

Occasional Paper Series
Paper 42

The way forward for educational research?

Stephen Gorard

Cardiff University School of Social Sciences

ISBN 1 872330 606

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
Patrick White
Helen Taylor (Project Administrator)

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

The way forward for educational research?

Stephen Gorard
Cardiff University School of Social Sciences
King Edward VII Avenue
Cardiff, CF10 3WT
02920-875113
email: gorard@cardiff.ac.uk

Abstract

Educational research in the UK has for some time been criticised in terms of both its relevance and its quality. Indeed, these issues of relevance and quality have been presented by critics as somehow linked with each other, and with a dearth of large-scale 'quantitative' evidence about teaching and learning. It is, therefore, interesting to consider the recent Hay/McBer research into teacher effectiveness, one of many responses, in the light of these criticisms. To what extent has the work used quantitative evidence in a high quality piece of research to address relevant issues for practitioners and policy-makers? The answers presented in this brief account of the research are that it is clearly important work, uses a variety of numeric and other data sources, but that the design is not one that the research community should set out to emulate. Despite considerable prior adjustment to the sample, in favour of the publicised findings, it remains the case that only 55% of the teachers rated 'outstanding' in a dichotomous classification by Hay/McBer actually achieved pupil-results higher than would be expected (once prior attainment and contextual variables were taken into account). This success rate in identification is insufficiently better than would occur by chance for the researchers to then present the characteristics of these supposedly outstanding teachers as ones that other teachers should take note of. If this report is being presented by the DfEE as the form of research that they and other client bodies are seeking, then we must conclude that quality is linked neither to research relevance, nor is it necessarily an outcome of the use of quantitative evidence *per se*.

The way forward for educational research?

Abstract

Educational research in the UK has for some time been criticised in terms of both its relevance and its quality. Indeed, these issues of relevance and quality have been presented by critics as somehow linked with each other, and with a dearth of large-scale 'quantitative' evidence about teaching and learning. It is, therefore, interesting to consider the recent Hay/McBer research into teacher effectiveness, one of many responses, in the light of these criticisms. To what extent has the work used quantitative evidence in a high quality piece of research to address relevant issues for practitioners and policy-makers? The answers presented in this brief account of the research are that it is clearly important work, uses a variety of numeric and other data sources, but that the design is not one that the research community should set out to emulate. Despite considerable prior adjustment to the sample, in favour of the publicised findings, it remains the case that only 55% of the teachers rated 'outstanding' in a dichotomous classification by Hay/McBer actually achieved pupil-results higher than would be expected (once prior attainment and contextual variables were taken into account). This success rate in identification is insufficiently better than would occur by chance for the researchers to then present the characteristics of these supposedly outstanding teachers as ones that other teachers should take note of. If this report is being presented by the DfEE as the form of research that they and other client bodies are seeking, then we must conclude that quality is linked neither to research relevance, nor is it necessarily an outcome of the use of quantitative evidence *per se*.

Background

During the last decade, the value and effectiveness of research as a contribution to the improvement of education has been increasingly called into question (e.g. Hargreaves 1997, Hillage et al. 1998, Tooley and Darby 1998). Educational research has been accused of being both 'second rate' and irrelevant to the needs and interests of practitioners. Millett, then the Chief Executive of the Teacher Training Agency, argued that 'despite the expenditure of over £65 million of public funding on educational research each year, there are surprisingly few studies which individually, or collectively, contribute systematically to the development of a comprehensive body of high quality evidence about pedagogy' (Millett 1997, p.2). Woodhead, then Her Majesty's Chief Inspector for Schools, claimed to have given up reading research as 'life is too short. There is too much to do in the real world with real teachers in real schools to worry about methodological quarrels or to waste time decoding unintelligible, jargon-ridden prose to reach (if one is lucky) a conclusion that is often so transparently partisan as to be worthless' (Woodhead 1998, p.51). This epistemological crisis of confidence is not confined to the UK (NRC 1999, NERPP 1999, 2000, Resnick 2000).

At heart, these criticisms address two main issues. The first is the claimed lack of real-world relevance of much research. Much educational research in the UK is not directly transferable into improved pedagogic practice - although whether it should be so remains a matter for debate, for many of those criticising the relevance of research have a very narrow, usually initial-school, view of what constitutes

learning, for example. The second issue is an apparent system-wide gap in expertise in conducting large-scale studies, especially field trials derived from laboratory experimental designs. Most educational research in the UK is small-scale, non-replicable, and interpretative, leading perhaps to too many insecure conclusions. Put bluntly, most researchers are avowedly qualitative in approach, but politicians and funders want to see the pendulum swing back towards a more balanced portfolio of skills. This does not mean that we necessarily need *complex* techniques. Our approach could start from a consideration of the importance of 'truth' (Bridges 1999), and a return to a political arithmetic tradition (Mortimore 2000).

