Educational Research and Policy: How can we ensure that our findings are used?

Sarah Maughan, National Foundation for Educational Research


Since the late 1990s the rhetoric about evidence-based or research-informed policy making has become much more widespread. The New Labour Government in England, in particular, is keen to state that its policies are based on sound evidence, collected through appropriate research. However, there is much debate about whether research should be designed to support the policy making process directly or whether a wider research focus is justified. Where it is decided that a research aim is to inform policy making, then the routes towards ensuring the findings are listened to, and used appropriately, are complex. There are few guarantees that research will have the intended impact, and in some cases it may even have negative or unintended consequences.

This paper provides a review of the literature on the impact that research has on policy and how this is achieved. The review will collect together and discuss existing measures of impact and attempt to draw conclusions about the extent to which research findings are used in policy development. It will also investigate the ways in which research is designed, or the particular features that are most likely to result in impact. Three case studies based on a range of NFER projects will be used as examples.

Introduction

The National Foundation for Educational Research (NFER) has been in existence since 1948, carrying out research projects to inform educational developments. NFER conducts around 200 research projects each year, spanning all sectors of education, from early years to lifelong learning. We aim to conduct high quality research, producing evidence for use by policy makers, managers and practitioners. Set against this background, NFER is keen to ensure that the projects that it undertakes have as much impact as is possible.

In the current economic climate it is likely that cuts in public sector spending will have an impact on research budgets in the coming years, so more efficient means of ensuring the same impact from fewer or smaller projects may be valuable. Whitty (2007) describes the evolution of research informed policy development, since 1997, under the current new labour government in England and Levin (2005) suggests that ‘interest in strengthening the impact and value of education research has been growing around the world, among governments and funders as well as scholars’ (p. 1).

The recently published Department for Children, Schools and Families’ Analysis and Evidence Strategy 2009-10: Delivering Evidence Based Policy (DCSF, 2009), suggests that the Department intends to continue its emphasis on using evidence to inform its decisions, and Ed Balls, Secretary of State for Children, Schools and Families is quoted as saying ‘An effective use of a robust evidence base is more important than ever before in helping us achieve our aspirations for children and families within a tight fiscal environment’. The document goes on to say that ‘the role of evidence and analysis is vital. It ensures that we set the right priorities and target our efforts and resources effectively’ (p. 5). This bodes well for the future role of researchers in informing policy developments, and continuing and expanding their influence.

However, that does not mean that all research should be designed specifically to have an impact on the policy making process. Over 30 years ago, Shavelson (1988) argued that ‘education research
can be justified as a legitimate inquiry in its own right’. He went on to say that ‘most of education research bears on theory or a particular line of empirical inquiry, as it should’ (p. 10, emphasis as in original). More recently, Black and Wiliam (2003) have argued similarly that not all educational research should be useful; they are supportive of the idea that research has a value in its own right, in part because we do not currently know what research will be useful in the future.

However, it could be argued that NFER is in a different position to some other research organisations, for example Higher Education Institutions or funded research centres, in that the majority of its work is commissioned on a project by project basis by government bodies, with the intention that it will produce findings that can be used for improving policy or practice. In these projects, it could be argued that NFER should use the ideas currently suggested in the literature for maximising impact, and design research with the specific aim of having as much impact as possible. A small proportion of NFER’s work is funded by other means, for example through its own research development fund. Perhaps projects funded in this way may be designed to meet other purposes, for example to develop new ways of thinking or new methodologies, without a direct impact on policy or practice per se. In general, however, for research organisations like NFER, producing research evidence that can be used directly by policy makers or practitioners may be of particular importance.

Against this context a number of key questions arise about what is meant by impact and how it is best achieved. These are discussed below.

What proportion of research has an impact?

Much of the literature on evidence based policy was reviewed for this paper, although it must be acknowledged that a systematic review of the literature was not conducted in the pure sense. One purpose of reviewing the literature was to gauge whether there is any agreement about the proportion of education research which does have an impact at the current time. Very few examples were found where an attempt has been made to quantify this impact. In fact, only one figure was found, Sebba (2007) quotes earlier research showing that only 70 out of 70,000 educational research projects in the US had a significant influence on policy (p. 136). Although no other examples were found against which to compare this, the figure seems very low. Perhaps this is due to the fact that the researchers were looking for a ‘significant’ impact, which raises the question of whether there is a consensus about what is meant by impact at all.

