Department for Education and Skills Research Conference 2004

Research in Education: What works?

department for
education and skills
creating opportunity, releasing potential, achieving excellence

Research Report: CR2004

Contents

DfES Research Conference 2004

DfES Research Conference 2004

The Department for Education and Skills held its fourth annual research conference on 19th November 2004 at the QEII Conference Centre, Westminster.

The day's theme was *Research in Education: What Works?* Keynote speeches were delivered by Robert Slavin from Johns Hopkins University and Sandra Nutley from the University of St Andrews. Robert Slavin discussed key issues in the development, evaluation and dissemination of the US Success for All programme, a comprehensive research-based reform model for primary schools. Sandra Nutley drew on the cross-sector work being undertaken by the Research Unit for Research Utilisation (RURU) at St Andrews to explore the delivery of research-informed practices.

The main part of the day was spent in a series of 12 workshops which covered the main DfES policy areas from children, young people and families to schools, further and higher education and lifelong learning.

This report contains a selection of the papers presented at the workshops. Where full papers are not presented, a summary of the workshop and discussion has been offered.

The 2005 conference will take place on Friday 25th November. Fuller details will be available on the research pages of the DfES website (www.dfes.gov.uk/research) in July.

Translating research into practice on a large scale: lessons from Success for All (US)

Keynote address by Robert Slavin, Johns Hopkins University

Our experiences in the United States with the Success for All programme – in terms of how government policy can help to introduce educational innovations that have been proven in rigorous research – goes to the heart of how research can influence practice at scale. Needless to say, this is not the usual process of innovation.

Change in education

The way change takes place in education falls into several categories, none of them good. One is pure faddism. For example, at the height of what we call 'whole language' and you call 'real books', there was a fad in the US for bathtubs in classrooms. You could go anywhere in the US and find bathtubs in primary classrooms. Why? They would be outfitted with pillows, and there would always be a child in there reading. It looked very nice. It was a lot of fun. There is nothing wrong with reading, and nothing wrong with bathtubs. But it was striking that within just a couple of years bathtubs were everywhere. A couple of years after that, you never saw another bathtub.

Another way that change typically takes place is the adoption of new approaches based on a theory – often a crackpot theory – but one with a positive view of children that people wish were true. When I was a special education teacher, long ago, there was a popular theory that children with special needs had somehow skipped steps in the development of neural pathways. The theory was that if you took them back to teach them how to crawl, and repeated the developmental steps they should have experienced, everything would then be fine. There was not the slightest shred of evidence to support this, but it was a very appealing theory. And so, every single day, the teacher in the classroom next to mine had all of her children – many of whom were rather large children with Down's Syndrome – crawling all over the place in hopes of moving them forward.

Change also takes place through direct commercial marketing. I have a friend who was the president of a large publishing company. They came up with what they considered to be a very innovative research-based book for high school biology, and had a series of focus groups to introduce this book. They brought together teachers and principals and school superintendents to look at the various books that had recently been introduced by various publishers, all set out on a table. One book from another publisher incorporated a fabulous innovation: it had a picture of a lizard on the cover, which was embossed so that you could feel the 'bumpiness' of the lizard. In focus group after focus group, people gravitated to the bumpy lizard. They felt it over and over again. They were quite serious, and ascribed to it all of the qualities it lacked, but which the other book actually possessed. The bumpy lizard book did fabulously in commercial terms. Did it make a difference for children? Only if you subscribe to the theory that somehow neural pathways are aided by feeling bumpy lizards; if not, then chances are it didn't make any difference whatsoever – but it certainly made a difference to the publishers.

Education also sometimes changes through well-meaning misapplication of good research. Educators often learn about the findings of good research in terms of isolated principles of practice. Then, rather than adopting something complete, well thought out and well designed, they implement their understanding of what that research is about. I have had a very long-standing and painful experience with this in the area of co-operative learning. Co-operative learning has long been known to have benefits for children's learning, but as early as 1980 it was known that those benefits are not automatic. They depend on the inclusion of a group goal and individual accountability. Put another way, co-operative learning works when children are working to teach each other, to help each other, to prepare each other for something like an external assessment. It demonstrably does not work when the children are asked to produce a single project or product, and one child does it while three children watch her do it.

