many things, the warrants for research findings are most noticeable by their absence, and partly to begin to suggest how prevalent such problems appear to be. Humes and Bryce (2001), for example, 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 in educational research (see below). 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. I use this example here as an extreme case to introduce a point about warrants from evidence. A similar situation of the misreading of basic findings is reported in Lomborg (2001: 3) – 'in reality we were merely discussing who could look up a number correctly'. 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.

The reasons why this error occurred and how it was able to pass peer review

Table 1 Mean SES of students by locality of school

<i>Locality/Year</i>	<i>1990</i>	<i>1991</i>	<i>1992</i>	<i>1993</i>
Local school	3.20	3.22	3.26	3.19
Adjacent school	3.02	3.27	3.40	3.36

Source: Table 4 in Waslander & Thrupp, 1995

for an apparently high-quality journal are not known. Potential explanations include general innumeracy among readers, the ideological opposition of both authors and referees to choice in education, and simple misprints. However, a clearer line of argument than actually presented would have, presumably, either persuaded the sceptical reader of the correctness of the conclusion, or helped the authors themselves realise the mismatch between their evidence and their conclusions. The text simply states that the table shows that more distant schools are increasingly used by more privileged families after 1990. In fact, the paper is written almost as though the text was produced by one person, while the tables were prepared by another. This impression is confirmed by the response of one author to subsequent queries, that he cannot answer any questions since the tables were solely the responsibility of the other author (personal communication). Interestingly, despite this denial of the necessary technical knowledge, that author still claims that the conclusions are justified. This shows quite clearly the difference between personal belief (which is opaque and often incorrigible) and warrant (which must be capable of inspection and argumentation).

Now consider another example of a deficient warrant for conclusions from the educational research literature, but this time from a tradition of 'qualitative' analysis. Reay and Lucey (2000) report findings from interviews with families choosing a new school for their child. The researchers have the 'qualitative' evidence of the views, sex of parent, sex of child, occupational class and ethnicity from 15 parents, and they present some interview extracts from five of these. These are their findings (and it is important to note that we have no reason to doubt the quality of these). The authors conclude that 'in the parental interviews, a majority of the working class parents concurred with their child' (p. 90), and 'a significant deviation from this class trend was mothers of mainly black working-class boys' (p. 90). These are their conclusions, based on their evidence. However, Reay and Lucey do not present a chain of reasoning from the one to the other. If they did, then again either a sceptical reader would have been persuaded, or the paper would not have been published.

It would have taken little effort for the authors to report the number of parents who agreed/disagreed with their child, and the number of each of sex of parent, sex of child, occupational class, and ethnic group. This could have been presented succinctly as a table, and would have formed an important basis for their warrant. In fact, not only are we not told these frequencies, we are not even told the classification system used to generate the frequencies. If we make a charitable assumption that the parents were roughly half mothers and half fathers then there would be seven or eight (half of 15) of each. If we assume that the pupils were half girls and half boys, then there would be three or four parents in each gender group. Again, if we assume that half of each of these groups were middle-class and half working-class (and assuming that only two categories were used) then there would be one or two cases in each gender/class cell for comparison. Finally, if we assume that half of each of these groups were black and half non-black (again making the favourable assumption of only two categories) then there would be, on average, less than one case in each cell for comparison. Finally when we come to compare those that concurred with their child's choice and those that did not, then we have

on average less than half of one case per cell. When Reay and Lucey state that mothers of mainly black working-class boys differed from a larger pattern, they could be talking about one mother. If they are talking about more than one mother, this must mean that many of the other cells with which they make explicit comparison are actually empty (and therefore can provide no basis for that comparison). Made explicit, their evidence does not, and cannot, support their reported conclusions. Again the reasons why this error occurred and passed peer-review for publication in an apparently high-quality journal are unknown, but they may include the continuing and over-used dichotomy between qualitative and other research.

