

Trials in Public Policy

Project Update

Randomised Controlled Trials in the Social Sciences: Challenges & Prospects

The first annual conference on randomised controlled trials in the social sciences was held at the University of York on 13th-15th September, 2006. The conference was jointly organised by Professor Stephen Gorard, Dr Carole Torgerson (who are leading the ESRC-funded RDI project on training in pragmatic social interventions) and Professor David Torgerson (Director of the York Trials Unit and Chair of the Social Science Trials Methods Group). The conference was opened and closed by keynote speeches from Professor Thomas Cook of Northwestern University, Chicago and Dr Philip Davies of the Treasury Office. Invited speakers were Jacque Mallander, Dr Robert Coe, Prof Laurence Moore and Professor James Middleton. Thirty four delegates attended the workshops on the first day, and seventy six delegates attended the main conference. There was a range of interesting oral presentations describing trials across the social sciences, including trials in politics, criminal justice, work and pensions, psychology, education and social welfare.

Prior to the main conference there were workshops on cluster randomized trials (led by Professor Martin Bland, Dept. of Health Sciences, University of York) and systematic reviews of randomized controlled trials (led by Carole Torgerson, with invited presentations from Dr Isabelle Boutron, University of Paris, David Torgerson and Dr Amanda Perry, Dept. of Criminal Justice, University of York). The conference also hosted the inaugural meeting of the Social Sciences Trials Methods Group, which is a national initiative designed to increase the use and support of RCTs in the social sciences.

The second annual conference on RCTs in the social sciences will be held at the University of York during the week commencing 12th Sept. 2007.

All researchers should want to conduct a 'fair test'

Stephen Gorard

What follows is a summary of my own conclusions drawn from participation in the international debate on education research methods:

a debate that is growing by the decade, as demand apparently grows for sound evidence on which to base educational policy and practice. I shall set out a number of summary propositions. Interested readers can trace the further basis for these propositions in my research writings – examples of which are provided.

A key ethical concern for those conducting or using publicly-funded education research ought to be the quality of the research, and so the robustness of the findings, and the security of the conclusions drawn.

Until recently, very little of the writing on the ethics of education research has been concerned with quality. The concern has been largely for the participants in the research process, which is perfectly proper, but this emphasis may have blinded researchers to their responsibility to those not participating in the research process. The tax-payers and charity-givers who fund the research, and the general public who use the resulting education service, have the right to expect that the research is conducted in such a way that it is possible for the researcher to test and answer the questions asked. Generating secure findings for use could involve a variety of factors including care and attention, sceptical consideration of plausible alternatives, independent replication, transparent prior criteria for success and failure, use of multiple complementary methods, and explicit testing of theoretical explanations through randomised controlled trials or similar experimental designs (Gorard 2002a).

It is helpful to consider the research enterprise as a cycle of complementary phases and activities, because this illustrates how all methods can have an appropriate place in the full cycle of research.

Experimental designs, like in-depth work or secondary analysis, have an appropriate place in the cycle of research from initial idea to development of the results. The main reason to emphasise experiments at this point in time is not because they are more important than other phases in the cycle, but because they represent a stage of work that is largely absent in education research.

Inside this issue:

Project Update

Fair Test

Top tips

Contact Details

RCT Helpline



If nearly all of education research were currently conducted as field trials then I would be one of the commentators pleading for more and better in-depth work or secondary analysis, for example. Other weak points in the cycle are currently the systematic synthesis of what we already know in an area of work, the design or engineering of what we already know into usable products for policy and practice, and the longer-term monitoring of the real-world utility of these products (Gorard with Taylor 2004, Gorard et al. 2004).

Working towards an experimental design can be an important part of any research enterprise, even where an experiment is not envisaged or even possible.

Sometimes a true experiment, such as a large randomised controlled trial, is not necessary, and sometimes it is not possible. An experiment is not necessary in a variety of research situations, including where the research question does not demand it, or where a proposed intervention presents no *prime facie* case for extended trialling. An experiment may also not be possible in a variety of research situations, including where the intervention has complete coverage, or has already been implemented for a long time, or where it would be impossible to allocate cases or clusters at random. However, a 'thought experiment' is always possible, in which the researchers consider no practical or ethical constraints except answering the research question as clearly as possible. In then having to compromise from this 'ideal' to conduct the actual research, the researcher may come to realise how much more they could be doing. There might then be more natural experimental designs, more practitioner experiments, and surely more studies with appropriate comparison groups rather than no explicit comparison at all (a situation which reviews show is the norm for UK academic research in education). There might also be more humility about the quality of the findings emanating from the compromise design (Gorard 2002b, 2003a).

Part of the problem of research quality lies in traditional research methods training and 'experts'.