However, if you visit classrooms in the US or here, you will find a very large proportion of teachers using some form of co-operative learning. Nine times out of ten they are doing exactly what for 20 years the research has clearly shown is ineffective. And yet, people cite the research and claim they are doing things in accordance with evidence, because it looks vaguely like what the evidence did support. And because, frankly, it offers a very positive view of children and how they learn. It also keeps the kids busy and out of the teacher's hair, so there are multiple side benefits.

Research into practice is rarely a simple translation.

Translating research into practice

I want to talk to you today about one of the best cases in which good research actually has been translated into widespread practice, about the history of this development, the policies that promoted it, and how these policies could work if you were to do something of this kind again.

Our Success for All programme is a primary school intervention model to improve literacy achievement in high-poverty schools. It was developed in part as a strategy to make the research-to-practice link much more solid than it had been. Basically, the programme came about because we carried out research reviews on each of many elements about instruction, curriculum, classroom management, parent involvement, assessment, and so on. We included in the programme the elements that we felt had the strongest evidence of effectiveness and, over time, we have continued to update and change the programme in light of our developing understanding of what is most effective in teaching children to read or teaching them other subjects. You might think of it as a sort of chocolate-box assortment of effective practices.

Success for All began in 1987, in one school in inner-city Baltimore. Today it is in about 1,400 schools in the US, serving about 800,000 children in 47 states. There are also 28 Success for All schools in England. It is an interesting model in terms of both incorporating research and being able to scale up to a substantial degree. It is also interesting in terms of the amount of research that has been done on the programme itself. More than 50 experimental controlled comparisons of Success for All have been done in the US, and a handful in Britain as well. A large-scale national randomised experiment evaluating the programme is now underway in the US. So it has been quite a story, not only in terms of the application of research, and research on the product of applications of research, but also in terms of dissemination.

How did this happen? And how should it have happened? Success for All came out of a group at Johns Hopkins University that had been working on co-operative learning and other instructional strategies at the classroom level in the 1970s and 1980s. We had begun in the 1980s to assemble the principles learned from this work into programmes for classrooms. Initially, we created a mathematics programme for elementary schools that incorporated co-operative learning and other principles of practice, and later programmes for the teaching of reading and writing.

In the mid-1980s, we created something we called the 'co-operative elementary school' that began to put these together, and to deal with some of the issues of school organisation and school-wide practice.

In 1986, we were approached by the Baltimore city schools. Baltimore is a large, very impoverished city near Washington, DC, with many children who are at risk. We were approached by the then Superintendent of Schools, who asked us – as the fancy university in our very impoverished town – 'What would you do if you had the opportunity to do whatever you thought necessary so that every child was going to succeed from the beginning of their time in school?'

The Superintendent had previously been the principal of an elementary school, and so she understood very well the progression that takes place for so many at-risk children. Children come into kindergarten (or reception), bright-eyed and eager to learn, full of ideas and enthusiasm, and smart as can be. And then, just a few years later, the very same children are in special education or falling behind or in serious difficulty. So her question to us, and our question to ourselves, was 'How could you stop that progression? How could you have children start with success, and then move from success to success?'

In taking on this challenge, we began to review the literature in all areas that we thought were going to be important.

We looked at research on reading methods, on reading content, on pre-kindergarten and kindergarten programmes, on assessment, parent involvement, classroom management, what you can do with children who are falling behind despite good instruction, which meant we looked at the tutoring literature. And we looked at ways to keep children out of special education, and to help them if they are in special education, and so on. Out of this, we assembled the model, which we then trialled and evaluated. The results were astonishingly good, in terms of making a substantial difference in the reading performance of children in some of the most impoverished schools in Baltimore.

We then revised the programme in light of our experiences and gradually expanded it, first within Baltimore and then in other districts. This is an important principle of a scale-up: don't put all your eggs in one basket. We began to work in other cities and to respond to additional needs that became apparent as we began moving out from our base. For example, the Baltimore schools are overwhelmingly African-American, so almost everyone speaks some form of English. In Philadelphia, we found many children who spoke Spanish and other languages, so we then had to develop programmes to respond to children who are not speakers of English.

We recognised the need to develop programmes in writing as well as our reading base. Later we developed programmes in mathematics, and in Spanish (because many children in the US are taught in Spanish initially and then transitioned to English), as well as programmes to help children taught in English who are not speakers of English.

The policies that made Success for All possible are worth some attention. First, we had a research centre at Johns Hopkins that had funded our work over an extended period of time. Almost uniquely within the United States – and I'm sure it would be unusual in Britain – our people did research full-time, rather than as something they did when not teaching classes. This is important, because it is very difficult to do demanding field research and development work when you are also trying to do many other things. We had the opportunity to pursue big ideas rather than just produce the next publication to get tenure. We also had some limited funding from national foundations to help us develop some of our initial prototypes.