What is impact?

In the literature there does not seem to be a single agreed definition of impact that is being used across the different papers. Policy makers may work in a context in which they want research to say that if you do x, y will definitely happen. However, the nature of social science research is such that there is rarely this level of certainty. In this context it is important to ensure that there are realistic expectations about the usefulness of the evidence that can be obtained from research. Shavelson (1988) argued that having an impact may not be in the form of ‘immediate and specific applications’ but may rather lie in ‘constructing, challenging, or changing the way policy makers and practitioners think about problems’. He describes how ‘research can alert policymakers and practitioners to problems, increase their commitment to working on a problem area, support a position held, legitimate decisions already made, or be used to persuade others about a position held’ (p. 5). Similarly, Nutley et al. (2003) suggest that ‘research impact forms a continuum from raising awareness of findings through knowledge and understanding of their implications, to changes in behaviour’ (pp. 34-35). They draw a distinction between ‘the ‘conceptual’ ... value of research, which brings changes of knowledge, understanding and belief; and its ‘instrumental’ ... value in directly changing behaviour’ (p. 11). The literature they reviewed is focused mainly on impact in terms of practical change, and this is likely to be the kind of impact being considered in studies producing figures of impact such as that quoted by Sebba (2007).
How can we measure impact?

In addition to the lack of an agreed definition about what is meant by impact, there is some debate about the best means of measuring whether research has had an impact. Nutley et al. (2003) review the literature and state that ‘diverse methods for assessing different kinds of impact are reported’ (p. 16). In most cases they found that the measurement of impact was based on self-reporting or by observation of participants, they found no cases of ‘rigorous evaluation’.

Canavan et al. (2009) propose six propositions for successful measurement of research impact.

- commissioner – researcher partnerships: sustained interactivity between the different parties will allow a focus on impact measurement;
- the embedded research methodology: processes and tools for achieving research impact should be embedded in the policies and standards of an organisation and this is likely to ensure that the means of measuring the impact also become embedded;
- planning for impact from the outset: agreements should include expectations about the research impact and its measurement;
- adopting facilitative research methodological strategies: the methodology chosen should incorporate consideration of the best means of collecting evidence of impact;
- following good practice towards research impact: as impact probably depends on the relevance of the material, there should be a greater emphasis on the extent to which the research is fit for purpose;
- connecting research measurement to performance management: a commitment to measuring impact can be linked to existing performance management systems which would mean that measurement becomes a part of day-to-day work.

What features of research mean it is likely to have an impact?

Using the continuum idea of impact as given by Nutley et al. (2003), and assuming that this can be measured objectively in the ways suggested by Nutley et al. (2003) and by Canavan et al. (2009), it should be possible to find agreement in the literature about features of research that make it more likely to have an impact. A number of research articles suggest features, and although again there does not seem to be a consensus, there does seem to be a great deal of overlap in what is being suggested.

The review of the NERF Bulletin Trial (Sanders et al., 2005) found that teachers liked ‘relevant research content coupled with practical applications’ (p. 21), this content should be presented using ‘accessible language’ and should include ‘case studies of practice’ (p. 21). Visual features of a publication or website such as a clear front cover, use of colour or illustrations and a clear layout were all seen as useful features. However, these features related more to practitioners’ preference for means of dissemination, which in this case is taken to mean the likelihood that the practitioners will pick up and read research findings, rather than features which mean that the practitioners will use what they read to change their practice.

Interestingly, Sanders et al. (2005) also report that members of their focus groups commented that a recent change in culture in the teaching profession means there is now more of an emphasis on using research findings. This was linked with a perceived increasing professionalisation of teaching.

Nutley et al. (2003) propose a four-stranded strategy for increasing research impact, made up of:

- supporting interaction between research and practice;
- developing skills in researchers so they conduct research in a way that is most useful to its potential users;
- developing skills in policy makers and practitioners to use research to inform their work;
• developing the organisational context to support the greater use of research in practice.