Relevance, quality and funding

Allowing political bodies and practitioner groups, but not academic researchers, to decide on a research agenda may produce greater short-term relevance but is almost certainly not going to produce research of higher quality. There may, in fact, be a tension between looking for rigour and looking for relevance in research. Central planning has therefore failed to produce better research in other fields (Hammersley 1997). Many researchers who have worked on consultancies, contract research and evaluation studies will have experienced the pressure, subtle and not so subtle, put upon them to produce results in accord with some pre-determined plan. It is almost as though 'research' is being conducted to find evidence for an already existing agenda. Perhaps when politicians with legal experience talk about evidence-based policy they do not mean what I understood the term to mean when I first heard it. I supposed, then, that it referred to making policy based on research evidence about what is likely to be most effective or the fairest outcome (the social science definition). An alternative, and equally plausible, interpretation (the legal definition) would be that evidence was what a policy-maker sought to help establish a case for a policy. It is therefore important for all concerned to decide which of the two versions is being supported in the push for evidence-bases. They are incompatible.

I have recently undertaken research for government departments, the National Assembly for Wales, and the QCA for Wales among others. In each case my draft report was adjusted for style and readability, but also for content. I had to work quite hard to keep passages about the inadequacy of official assumptions (see Gorard et al. 2001a), about the conveyor belt effect (see Gorard et al. 2002), or the actual nature of differential achievement between boys and girls (see Gorard et al. 2001b), from being lost in the official reports (where these were published). I have recently seen a DfEE invitation to tender for a project ameliorating the underachievement of boys (taking it for granted that such a phenomenon exists, and that it has a pedagogical root), and a BECTa invitation to conduct research to 'find out why computers are so useful' in school work (taking it for granted that they are). Anyone wishing to apply for this money who was not sure whether boys were being led to underachieve, or whether computers were useful in schools would be excluded automatically. Thus, the DfEE and BECTa will be able to shield itself from unwelcome findings. They are working, perhaps unwittingly, with the legal definition of evidence. Conversation with colleagues, and at conferences, lead me to believe that most readers will recognise this list, and be able to add to it from their own experience. The line between relevance and censorship is a relatively fine one.

Partly as a response to the UK debate about the value and relevance of academic educational research, government appointed bodies have tried to move some research funding into the hands of practitioner groups, while government departments have increased their direct research and development funding. The result has so far been unimpressive. I would not wish to try and defend the overall quality of much UK academic educational research, but research by teacher-researchers in isolation could be even worse. The Teacher Training Agency (TTA), for example, award small grants to teachers to carry out research 'relevant' to their needs. Judged on the basis of the cream of the pilot studies published so far, the research so funded is often actually not research at all, not being aimed at producing knowledge based on the collection of evidence. It often simply describes current practice or uncontrolled attempts to change it (so-called 'action research'). Not all studies produce a report, but even the rest omit crucial details such as what they had done, or what their evidence was. Their apparent conclusions are mainly repetitions of previously held opinions (there being no sign of the surprise that for me is the hallmark of real discovery), but their 'lack of critical scrutiny also allows the presentation of questionable ideological views about the nature of practice under the banner of scientific research' (Foster 1999, p.396). I would not wish to base best-practice on these tiny case studies (although it should be noted that the examples used here come from the early stages of the scheme)

Perhaps the best such study I have seen concerns an 'experiment' on the teaching of mental arithmetic (TTA 2000). In the school concerned the 1998 Key Stage One cohort (54 pupils) were taught mental arithmetic using a new method. A higher proportion gained level 3 than in the 1997 cohort (55 pupils) taught using another method. Thus, the researchers conclude that the new method is better, and the TTA have supported this view by publishing it. Despite being the best study of the first 26 funded (by presenting evidence and method for example, and having a reasonable number of cases), this is precisely the kind of evidence we do not need or want in education. The findings are simply not safe. The cohorts were different and unmatched, the proportions gaining level 3 are growing year by year nationally anyway, the two groups sat different test papers, the teachers were not matched for the different cohorts (the Maths co-ordinator taking the second group), and there is no consideration of either the Hawthorne or experimenter effects (see Gorard 2001).

The Hay/McBer model of teacher effectiveness

Perhaps the clearest example of increased political control of research-funding comes from the recent Hay/McBer study. The DfEE apparently spent nearly £4 million on the Hay/McBer research into teacher effectiveness. I have only had access to the introductory report on the website, the summary of findings for Phase II, and the report of numerical aspects of the study. The following comments are based on this publicly available information, and should be read, not as definitive, but as the kind of comments that might have been made by a referee or peer-reviewer if the work had been subject to such. In many cases, I simply felt I needed to know more - about the methods, logic, decisions and justifications. It is only after being convinced by these that I will engage with, and trust, the substantive findings.