In the UK, traditional methods training for new researchers in university departments of education generally starts by introducing students to differences between types of research, and emphasising the purportedly incommensurable values underlying the variety of approaches to discovery. Most obviously, researchers are introduced to a supposed paradigmatic division between 'qualitative' and 'quantitative' studies in a way that encourages methods identities based on a choice of only one of these 'paradigms'. This leads many to indulge in paradigmatic strife, or to write off entire fields of endeavour – as being 'positivist', for example. Some commentators try to heal these schisms after they have been created, but there is a shortage of texts and training resources that take the far superior approach of assuming that there is a universal underlying logic to all research.

Such an approach leads to a focus from the outset of training on the craft of research, thus bringing design, data collection, analysis, and warranting results to the fore, leaving little or no place for paradigms (Gorard 2003b, 2004a).

Part of the problem of research quality may lie in a lack of appropriate integrated use of numbers.

Since experimental designs are seen by many, incorrectly, to be 'quantitative' in nature, a lack of researchers willing and able to work with numbers could also be part of the reason for the lack of experimental work. There may be a range of influences at play here, including poor maths teaching in schools, lower ability of social science students in comparison to other disciplines in terms of maths, selection of methods courses by students in terms of perceived ease, and the widespread misunderstanding that being a 'qualitative' researcher means never having to deal with numbers. However, I am coming increasingly to the view that a major share of the blame lies with 'quantitative' researchers. They seem to prefer devising more and more complex methods of analysis rather than devoting their energy to creating higher quality datasets that are easier to analyse. They often present their research in exclusive and unnecessarily technical ways. They generally assume, incorrectly, that numbering is the same as measuring, that reliability is the same as validity, that probabilistic statistics can be used with purposive samples, and that any use of numbers must be based on sampling theory. This is not the way forward (Gorard 2006a, 2006b).

Part of the problem of research quality lies in an unwillingness to test cherished theories.

Another element of the methods crisis stems from a love of specific theories, and a consequent unwillingness to test them for failure. A typical piece of evaluation in UK education is either commissioned by, or conducted by, those responsible for the programme being evaluated. There may then be pressure from funders to 'finesse' the results. I have certainly been contacted by evaluators seeking some new kind of analysis that will gainsay the surface findings, and which will support instead their underlying belief that the programme must be being effective. This is no different, in principle, to the dredging of data that goes on *post hoc* in other forms of research as well. I have also experienced too many cases in which researchers appear to distort data in order to help preserve their prior beliefs. Some methods experts in the UK actually advise researchers to 'take sides' before conducting research, and not to publish unhelpful results. Of course, it remains true that the evidence-based approach to policy-making and practice is itself untested in education, and still far from satisfactory in fields such as health sciences. But this is a reason to test it, not reject it out of hand (Gorard 2004b, Gorard and Fitz 2006).

Much of the solution lies in greater scepticism, because the problem is not really one of methods at all.

Some of the criticism of education research in the US, UK and elsewhere during the 1990s was concerned

with relevance. But education is a very applied field of research. I do not find, as I review evidence for different projects, much published research that has no relevance to some important or useful component of education. The criticism should more properly be directed to the poor quality of much research, where even though the findings may have relevance they still cannot be used safely. In response to these perceived deficiencies, formal capacity-building activities have tended to focus on solutions in terms of methods, such as having more quantitative work, more systematic reviews, or more experiments. These, in themselves, are not the answer. The answer for me lies in genuine curiosity, coupled with outright scepticism. These characteristics lead a researcher to suit methods to purpose, try different approaches, replicate and triangulate, and to test their findings. It leads them to consider carefully the logic and hidden assumptions on the path from evidence to conclusions, automatically generating caveats and multiple plausible interpretations from the standard query – ‘if my conclusions are actually incorrect, then how else could I explain what I have found?’. Genuine curiosity leads, almost inevitably, to a desire to conduct a rigorous test or trial of any tentative explanations of results. Thus, the concept of a ‘fair test’ of an idea is relevant to all researchers (Gorard 2002c, 2005).

References

- Gorard, S. (2002a) Ethics and equity: pursuing the perspective of non-participants, *Social Research Update*, 39, 1-4
- Gorard, S. (2002b) The role of causal models in education as a social science, *Evaluation and Research in Education*, 16, 1, 51-65
- Gorard, S. (2002c) Fostering scepticism: the importance of warranting claims, *Evaluation and Research in Education*, 16, 3, 136-149
- Gorard, S. (2003a) *Quantitative methods in social science: the role of numbers made easy*, London: Continuum
- Gorard, S. (2003b) Understanding probabilities and reconsidering traditional research methods training, *Sociological Research Online*, 8,1, 12 pages
- Gorard, S. (2004a) Scepticism or clericalism? Theory as a barrier to combining methods, *Journal of Educational Enquiry*, 5, 1, 1-21
- Gorard, S. (2004b) Three abuses of ‘theory’: an engagement with Nash, *Journal of Educational Enquiry*, 5, 2, 19-29
- Gorard, S. (2005) Current contexts for research in educational leadership and management, *Educational Management Administration and Leadership*, 33, 2, 155-164
- Gorard, S. (2006a) *Using everyday numbers effectively in research*, London: Continuum,
- Gorard, S. (2006b) Towards a judgement-based statistical analysis, *British Journal of Sociology of Education*, 27, 1, 67-80
- Gorard, S. and Fitz, J. (2006) What counts as evidence in the school choice debate?, *British Educational Research Journal*, 32, 6, 797-816
- Gorard, S., Rushforth, K. and Taylor, C. (2004) Is there a shortage of quantitative work in education research?, *Oxford Review of Education*, 30, 3, 371-395
- Gorard, S., with Taylor, C. (2004) *Combining methods in educational and social research*, London: Open University Press