Hemsley-Brown and Sharp (2003) also conducted a review of the literature and found, as many others have, that there is no direct relationship between systematic dissemination of findings and impact on policy or practice. They conclude that ‘the most significant factor affecting impact on practice in education is legislation because schools will adopt policies, whether or not they are evidence based. However, if such initiatives are not adequately funded, they are unlikely to be adopted.’ (p. 461)

Whether dissemination strategies are important in their own right is the subject of some debate in the literature, as although it may mean that the research is read, it does not necessarily mean it results in a change in policy or practice. However, Levin (2005) makes a convincing argument for the importance of dissemination. He describes in some detail the political context in which policy is made and the drivers which affect political decisions. An important aspect of this is the requirement for politicians to be re-elected. This is dependent on the views of the general public. In simple terms, this can depend on the public being interested in a particular issue, knowing about the research findings, and wanting these findings to be acted upon by their policy makers. Dissemination of the results widely through the media and with easily understood messages, is a key aspect of this.

How do we ensure research has an impact?

Nutley et al. (2003) review a large amount of literature from education, healthcare, social care and criminal justice. They draw conclusions about the types of processes that can be used to improve the likelihood of impact from research and summarise the four most effective as:

• active dissemination: tailored materials for targeted audiences;
• closer links between researchers and practitioners: establishing links through project specific and on-going relationships;
• facilitation of research use: through programmes of training and development for researchers and/ or practitioners;
• recruiting expert and peer opinion leaders: these people play an important role in raising awareness of and shaping attitudes to research findings.

The Department for Children, Schools and Families similarly describe different processes that they use to increase the likelihood that their research will have an impact: ‘the structure of the analytical community within DCSF – with many analysts embedded within policy Directorates and close working between analysts, policy makers and communicators – is the foundation for ensuring our work has impact.’ (DCSF 2009, p. 14)

In Harnessing Knowledge to Practice: accessing and using evidence from research, researchers from the Centre for the Use of Research and Evidence in Education (CUREE) bring together information about a number of different projects which aim to make research evidence easily available to practitioners. This publication provides a convincing case that much has been done over recent years to structure research findings in such a way that they have an impact on practice. It also emphasises a shared view across many of these initiatives about the best means of making an impact.

The evidence from the literature in response to this question suggests that there is some agreement about the research features that may make findings more likely to be used. As such, this may be the area in which most progress has been made and most has been achieved in the development of understanding about impact.

NFER Case Studies

Given this broader context set by the literature, what can be learnt from examples within NFER’s work? As mentioned above, NFER conducts approximately 200 projects each year, and there is a
great deal of variety within these projects. The projects can be grouped into a number of different
categories, each of which may require or intend to have different degrees of impact, and for some,
different means of ensuring impact may be necessary.

Some projects are very large, spanning a number of years, and producing a large number of
outcomes, some are small and defined, lasting only a few months; some projects are commissioned
with a specific aim of having an impact, others aim to gather background information in a given area
(which may ultimately have an impact), whereas still others are conducted using NFER funding,
perhaps to inform internal practices, or perhaps with the intention of having an impact on a policy
area. Three case studies of projects of different types are given below.

TIMSS

One very large project that NFER has been involved in for some time is the Trends in International
Maths and Science Survey (TIMSS). NFER has had a long involvement with the international surveys
and was a founder member of the International Association for the Evaluation of Educational
Achievement (IEA), which was set up in the 1960s and organises international comparative studies of
educational achievement.

The TIMSS study is an international survey comparing performance in maths and science across a
number of different countries over time. It is conducted by the IEA and sits alongside the Progress in
International Reading Literacy Study (PIRLS) tests which are reading assessments. The TIMSS
survey has been running since 1995 in its current form and England has participated from the start
(see Whetton et al. 2007 for a review of the studies and performance of primary aged pupils in them).

NFER currently administers the TIMSS study in England and Scotland, as well as fulfilling a similar
role for the PIRLS study, and the Organisation for Economic Co-operation and Development’s
(OECD) Programme for International Student Assessment (PISA) survey. As well as administering
the surveys, NFER also contributes to the development of the tests for TIMSS and PIRLS as part of
the international consortium.