The main result of the study was a description, and by implication a prescription, for what effective teachers actually do when teaching (Hay/McBer 2000). The relevance of this work cannot be doubted.

What is much more doubtful is any claim to knowledge stemming from this colossal (in educational research terms) expenditure. In essence, the researchers are saying that they have described, via a range of data collection techniques, what 96 effective teachers are like. However, even assuming that the researchers can successfully identify an effective teacher (see below) this description is of little practical use. It is like the 'potted plant theory' in school improvement. If these 96 teachers tend to dress smartly for example, then an argument is advanced that dressing smartly can lead previously less successful teachers to improve. This is, of course, nonsense but it is very popular kind of nonsense (Davis 2001). The effective teachers were largely selected by a combination of pre-existing ideas and asking other people 'who is effective?'. So the ensuing description of the attributes that led to their identification is that of other people. The study tells us what an unpublicised group of 'other people' think makes a successful teacher. It is no surprise therefore that, like prescriptions for school improvement, the ensuing model contains elements of tautology, the blindingly obvious, and the impractically nebulous.

In fact Hay/McBer defined effectiveness both in terms of ratings and observations, *and* from pupil progress scores. In order to arrive at the 96 cases, it was only possible to use schools with high quality pupil records and other datasets. The population for the study is therefore not, as it appears at first sight, all schools but all schools with high quality datasets (i.e. all those with a non-zero chance of being in the sample). It is perfectly possible that such schools, and the teachers within them, will tend to be different from the rest in important ways.

Even given this restriction on data quality, the number and range of actual datasets encountered in the sample schools were considerable (Reynolds 2000). Half of the schools had unique home-made schemes for calculating value-added scores, and the remainder used various scoring systems such as Yelsis or the optional QCA SATs (ten different examples are presented in the relevant appendix). Most also used some socio-economic background or contextual variables, but again these were not consistently collected and recorded from school to school. Making the best use of the compromise data available, as is always necessary in research, Reynolds (2000) and a colleague converted the pupil gain score for each teacher into categories such as A+ (very effective) to C (poor), and the judgement of these categories was based on pupil scores for each teacher relative to context and to prior attainment. The Hay/McBer study also grouped these teachers, but not using the same categories preferring instead a range of 'Outstanding+' to 'Poor'. These judgements were made on the basis of observations and the rating of others (but unlike the Reynolds classification we are not told who made the judgements, nor how). Teachers judged ineffective or poor were dropped from the original sample of 172, leaving only 128. So now the sample, and its population, consisted of teachers not rated ineffective by others, from schools with adequate datasets (and in a limited range of subject areas, of course). It is also the case that both approaches used a classification system primarily because they had not generated reliable scores for each teacher. Whether collapsing the scores into groups overcomes this unreliability in any way is not an issue addressed anywhere in the public reports I have seen.

The ratings score was correlated with the value-added score, yielding a coefficient of +0.43. This means that if, for the present, we accept the validity of both approaches then only 18% of the variance in these already heavily edited groups is common to both. If outstanding teachers are meant to lead to greater progress for their pupils then this 18% is *prime facie* evidence that one or both of the classification systems is heavily in error. While a definition of an effective/outstanding teacher may be about a lot more

than pupil attainment, it might be assumed that all definitions would include some reference to it. In the circumstances it was up to the Hay/McBer to explain why the correlation was so 'moderate', and why they felt able to continue in spite of it. Given that the 'Reynolds' classification takes into account contextual figures whereas the 'Hay/McBer' does not this may suggest that it is the latter which is more at fault. Alternatively it *might* be possible to argue that the acknowledged limitations and variations in the datasets used by Reynolds are somehow absent from the more qualitative material used in the main report.

However, rather than suggest anything like this the study edits the 'sample' again. It only proceeds to analyse and characterise those individuals whose ratings intersect to a large extent. Put another way, 32 teachers who have A+ pupil progress scores but were not considered to be good teachers by others, or who had weak pupil progress scores but were considered 'outstanding' by others, were dropped (and remember that all purportedly poor teachers had already been dropped prior to the initial correlation). Not surprisingly the correlation between the two classifications for the remaining 96 cases rose to +0.79, but I consider it inappropriate (or worse) for the study to even attempt to cite this figure. Running a correlation, removing cases from the awkward diagonal, running it again and then reporting the final correlation coefficient is not good science.