A very similar problem appears in the work of Pollard and Filer (1996) in one school in a middle-class community involving 10 pupils, one of whom refused to continue with the study, and two of whom moved schools during it. The evidence presented comes from interviews with only five of the remaining seven pupils. Having listened to these children, the researchers conclude that 'the implications of this study are that young children become effective learners when their self-confidence is high, the classroom social context poses manageable risks and they receive sufficient appropriate instruction and support' (p. 311). The key problem here is that this conclusion is simply not warranted by their evidence (although it may well be true). They do not subject these ideas to any firm test, nor even consider reasonable alternative explanations of their case studies. Does self-confidence produce effective learners as they claim, or do successful learners exhibit greater confidence? What is the underlying causal model here? What does it mean for support to be 'sufficient and appropriate'? If this is not defined then there is a danger of tautology (e.g. 'it was not sufficient because successful learning did not result'). Above all, how is it possible to reach such a conclusion from five (or perhaps seven) case studies? Presumably such a conclusion requires a bare minimum of one pupil each with self-confidence and without, in each of a classroom posing manageable and unmanageable risks, in turn receiving and not receiving the requisite support. Even if they had this minimum of eight pupils (which they did not) it is difficult to see how the comparison could be justified from one class in one middle-class school. Or how the researchers could eliminate any other unobserved factors affecting their judgement about the one pupil in each of their eight cells for comparison.

Of course, such problems are not confined to education, or even social science. In judicial proceedings (and media reporting), when forensic evidence (such as a fingerprint or DNA profile) is used to make a match with a suspect, prosecutors tend to use the probability of such a match as though it were a probability of guilt. This is what Gigerenzer (2002) describes as the 'prosecutor fallacy'. The actual evidence is the probability of the match (such as 1 in 10,000), but the warrant for the prosecutors' conclusion (there is therefore a 9999 in 10,000 chance of this individual's guilt) is very weak. They have to also argue that there is no human error in the matching process, that the match signifies presence of the suspect at the crime scene, that presence at the scene necessarily entails guilt, and so on. Above all, they have to demonstrate that the number of potential suspects is so small that a 1 in 10,000 chance is the equivalent of 'beyond reasonable doubt'. If the crime took place in a city of

1 million people, and if we make the favourable assumption that potential suspects are limited to residents only, then 1/10,000 means that 100 residents will have just such a forensic match. Thus, the suspect, *ceteris paribus*, has a 1/100 probability of guilt. This is much higher than for an average resident of that city (and therefore germane to the case without being conclusive), but much lower than 9999/10,000. The importance of this error, and others like them, is hard to overestimate in law, medicine and beyond. Yet, like the examples above it is created largely by an incompletely transparent chain of reasoning from evidence to conclusion. Or put another way, once the warrant is made explicit it can be easily seen to be false (Dawes, 2001).

Finally, consider an example of a warrant involving a causal model. Death rates due to cancer (of all types) increased over the course of the twentieth century in the UK, and they look set to continue to rise. One possible conclusion is that 'modern' lifestyle is to blame, including perhaps the food we eat and damage to our environment. The warrant here would be largely based on causation as correlation. Two sets of events, growth of cancer and lifestyle changes, are contemporaneous. Therefore, we assume that they are causally related and, of course, they may be. But we should also automatically start seeking alternative explanations, and see how these shape up. Such an approach fosters a healthy scepticism, and should be an almost instinctive response for all researchers. Another very plausible alternative is based on the fact of mortality. We all die. Therefore, a change in the probability of death by any one cause affects the probability of death by all other causes (put in statistical terms – the degrees of freedom of our model are fixed). As death rates due to typhoid, smallpox and war have declined so the death rates due to heart disease or cancer must be expected to rise (this is progress). If we add some more evidence, that people in the UK now live longer, on average, than at the start of the twentieth century, then the lifestyle theory becomes a much poorer explanation for the rise in cancer than the simple reduction of other avoidable causes of death. The latter explanation makes fewer assumptions for which we do not have direct evidence, and is therefore currently more 'scientific'. This example highlights another characteristic of a desirable warrant. It should be simple, as well as transparent.