The decision to participate in TIMSS is made by the government on a survey by survey basis, and the
research team to conduct the survey in England is commissioned by the Department for Children,
Schools and Families. The purpose of the survey is to assess the performance in maths and science
in the participating countries so that comparisons can be made about differences in standards
between those countries. Although the reasons why the government in England chooses to
participate are not published, the standard of maths and science in a country is seen as important for
the economic strength of that country, and outcomes from the survey can be used to inform
developments to maths and science curriculum and assessment with the aim of improving these
standards.

At the policy level, a great deal of impact can be traced back to the TIMSS survey results. In 1995
England did not perform as well as hoped in the maths strand of the survey which led to negative
coverage in the media. This may, in part, have contributed to the establishing of the national
numeracy strategy in the late 1990s. More recently, the results for England have shown some
improvements and Whetton et al. (2007), in a report written for the Cambridge Primary Review, state
there has been ‘a clear and marked increase in performance from 1995 to 2003 in primary
mathematics performance in England’ (p. 9). This can be used as evidence to support government
initiatives and dismiss charges that the various initiatives have not been successful. As Whetton et al.
(2007) say, it ‘would be hard not to attribute this change in mathematics performance to the influence
of the National Curriculum in England from 1989 and the associated Numeracy Strategy in the late
1990s, both of which formalised the requirements on teachers and perhaps raised their expectations
of pupils’ (p. 10). Performance in science in England was strong when the surveys were introduced,
and has continued to improve over time.
It is possible to relate a number of features of TIMSS back to theories proposed in the literature about impact. Canavan et al. (2009) propose that one way to increase impact is by building close links between the researchers and the research commissioners. In this case, this link is not a key factor in ensuring impact. The strict international guidelines for the way in which the survey is administered in each country leave little scope for the researchers and the research commissioners to work together to plan a research design that will achieve maximum impact, although there is close liaison to ensure that the survey is administered appropriately, and that the target sample is achieved (a requirement for international reporting).

Significant changes that have been made in classrooms in the teaching of maths and science could indirectly link back to the results from the TIMSS study. However, the impact is likely to have been as significant as it is because it has been legislated through the introduction of the national strategies, as Hemsley-Brown and Sharp (2003) suggest, this is a key factor in the use of research evidence by practitioners.

One of Nutley’s suggestions (Nutley et al. 2003) for increasing the likelihood of having impact is to tailor dissemination materials to different audiences. This is done as part of the TIMSS survey, with materials being developed for policy makers, practitioners and the press. In recent rounds of TIMSS there has been an active dissemination programme for the survey findings aimed at practitioners. This dissemination has included authoring of papers and presentations for conferences (one of which is being given at BERA this week), and conferences set up specifically to give the outcomes from TIMSS to teachers. In addition, there have been a number of press releases, as well as the required publication of national reports.

**Gifted and Talented Evaluation**

In 2003, NFER conducted a systematic literature review commissioned by the Local Government Association collecting evidence from the literature about different practices for supporting the teaching of gifted and talented pupils (White et al., 2003). This project set out to use the evidence to make practical recommendations about the successful processes which contribute to gifted and talented education, so could be said to have been designed to have a direct impact on policy and practice. The review of the literature concluded that there were different processes that contributed to gifted and talented education and these were categorised as: methods of identification, differentiation, enrichment, acceleration, classroom organisation, monitoring and evaluation, as well as the roles and responsibilities taken on by various individuals. However, the main conclusion from the literature was that there appears to be little unique practice in the area of gifted and talented education, as most of the strategies found apply equally to effective curriculum management for all types of pupils. The final project report made a number of recommendations for the future development of policy for gifted and talented pupils, which included the need for further research into what is unique in this area; the identification of evidence of effective practice; an exploration of the context in which effective practice develops and the consideration of what kind of discrete training might be useful for practitioners in this area.

The findings of the report were sent to all Gifted and Talented coordinators. However, although this is an effective means of dissemination, it does not necessarily mean that practice has been changed. This is an example of a project where, perhaps, more could have been done to ensure there has been an impact. It may have been useful to hold seminars with gifted and talented coordinators to discuss the findings, and their implications for practice. A more straightforward, but possibly less effective method would have been to develop guidelines for good practice based on the findings are targeted specifically at coordinators, and distribute these.
National Monitoring Surveys

A third project conducted by NFER is a small review of international and historical practice related to national monitoring surveys. In October 2008, it was announced that key stage 3 tests would be abolished and in their place a system of national sampling would be introduced. NFER was one of the lead agencies in the Assessment of Performance Unit (APU) in the late 1970s and 1980s, the previous national monitoring system in England which ended with the introduction of the national curriculum and the national tests. NFER staff are also very familiar with national monitoring systems in other countries and with the international surveys, and so it seemed valuable to use this experience and expertise to inform the introduction of any new survey in this country. NFER used its own funding to conduct this research.