In 1991, something very significant happened. By then we were working in about 30 schools, and pleased at being able to work at scale. There are, however, 100,000 schools in the US, so 30 is not a significant number. But in 1991, the first Bush administration facilitated the assembly of a group of business leaders, who decided to do something serious about education reform. They formed the New American Schools Development Corporation, later shortened to New American Schools, and ultimately raised about $250 million from corporate sponsors. A lot of this money initially went to help various developers create effective comprehensive models to reform schools in high-poverty areas. Basically, they were saying 'Let a thousand flowers bloom – you can have whatever model you think will be most effective. We'll provide the resources to develop and pilot effective programmes, and then help with the process of scaling up and building up an infrastructure to support this kind of work.' This is necessary, because after you have proven something to be effective in your research, you have to acquire some of the wisdom of operating a small business. You have to deal with issues of capital and personnel and other things that academics typically don't want to have anything to do with. So we had advice and assistance to help us deal with those kinds of issues.

New American Schools initially funded eleven programmes and then, by design, winnowed out programmes that didn't seem to be going anywhere, and ultimately settled on seven that seemed to have potential. This happened in the early 1990s, and six of those seven programmes are still operating today, and collectively are serving about 3,000 schools across the US. This is really quite an extraordinary, unparalleled process, in terms of investing in the research and development process at the front end, and then following it all the way through.

There was a big mistake, however, in the way that New American Schools proceeded. They did not insist that the programmes carry out research on the effectiveness of their programmes. They just assumed that, as they were wildly effective, any possible way of looking at them would show wonderful results. Therefore, in some ways, the programmes were developed to look effective, but because they weren't subjected to that kind of test, it affected both the amount of research done on most programmes, and the way in which many programmes were designed. We were the exception. Because we were in a university setting, we felt this research was important, so uniquely among these seven, our programme built up a substantial set of valuable data.

In 1997, something came along that also made a huge difference. The US Congress established the Comprehensive School Reform (CSR) demonstration. This provided money to individual schools to enable them adopt programmes that were 'both comprehensive and proven'. Unfortunately, because of the lack of evidence about many of these programmes, there was much more emphasis on 'comprehensive' than on 'proven'. The latter standard would have limited investment to two or three programmes (mostly ours). CSR provided schools with $50,000 a year for three years to choose whatever comprehensive programme they thought was most beneficial to their particular circumstances. There were extra points for applications from high-poverty schools and, as a practical matter, almost all of the schools that have adopted these programmes have been in poverty areas.

The government also provided funding for development and capacity-building to help existing and new Comprehensive School Reform models establish themselves, become more effective, and learn how to work at scale. This was basically seed money, seeding the market. The amount of money provided was not enough for most programmes, but the concept was that schools would bring in other funding to enable them to implement the various models. At state level, design fairs were held to which schools sent representatives to learn about the programmes, and to make an informed choice among them. That process in itself was exciting: schools coming to grips with what they need, what is most important to them, their values and capabilities, so that they could choose programmes they felt could help them to achieve better results with their students.

In about 2000/01, the whole movement hit a plateau. Part of the problem was complaints about the lack of research for programmes other than Success for All, Direct Instruction and one or two others. And part of it was a pendulum swing away from site-based management, towards much more control by the district or the LEA. The Comprehensive Reform movement depended substantially on the idea that individual principals (or head teachers) would review the various programmes with teachers, and make a choice consistent with their philosophies and their needs. That might be a different choice from those made by other schools, even in the same area. The movement towards much more control by superintendents and much more consistency across school districts conflicted with this concept. In our own programme, as in most of these programmes, we have insisted that the teachers vote by secret ballot, with a positive vote of at least 80% in favour of taking on the model, so that there would be a clear mandate by the staff. That is very difficult to do when a school superintendent demands that everyone does the same thing.

We are now operating under a policy called No Child Left Behind, which has greatly changed the landscape of education in the US. It has got a wonderful element to it, which is of course honoured more in the breach than in the observance. Nevertheless, it is wonderful to have the words there: something like 'based on scientifically based research'. It's not Shakespeare, but that formulation is in the No Child Left Behind Act 110 times. It refers to the kinds of programmes that should be used under any kind of government funding, and it means that schools should focus on programmes based on scientifically based research.