Table 1 - Crosstabulation of teachers effectiveness: Value-added versus Hay/McBer

	Value-added B/C	Value-added A	Total
Hay Outstanding	23	28	51
Hay Typical	62	15	77
Total	85	43	128

Returning to the already heavily-edited sample of 128, for which Reynolds (2000) reports a 'moderate correlation' between the two assessments of teachers effectiveness, 51 of these were rated outstanding by Hay/McBer, while 43 were given very effective grades by Reynolds. Only 28 of these individuals were in the intersection. Collapsing the various categories in the two different scales into the simplest classification of two classes each - whether outstanding or not - leads to Table 1. Therefore, only 55% (or 28/51) of teachers rated outstanding by Hay had better than typical pupil progression. This is, presumably, not sufficient to enable the study to claim that they can describe, and perhaps prescribe, the characteristics of an effective teacher. Trimmed to this summary, the report is a list of the attributes that some people already think a good teacher has, backed up by rather poor statistical manipulation (and here I am not referring to the value-added scores themselves which are commendable, having a clear method combined with a report of their inevitable defects).

Conclusion

We need to consider dispassionately the possibilities and limitations presented by the new climate of evidence-based policy and practice, and the appeal for more large-scale designs in UK educational research. What we should not do is unnecessarily conflate the issues of relevance, quality and funding. There is a danger that the perceived lack of rigour in current/recent research designs is being used,

perhaps unwittingly, to argue for greater political control of research funding, thereby ensuring relevance. The recent Hay/McBer study, and others like it, serve to remind us that use of numbers, political control, relevance to practitioners, and even lavish funding do not necessarily lead to high quality research or secure findings. This is not the way forward. Nevertheless the pressures are so great, the situation in UK educational research is serious enough, and the threats of irrelevance or political control of findings real enough, for the research community to take the job of increasing their own research-capacity seriously. Some would say that this may be the last chance for researchers to be allowed to 'police' themselves.

Notes

This work was completed as part of an ESRC grant (L139251106) aimed at auditing and building capacity for research into teaching and learning in the UK. A draft was presented at the BERA Hay/McBer Methodological Conference, London Institute of Education, May 2001

References

- Bridges, D. (1999) Educational research: pursuit of truth or flight into fancy?, *British Educational Research Journal*, 25, 5, 597-616
- Davis, A. (2001) Effective teaching: some contemporary mythologies, *Forum*, 43, 1, 4-10
- Foster, P. (1999) 'Never mind the quality, feel the impact': a methodological assessment of teacher research sponsored by the Teacher Training Agency, *British Journal of Educational Studies*, 47, 4, 380-398
- Gorard, S. (2001) *Quantitative Methods in Educational Research::The role of numbers made easy*, London: Continuum, ISBN 0826453074, 256 pages
- Gorard, S., Selwyn, N. and Rees, G. (2001a) Privileging the visible: Examining the National Targets for Education and Training, *British Educational Research Journal*, (forthcoming)
- Gorard, S., Rees, G. and Salisbury, J. (2001b) The differential attainment of boys and girls at school: investigating the patterns and their determinants, *British Educational Research Journal*, 27, 2, 125-139
- Gorard, S., Rees, G. and Selwyn, N. (2002) Work-place learning and the 'conveyor belt effect': an assessment of the impact of National Targets for Lifelong Learning, *Oxford Review of Education*, 28, 1
- Hammersley, M. (1997) Educational research and teaching: a response to David Hargreaves' TTA lecture, *British Educational Research Journal*, 23, 2, 141-162
- Hargreaves, D. (1997) In defence of research for evidence-based teaching: a rejoinder to Martyn Hammersley, *British Educational Research Journal*, 23, 4, 405-419
- Hay/McBer (2000) *Research into teacher effectiveness: Phase II report*, (mimeo via BERA office)
- Hillage, J., Pearson, R., Anderson, A. and Tamkin, P. (1998) *Excellence on research in schools*, Sudbury: DfEE
- Millett, A. (1997) *Speech to TTA Research Conference*, 5.12.97. London: TTA
- Mortimore, P. (2000) Does educational research matter?, *British Educational Research Journal*, 26, 1, 5-24
- National Educational Research Policy and Priorities Board (1999) *Investing in learning: a policy statement with recommendations on research in education*, Washington DC: NERPP
- 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
- Resnick, L. (2000) *Strengthening the capacity of the research system: a report of the National Academy of Education*, presentation at AERA Conference, New Orleans, April 2000
- Reynolds, D. (2000) *The pupil progress project: final report*, (mimeo via BERA office)
- Tooley, J. and Darby, D. (1998) *Educational research: a critique*, London: OFSTED
- TTA (2000) *The Teacher Research Grant Scheme: summaries from the second year of the scheme*, London: Teacher Training Agency
- Woodhead, C. (1998) Academia gone to seed, *New Statesman*, 26 March 1998, pp. 51-52