The project involved reviewing websites and publications related to current national surveys in Scotland and America, current international surveys: TIMSS and PISA, as well as archived documents at NFER about APU. The information was collated and used to raise a number of key issues that should be considered when introducing a national monitoring survey, and the report made 17 concrete recommendations for a survey to be introduced in England at key stage 3.

Going back to Nutley et al.’s (2003) four ways of increasing impact, it was not appropriate in this project to involve policy makers in the research design process, although NFER did discuss the project with policy makers as work progressed. In particular we discussed the findings with a member of the Expert Group on Assessment, set up by the government to consider changes to national assessment.

The research findings were, however, carefully tailored to the needs of the target audience. The report was written in a concise manner, and also had a brief executive summary and upfront recommendations which could be picked up without an extensive investment of time. In addition, we presented the outcomes from the review at a seminar, set up by NFER, Cambridge Assessment and the Nuffield Foundation, and attended by key players in the English assessment world, assessment policy makers and the press. The 17 recommendations were provided as a handout at the seminar and picked up in a subsequent article on the BBC website. Outcomes from the seminar and the national monitoring report were published on the NFER website, and also sent to each member of the Expert Group on Assessment.

In terms of impact on policy from this project, we were told by more than one member that the Expert Group discussed our report in some detail. In the final report from the group, some of the small number of recommendations given reflected those in the NFER paper. However, the longer term impact of the paper is not yet known: plans for the national monitoring survey continue to be discussed, but no definite decisions about a way forward are known as yet.

Discussion and Conclusions

It is apparent from both the literature and from the NFER case studies that making an impact on policy or practice is not straightforward; there are many different kinds of research project, aimed at different audiences, and with different requirements for impact. There do not appear to have been many attempts to quantify the amount of research that does have an impact, and the one example that was found, in which Sebba (2007) quotes previous research suggesting that 70 out of 70,000 research projects in the US had an impact, seems unduly pessimistic. For the measurement of impact to become more systematic it would seem to be useful for the issue to be problematised in a way that would support future investigation.

For this to be possible, the first step needs to be to agree within the research community what is meant by the term impact: is it about directly influencing a particular policy agenda, or is it a much
wider definition, of the kind given by Shavelson (1988) or Nutley et al. (2003), in which a contribution to a debate is at least a recognised aspect of impact? There seems to be more of a consensus in the literature that it is the latter, albeit there is no single definition.

Assuming this is what is meant by impact, is there an agreed way in which this can be measured? Again, in the literature there does not appear to be a shared view of what this form of measurement might be. Canavan et al. (2009) suggest that it could be measured via performance management processes, although it is clear that this approach will not work in all cases, nor would this be appropriate for all of the different types of project exemplified in the NFER case studies. It would be invaluable to agree both a measure and a current view of the proportion of current projects that do have an impact, either within NFER or across the research community more widely, so this can be used as a benchmark against which future changes in approach can be evaluated. Without this it is difficult to be able to draw conclusions about which features of research really do improve the likelihood that a given project will have an impact. In simple terms, as an initial step, some kind of checklist could be agreed which can be used to record both the kinds of impact that is intended from a given project, and the kinds of impact that were actually achieved.

Without this quantification of current evidence, it does appear that some kinds of research are more likely to have an impact than others. As Hemsley-Brown and Sharp (2003) suggest, research that leads to a change in legislation is most likely to have an impact on practitioners, similarly research that has the backing of opinion leaders may well stand more chance of having an impact. However, beyond legislation there does not appear to be any guarantee that research findings will lead to a change in policy or practice. Other features can only go as far as making it more likely that the findings will be used.