However, the policy-makers then asked what was meant by 'based on scientifically based research'. It can mean two very distinct things. It can mean that the principles underlying this practice are scientifically valid, or that the programme itself had been rigorously researched. Many have pointed out that 'based on scientifically valid evidence' is something you could drive a lorry through. It is a very 'squishy' sort of definition. Anybody can claim that anything is based on some rigorous research that was done somewhere at some time. If, however, you required that the programme itself was rigorously evaluated, you would require the entire nation to use either Success for All or a programme called Direct Instruction. That was unacceptable, so they went with the squishy definition. This has been disastrous, because it has allowed each state to decide what is research-based. And the states do what they always do – they go for the bumpy lizard on the cover. Full circle, but worse. At national level, the commercial publishers are not terribly powerful. At the state level they are extremely powerful because, not to put too fine a point on it, they are able to take people out to dinner, to form personal relationships, and this becomes the basis on which education innovation really takes place.

There are some positive stories about this as well.

The No Child Left Behind Act, which was intended to focus schools on scientifically based programmes, in fact ended up favouring the traditional textbooks that have always dominated practice in US schools. Is there rescue on the horizon? Well, maybe. The Institute for Education and Sciences (IES), the research branch in the US Department of Education, has a most extraordinary leader. Russ Widehurst truly believes that science must genuinely be the basis for educational practice, and somehow he has talked the Bush administration into letting him promote that idea. He has emphasised, in particular, the use of randomised experiments as a basis for rigorous evaluation, and he is trying to move toward requiring that programmes have been rigorously evaluated – not merely based on scientific principles, but actually rigorously evaluated in randomised experiments.

Why randomised experiments? There are, of course, methodological reasons why randomised experiments are preferable to other forms of evidence, but there is also an extremely important political reason. If you are going to ask Congress for money comparable to the funding that goes into research in medicine or agriculture, you'd better have methods as rigorous as those used in medicine and agriculture.

Education research in the US, as in England, doesn't get a lot of respect. And therefore it doesn't get a lot of funding. Because it doesn't have much funding, it doesn't get much respect. And because it doesn't have much respect…

The concept was to break that cycle, and demand respect, by coming forward with randomised experiments of the highest quality that would qualify in any field as rigorous and conclusive. And so IES has begun to fund a large number of studies that use randomised designs, studies of practical programmes and practices in reading, in mathematics, in science, promoting graduation, even character education, and many other fields. The intention is that, as these studies begin to bear fruit, tighter definitions of what it means to be scientifically based will help to promote the widespread adoption of programmes that have strong evidence of effectiveness. A key aspect of this is the What Works Clearing House. This is a facility that reviews research in many areas on replicable programmes. Their first report, on mathematics in middle school, has just come out, and several more are due during 2005. We believe this will fundamentally change the discussion

about what works and about the relationship between research and practice. The reviews are intended to be definitive, using extremely high standards of evidence, and because it is supported by the government, it will be seen as noteworthy. It will affect policy.

Will this actually happen? It might. There are many many forces in opposition to it, but it could very well happen, and something like this has to happen sooner or later, if we are truly going to move education forward, and move good science into good practice.

So how do you do it right? If I had the opportunity to start all over again, for example in Britain, I would do the following:

- Fund development of many replicable programmes to solve the key problems in your education system;
- Engage in a process of identifying the problems;
- Identify the most important of these that can be solved through research;
- Invite proposals, perhaps from fields other than education, from anyone who has a solid idea, and a good theoretical reason to expect that something might be effective;
- Provide them with at least seed funding to begin the process of developing and piloting, and evaluating programmes.

I would have the expectation that:

- Programmes would be ready for evaluation in as few as three years;
- Funding should be sufficient to enable concentrated attention on each programme, so that each can be ready for evaluation in three years;
- Randomised evaluations would then be funded, and carried out first by the developers themselves if they are proficient in research, but then by third parties.

Within a reasonable period of time, we would know how effective these programmes are and which of them should be taken to the next level. Once proven programmes have been developed and researched, it will be important to provide schools with resources, like the Comprehensive School Reform resources, although perhaps more tightly focused than a whole-school reform model. The only criterion should be 'does it work?' Programmes with strong evidence of effectiveness, that respond to national priorities and standards, should be made available to schools without telling schools exactly which to choose. Part of the genius of the Comprehensive School Reform model was that schools weren't told to use any particular programme; they were told to use something with strong evidence of effectiveness. 'If you want to do something that we've never heard of, show us the evidence, but choose it based on your needs and your philosophy.'