The 2002 summary report form for the individual projects within the ESRC-controlled Teaching and Learning Research Programme contained a section on 'warrant' (TLRP Annual Reports, 2002). One of the first observations that can be made about the use of this section is that the 14 projects found it difficult to complete. There may be several reasons for this. Researchers may be unfamiliar with thinking in this way, the term 'warrant' does not have a clear common meaning, and a warrant is very difficult thing to discuss in the abstract – it is project- and context-specific – and therefore difficult to summarise in a paragraph or two. Perhaps the most important issue was the structure of the report form itself. The form required researchers to describe their 'key findings' first, and then their warrant, but, however disputed the term is, all descriptions of warrant agree that the logic of argument here involves three elements – evidence, conclusions, and the warrant(s) connecting the two. Put another way, researchers should be required to state what they actually found first, and then their warranted conclusions. However, in most cases TLRP

researchers used the key findings section to describe their conclusions, rather than their evidence. Thus, when they came next to describing their warrant they were somewhat at a loss. Many illustrated standard fallacies encountered in warranting conclusions, such as resting a conclusion on an appeal to popular sentiment, or to emotion, an *ad hominem* argument, irrelevant premises, the absence of contrary premises, or simply a restatement of their initial theories. In essence they ignored a key question for all warrants, and which they and their readers should be addressing – if the conclusions we have drawn were not true then how else could we explain the findings?

Most researchers merely rehearsed or emphasised their research methods (already described in full elsewhere on the form). This description was usually a summary of their data collection and/or analysis rather than the design itself, and a key claim appeared to be that the work used both what were termed ‘qualitative’ and ‘quantitative’ data (although how these were combined was not addressed). One project argued that their conclusions must be justified because they were based on ‘multi-faceted, rich and detailed data’; another simply stated that the research was of ‘high quality’. The most common approach that comes closer to addressing an issue of warrant is that of endorsement (although there was also a certain confusion here with issues of validity and reliability). For example, some projects referred to a close link between their findings and pre-existing literature or theory (although none actually specified what this link was). Some claimed that different datasets, or different nodes in a networked project, all pointed in the same direction, or that the judgement of several researchers agreed, or that practitioners had endorsed the findings in some way. One pointed out that their project was based on the logic of an experimental design (in which the warrant is implied and relatively straightforward), but then appeared not to accept the conclusion of the experiment (which was that the phenomenon sought was ineffective) and sought advice on how to find a weak effect that may still be hidden in the data collected. One project addressed the issue of warrant directly, using questions about their account – such as ‘is it convincing and coherent as well as persuasive to others?’ Their summary answer is largely based on a transparent and practically useful audit trail. What else could the others have done?

What Can We Do?

The first question to be asked of any evidence presented in support of a model of a social process is ‘but what else might this mean?’ The ability to discern rival explanations, while varying considerably between individuals, probably grows with practice (Huck & Sandler, 1979). It is a key skill for good research (but manifestly not a necessary one for ‘success’ in a research career at present). But, perhaps more importantly, it is a key skill for everyone to have as a consumer of the research of others. One way of improving this skill is to learn to recognise common forms of misleading argument. For example, the ‘fallacy of affirming the consequent’ is quite commonly encountered in social science. The fallacy argues that if A is true, then B will follow. Then if B appears it is taken by some researchers to mean that A is true. While seductive, there is no logic to this argument unless it starts more strongly with

'only if'. Otherwise exactly the same argument can be made with Z (or anything else) substituted for A.

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, as it is with a closely controlled experiment, but it is independent of any particular method of data collection (de Vaus, 2001). The National Research Council (2002) suggest principles for scientific research in education, of which the fourth is: 'Provide a coherent and explicit chain of reasoning' (p. 4). An important part of this involves 'systematically ruling out plausible counterexplanations in a rational, compelling way' (p. 4). The results should be disclosed to critique, and the warrant is intended to be persuasive to a sceptical reader (rather than playing to a gallery of existing 'converts'). Gorard *et al.* (2001), for example, presented a set of findings about changes over time in the social composition of UK secondary schools. They followed this with eight separate competing explanations for these findings, and spent the rest of the paper considering the relative merits of each. This is conducive to the scepticism felt necessary for research to prosper according to Shavelson *et al.* (2003: 27), whose guiding principles for warrants are:

To what extent can rival narrative accounts of the same action be ruled out? To what extent would another narrator replicate the account? To what extent does the narrative generalize to other times and places? There is nothing in the use of narrative form, by itself, that guarantees the veracity of the content of the account or which vitiates the need for the usual epistemic warrants used in science. How can it be determined that the narrative being used is complete, or does not misrepresent events?