In the literature, there are a number of attempts to suggest some features of research that do increase the likelihood that the outcomes will have an impact on policy or practice. Indeed, this is one of the areas in which there appears to be agreement, for example tailoring outcomes to the needs of different audiences is mentioned by Hemsley-Brown and Sharp (2003), Nutley et al. (2003) and Levin (2005).

These findings from the literature seem to be reinforced by the experience within NFER given in the case studies, where we found that directly tailoring the outcomes of the National Monitoring review to the needs of policy makers and ensuring that the right people knew about the research ensured that it was discussed, and presenting findings at a seminar of experts ensured it was valued. However, as mentioned above, it is not clear from the literature or the case studies which features have the greatest effect, as this has not been quantified. Similarly, it is not clear that all the features do have an effect, as many are difficult to isolate in practice. The case studies from NFER would suggest that again this area is not straightforward. There are a number of different types of research project, with different purposes and different audiences. TIMSS is commissioned by policy makers with the aim of having an impact, and is likely to have contributed to significant changes in practice through legislation. In this case, the active dissemination to practitioners, and materials tailored to their needs, is likely to have had a lesser impact.

For national monitoring in contrast, where legislation and working directly with research commissioners were not possible, the tailoring of the research findings to the needs of potential users is likely to have had a much greater impact on the potential for the information to be used. In addition, in this case, a key aspect is also likely to be the timeliness of the information, a feature that is not widely discussed in the literature (although it is mentioned in Hemsley-Brown and Sharp, 2003). It is clear that the means of ensuring impact for this kind of non-commissioned research are different to those required in studies such as TIMSS.

As both Whitty (2007) and Black and Wiliam (2003) suggest, it is not necessary for all research to have an immediate impact, and the NFER case studies demonstrate different requirements. The
evaluation of gifted and talented methods pulled together evidence from a number of different sources, but in the end recommended further work as much as directly changing practice and did not pursue an active policy of ensuring that its findings were used. However, it has contributed to the body of knowledge that will hopefully improve methods and practices over time.

In 2008, Caroline Sharp and Wendy Trury developed a useful conceptualisation of research impact in an internal NFER document. Using the diagram below, they suggest that there are different levels of impact and different ways of achieving these levels of impact. These different levels of impact may be appropriate to different types of research project.

The first level of impact reflected in the diagram is raising awareness about an issue via research. With this purpose in mind, the research needs to identify clear messages with something new or insightful to say. Publicising the information to the intended audience is of particular importance.

The second level is about contributing to knowledge, and changing beliefs and attitudes. In order to achieve this, a slightly deeper engagement is needed between the research and the intended audience. Some element of theory needs to be presented alongside findings, so that people can understand how different elements relate to one another. Ideally, people need an opportunity to consider the implications of the research for them (for example, by discussion following a presentation).

The third level needs a more sustained programme of involvement by the research users. This could include using research for consultancy, referring to research in training and encouraging practitioners to adopt new skills by providing training, opportunities to practice the skills and feedback.

The deepest and most difficult level of impact to achieve is to change behaviour. The best way to achieve this is to embed research information in coaching, training materials and/or policy documents.

Another point from the literature which warrants discussion is one made by Shavelson in 1988. He points out that much of the debate suggests that research findings reflect some kind of ‘timeless truth’ that once discovered holds and can be used at any point. In fact, research findings are part of an evolving picture, with new developments and new research continually adding to the overall image in a way that causes it to gradually change over time, or perhaps even change dramatically at a given
point in response to a major breakthrough in thinking, or a dramatic policy shift. Shavelson concludes that ‘much of what education research produces is, and probably will always be, time and context bound’ (p. 6).

To sum up, it is likely to be the case across different research projects, that there will be different requirements for impact, and different means of ensuring that this happens. The diagram produced by Sharp and Trury in 2008 gives a useful conceptualisation of this variability. A substantial body of literature is starting to emerge in this area, but there remains the need to develop a more systematic approach to the problem, with shared definitions of impact, and shared views of how it should be measured. Most important is the need to develop benchmarks of where we are now, that can be used to measure impact over time and evaluate the methods that are likely to have greatest effect. There is likely to be increasing pressure to conduct research with smaller budgets, whilst having the same or greater impact, so a clearer understanding of this area will be invaluable to demonstrate this value for money.

References


This document was added to the Education-line collection on 19 January 2010