Top Tip

What follows is the second in our feature where participants in the project share what they consider to be ‘top tips’ for those of us conducting trials (see Issue 3). Tips can be about anything of practical use, and we encourage readers to send in suggestions (contact educ-trials-pp@york.ac.uk). In this issue, Carole Torgerson explains some of the ways to overcome the common problem where everyone wants to take part in the treatment group and no-one wants to be in the control.

Random allocation: The ‘waiting list’ design and the ‘stepped wedge’ design Carole Torgerson

There generally has to be a *prime facie* case for trialling some new intervention in public policy, perhaps based on prior but less rigorous evidence of its effectiveness. One potential problem then faced by a randomized controlled trial is that the control group is unhappy with being denied this promising intervention. When we are evaluating an intervention where the evidence is uncertain then there is no ethical problem: indeed, it is ethically correct to offer the participants a chance (random possibility) to be offered the most effective intervention, which may very well be the control condition. On the other hand, potential participants may not be convinced that having the control intervention is as likely to be as beneficial as the novel condition. Because of this anticipated benefit, those allocated to the control may suffer resentful demoralisation, and either refuse to continue with the experiment or deliberately or subconsciously do worse merely because they have been refused the novel intervention. We may also wish to evaluate the implementation of an intervention that has been shown to be effective in a laboratory type RCT (explanatory or efficacy trial) and we may wish to evaluate its effectiveness in the ‘real’ world. Finally, a national policy may be implemented when there is dubious or no real evidence of effectiveness but the political imperative is to be seen to do ‘something’. One way of addressing these problems is to use either a ‘waiting list’ or a ‘stepped wedge’ design.

In a waiting list study participants are told explicitly that they will receive the intervention; however, some will receive it straight away, whilst others will receive it later. We can then evaluate the effectiveness of the intervention by measuring both groups at pre-test, implementing the intervention in one group, giving a post-test measurement (after which the intervention is then given to the controls). As an example consider the RCT by Brooks and colleagues (2006). This evaluated the use of a computer software package in a secondary school. The package was usually implemented arbitrarily, as there were insufficient lap top computers for all pupils to receive the intervention at the same time. For the evaluation the researchers changed the arbitrary assignment to random allocation and adopted a

waiting list design which permitted a rigorous evaluation of the software package. The use of the waiting list in this instance allowed all the children to receive the package and may have reduced any demoralisation either on the part of the children or their teachers.

The stepped wedge design differs from the waiting list design in that it operates as a series of waiting lists. For example, the Sure Start evaluations (which did not use a randomised controlled trial design) could have used the stepped wedge approach. In this study there was huge political pressure to implement 'something'. A RCT design could have been incorporated into the evaluations by randomising geographical areas into several implementation phases. Baseline pre-tests would have occurred at the beginning of the study, and post-tests would have been included each time a new geographical area implemented the intervention. In this manner it would have been possible to control for potential confounding factors that may have undermined the non-randomised evaluation method that was used for Sure Start. More information on the stepped wedge design can be found in a recent systematic review of the method (Brown and Lilford, 2006).

References

Brooks, G., Miles, J.N.V., Torgerson, C.J. and Torgerson, D.J. Is an intervention using computer software effective in literacy learning? A randomised controlled trial, *Educational Studies*, 32(1), 2006.

Brown CA, Lilford RJ (2006) The stepped wedge design: a systematic review. *BMC Medical Research Methodology* 6, 54.

RCT Help Line

If you have a query or would like help or advice on any aspect of designing, running or evaluating randomised controlled trials, please contact us. Where appropriate, a member of the project will be happy to visit the site to provide personal assistance.

Contact Us:

Tel: 01904 433466 or

Email: educ-trials-pp@york.ac.uk

Our thanks again to all who came and contributed to our Conference in September. However, we want to encourage more people to be involved in face-to-face events, and in virtual participation, from all areas of public policy. In particular, we want to hear from national, regional and local policy-makers and practitioners who do or could use evidence from rigorous evaluations in their fields. And from research methods trainers, struggling with the place of trials methods in their courses. The first two events were in York, but we are happy to hold or help organise events wherever they are wanted. Please contact us with your comments and suggestions.

The RDI Trials Project Administrator

Department of Educational Studies

University of York

Heslington

York

YO10 5DD

Phone: 01904 433466

Fax: 01904 433459

Email: educ-trials-pp@york.ac.uk