This 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), in the same way that an infinite number of equations can join any two points on a graph. Therefore, no one can consider all possible theories to explain any finding – so that in social science, as in natural science, every 'law' that is ever proposed is quite literally false. This is also referred to as the underdetermination of theory by data, which perhaps expresses better the need to add something to the data in order to draw conclusions. This is the 'warrant'. The purpose of the warrant is show readers that the proposed explanation is the best we have at this point in time.

A useful short-cut is to employ parsimony to eliminate many of the potential alternatives. Parsimony is the canon attributed to Morgan (1903: 53), 'In no case may we interpret an action as the outcome of the exercise of a higher psychological faculty, if it can be interpreted as the outcome of one which stands lower in the psychological scale'. 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, whip-lash, tuberculosis and acute porphyria. 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 in their chain of reasoning the fewest (ideally none) assumptions for which we have no direct evidence.

Objections to Scientific Approaches

There will be commentators who oppose the call for greater rigour in warranting claims because of its link here to the term 'science'. 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: 95), but other than that it signifies very little. For more on this, see Gorard (2003). 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 be believed? 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. But 'interpretivist methods and analyses are sometimes abused to justify a lack of rigour' (Denscombe, 2002: 22).

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 & Darby, 1998) and these 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 (NRC, 2002). There are certain shared assumptions underlying all research, whatever methods are used, and there are no pure ontological or epistemological divisions in practice (Denscombe, 2002; also see above, and Gorard, 2002b). Research requires rigorous reasoning supported by a mixture of methods, and findings leading to testable models or theories. Despite the fact that science 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). Dewey (1916) warned against 'our predilection for

premature acceptance and assertion ... Even in the case of failure, we are inclined to put the blame not on the inadequacy and incorrectness of our data and thoughts ... Science represents the safeguard of the race against these natural propensities and the evils that flow from them.'

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: 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: 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 kinds of warrants used by teachers and policy-makers may differ both from each other (Lewis, 2001), and from those used by researchers (McNamara & Corbin, 2001). For example, teachers largely ignore 'evidence' (as opposed to experience) 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). Similarly, archives of research knowledge are not used by practitioners to improve their teaching (Hiebert *et al.*, 2002). Perhaps one way to overcome this lack of engagement is for potential users to be involved from the 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 do just that by involving the 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, education can claim to be an obvious field in which academics, as lec-

turers, are also routinely practitioners in their area of study. The ISS model may therefore be more important for disciplines like economics, geography and especially 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 & 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 to suppose that ISS, of itself, can improve quality, while it could lead to the charge of creating evidence on demand. There is even continuing dispute about what counts as a valid warrant, and politicians, lawyers and researchers, routinely use dubious *ad hominem* appeals or draw on personal experience/anecdote, as well as pointing out the irrationality of their opponents (Gergen, 1989).

Conclusion

One conclusion from this brief review 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 re-reading, about the nature of the evidence produced, but about the way in which it is converted into 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 or, as in some examples above, plain error). Second, perhaps partly because of this overclaiming, users and consumers of research largely do not worry about the relationship between evidence and conclusions, but use warrants of their own devising. This might be part of the reason why the actual impact of research findings is weak, and apparently not related to research quality. Findings may be used because they are convenient, or as rhetorical justification for an existing position, rather than warranted. For those who wish this situation to change, perhaps demanding an explicit warrant in all papers as a *sine qua non* for the peer-review process would help. As suggested above, whether we are told enough to make a sensible judgement about reported conclusions is more important than whether we agree with them.

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).

Correspondence

Any correspondence should be directed to Dr S. Gorard, Cardiff University School of Social Sciences, Glamorgan Building, King Edward VII Avenue, Cardiff CF10 3WT, Wales, UK (gorard@cardiff.ac.uk).