The next step would be large-scale third-party evaluations. In the US we're promoting the concept of awarding grants to schools in pairs: each pair of schools comes forward from a given LEA, and it's decided at random which one gets to use the programme right away, and which has to wait for a year. In the process of funding schools to use these programmes, you would thus be building large-scale, very high-quality evaluations. This process will continue indefinitely. There are always new ideas coming forward and new problems to be solved.

Government needs to develop a commitment over a long period to the constant development and evaluation of effective practices, moving them into as many schools as possible, and constantly improving the quality of programmes with strong evidence of effectiveness that are made available to schools.

Delivering research-informed practice: lessons from the crime reduction and social care fields

Keynote address by Sandra Nutley, University of St Andrews

This talk draws on the work of the Research Unit for Research Utilisation at the University of St Andrews, which is funded by the Economic and Social Research Council. The research unit is a cross-sector resource on research utilisation and it has contributed to several reviews of research use in different sectors, including a review of models of research impact for the Learning and Skills Development Agency. More recently, we have been involved in reviews that have investigated the improvement of research use in social care, and the implementation of the Crime Reduction Programme. Lessons from the latter two reviews are the subject of this talk.

The research unit seeks to promote cross-sector learning, particularly across the boundaries of education, health care, social care and criminal justice. Put very simply, we are trying to understand two things: Why does it take practice so long to adopt best evidence? And what can researchers do to increase research impact?

Improving research use in social care

To pull together some lessons from the social care and criminal justice fields, let me turn first to the review on improving the use of research in social care. We conducted this review, in conjunction with Barnardo's Research and Development Team, for the Social Care Institute for Excellence (Walter et al 2004). It was predominantly a literature review, which was supplemented by consultation seminars and some interviews with key people in the field in order to capture emerging activities and knowledge which may not yet feature in the published literature.

We found a great deal of activity around trying to increase research use in social care and a general commitment to the idea of research use was evident, even though in some people's judgement current practice in social care is not well informed by research. We grouped the activities to promote research use under seven main headings:

- **Ensuring a relevant research base:** developing a research base that can inform policy and practice in social care by commissioning research in specific areas and synthesising the results of existing research.
- **Improving access to research:** including efforts to improve research dissemination channels and the provision of user-friendly databases of research findings.
- **Making research understandable:** an emphasis on moving from often rather turgid research reports to accessible, easily understood summaries of research.
- **Drawing out the implications for practice:** for example Barnardo's 'what works for children' series of 'evidence nuggets'. These are readable summaries of research findings that also draw out implications for practice. The production of guidelines for practice is also included under this heading.

- **Developing best practice models:** this involves developing pilot or demonstration projects based on findings from research.
- **Requiring research informed practice:** some initiatives require that individuals and organisations demonstrate research-informed practice by specifying this in job descriptions and appraisal systems.
- **Developing a research-minded culture.** The areas of activity listed above are very much about trying to get specific lessons from research into practice. This category is much more about selling the idea of research use, changing people's attitudes to research rather than getting a particular research finding or product adopted.

As you would expect, many of the initiatives aimed at increasing research use combine activities from several of these categories. In order to understand these combinations, we explored the ways of thinking about research use that appeared to underpin them. In doing so, we developed three models of research use in social care, which seemed to capture the very different ways in which people were thinking about research use and how research use could be developed. These are: the research-based practitioner model, the embedded research model, and the organisational excellence model.

The research-based practitioner model

Here it is very much the role and responsibility of the individual practitioner to keep up to date with research. The model grew out of the evidence-based medicine movement. Research use is seen as a linear process: the research exists; you then need good systems for accessing it, appraising it, and applying it in day-to-day practice. Because the individual is responsible for the use of research, it is individuals that you target. It assumes that individuals have professional autonomy - that they can change their practice in line with research evidence, if they've accessed it and think it is appropriate. In putting this model into practice, the emphasis is frequently on initial professional education and continuing professional development.

The embedded research model

In this model, individual practitioners may remain totally unaware of the research that is guiding their practice. Research instead becomes embedded in the systems and processes within which they work and in the practice tools that they use, such as assessment checklists. The people who are responsible for embedding research in those systems, processes, guidelines and tools, work in policy roles at national or local levels. Hence the responsibility for research use lies with policy-makers, and to an extent with managers at a local level. Research use is still viewed as a linear process: the research exists, it is converted into systems and tools, and then applied. To encourage the use of guidelines and tools, performance management and scrutiny systems may be put in place.