References

- Bonjour, L. (1998) *In Defense of Pure Reason*. Cambridge: Cambridge University Press.
- 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.
- Brignell, J. (2000) *Sorry, Wrong Number! The Abuse of Measurement*, European Science and Environment Forum.
- Dawes, R. (2001) *Everyday Irrationality*. Oxford: Westview Press.
- de Vaus, D. (2001) *Research Design in Social Research*. London: Sage.
- Denscombe, M. (2002) *Ground Rules for Good Research*. Buckingham: Open University Press.
- Dewey, J. (1916) *Democracy and Education*. New York: Macmillan.
- Gergen, K. (1989) Warranting voice and the elaboration of self. In J. Shotter and K. Gergen (eds) *Texts of Identity*. London: Sage.
- Gibbons, M. (2000) Mode 2 society and the emergence of context-sensitive science. *Science and Public Policy* 27, 3.
- Gigerenzer, G. (2002) *Reckoning with Risk*. London: Penguin.
- Glymour, C., Scheines, R., Spirtes, P. and Kelly, K. (1987) *Discovering Causal Structure*. Orlando: Academic Press.
- Gorard, S. (2002a) The role of causal models in education as a social science? *Evaluation and Research in Education* 16 (1), 51–65.
- Gorard, S. (2002b) Can we overcome the methodological schism? Combining qualitative and quantitative methods. *Research Papers in Education* 17 (4), 345–361.
- Gorard, S. (2003) *Quantitative Methods in Social Science Research*. London: Continuum.
- Gorard, S., Fitz, J. and Taylor, C. (2001) School choice impacts: What do we know? *Educational Researcher* 30 (7), 18–23.
- Hiebert, J., Gallimore, R. and Stigler, J. (2002) A knowledge base for the teaching profession: What would it look like and how can we get one? *Educational Researcher* 31 (5), 3–15.
- Hillage, J., Pearson, R., Anderson, A. and Tamkin, P. (1998) *Excellence on Research in Schools*. Sudbury: DfEE.
- Huck, S. and Sandler, H. (1979) *Rival Hypotheses*. New York: Harper and Row.
- 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.
- Kennedy, M. (1997) The connection between research and practice. *Educational Researcher* October 1997, 4–12.
- Lewis, J. (2001) Getting social science research into practice. *Learning and Skills Research* 5 (1), 17–18.
- Lomborg, B. (2001) *The Skeptical Environmentalist: Measuring the Real State of the World*. Cambridge: Cambridge University Press.
- 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.
- Morgan, C.L. (1903) *Introduction to Comparative Psychology*. London: Walter Scott.
- Musgrave, A. (1993) *Common Sense, Science and Scepticism*. Cambridge: Cambridge University Press.
- 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.
- National Research Council (2002) *Scientific Research in Education*. Washington, DC: National Academy Press.
- Phillips, D. (1992) *The Social Scientist's Bestiary*. Oxford: Pergamon Press.
- Plantinga, A. (1993a) *Warrant and Proper Function*. Oxford: Oxford University Press.
- Plantinga, A. (1993b) *Warrant: The Current Debate*. Oxford: Oxford University Press.
- Pollard, A. and Filer, A. (1996) *The Social World of Children's Learning*. London: Cassell.
- Postan, M. (1971) *Fact and Relevance: Essays on Historical Method*. Cambridge: Cambridge University Press.
- Rappert, B. (1997) Users and social science research: Policy, problems, possibilities. *Sociological Research Online* 2 (3).
- 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.
- Reay, D. and Lucey, H. (2000) Children, school choice and social differences. *Educational Studies* 26 (1), 83–100.
- Shavelson, R., Phillips, D., Towne, L. and Feuer, M. (2003) On the science of education design studies. *Educational Researcher* 32 (1), 25–28.
- Stokes, D. (1997) *Pasteur's Quadrant: Basic Science and Technological Innovation*. Washington, DC: Brookings Institution Press.
- TLRP Annual Reports 2002 (2003) *TLRP Annual Reports 2002*. Cambridge: Faculty of Education.
- Tooley, J. and Darby, D. (1998) *Educational Research: A Critique*. London: OFSTED.
- Toulmin, S. (1958) *The Uses of Argument*. Cambridge: Cambridge University Press.
- Toulmin, S., Rieke, R. and Janik, A. (1979) *An Introduction to Reasoning*. New York: Macmillan.
- Waslander, S. and 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.