The organisational excellence model

Here the responsibility for research use lies with local service delivery organisations. The leadership and management of these organisations are seen as the keys to better research use. What is envisaged is not a linear view of research use, where the research exists and is then applied in practice. Instead, research use is viewed as an ongoing learning process, where research is commissioned locally and adapted locally. An important aspect of the model is developing a research-minded local culture and to promote this, local service delivery organisations often work in partnership with universities and other intermediary organisations.

The organisational excellence model suggests that it is unhelpful to have a strict separation of roles, where research experts generate evidence and practitioners use that evidence. Instead, the idea is to work in partnership, where the whole process of knowledge creation, validation and adoption is shared between researchers and users. The 'Success for All' programme described by Robert Slavin fits within an organisational excellence model, certainly during its early years.

Given these three models, a crucial question is whether there is any evidence that one model is better or more appropriate than another for improving research use. We found that there was not enough evidence to suggest that any one model was better than the others. Furthermore, it seemed unlikely that any one model would ever be preferred over the others. The models seem to be suited to different situations, depending on issues such as the type of research evidence and the sort of staff group involved in applying that research evidence. The social care workforce is very diverse, 80-90% of whom are not professionally qualified, and what may be an appropriate model for professionally qualified staff may not be best suited to other staff groups. Hence there is a need to match research use models to circumstances.

If no one model fits all circumstances, there is a need to think about how one combines the three models. In the review, we did this by considering the whole system of social care and what each part of this system can and should be doing in relation to planning, undertaking and using research. Our conclusion was that there is a need to think holistically about research use across the system, without being totally wedded to a particular model of research use.

Promoting research use in the crime reduction field

I now want to turn to the review of the implementation of the Crime Reduction Programme, as this programme was an attempt to engage a wide range of actors in whole-system change and in a more evidence-based approach to crime reduction.

The Crime Reduction Programme was incredibly ambitious. It ran for four years, but was originally planned as a three-year, £400 million cross-government commitment to reducing crime, beginning in 1999. It was announced grandly as the biggest single investment in an evidence-based approach to crime reduction that has ever taken place in any country. Here was an attempt to get research to influence practice at scale.

This was to be a programme that would use existing evidence on what works, would develop that evidence base and would, simultaneously, deliver reductions in crime and meet government targets, such as a 30% reduction in burglary rates. The process started with a review of the

existing evidence base – a review of what was known about ways of reducing crime. This evidence was sifted into areas where there was strong evidence, areas where evidence was promising, and finally areas of weak or little evidence. On the basis of this evidence base, the programme sought to develop a portfolio of initiatives, which involved larger-scale testing where there was reasonably strong evidence, smaller-scale testing where there was promising evidence, and, innovative, evidence-building projects where there was weak or little evidence.

Eventually some 1,500 projects were funded under the Crime Reduction Programme. These fell into five main categories:

1. Working with families, children and schools to prevent young people becoming offenders of the future;
2. Tackling crime in communities, particularly high volume crime such as domestic burglary;
3. Developing products and systems that are resistant to crime;
4. More effective sentencing practices;
5. Working with offenders to ensure that they do not re-offend.

A ring-fenced budget of £400 million, was established to move the programme forward. This budget funded both the implementation and the evaluation of the portfolio of initiatives. Unlike the example cited by Professor Slavin, there was a major commitment to evaluation: 10% of the budget was set aside for evaluation purposes, from the outset.

Implementing and evaluating the programme was a twin-track process. On the one hand, the process involved engaging delivery organisations in the process of bidding for funds and getting things happening on the ground. On the other hand, there was the development of an evaluation strategy, which was itself very ambitious. The evaluation process was to assess not only the effectiveness and cost-effectiveness of projects but also consider how successful initiatives could be transferred to mainstream programmes and at the sustainability of any benefits in the longer term.

The plan was that as a result of this twin-track implementation and evaluation process, initiatives would be refined during the course of the programme. At the end of the programme, the accumulated learning would feed through to the mainstream programme and it would also extend the existing evidence base.

Implementation challenges

Clearly, a lot could be said about the successes and the failures of the Crime Reduction Programme, but that is beyond the scope of this talk. My focus here is on the challenges of implementing such a programme. Four of these are highlighted in the review report (Homel et al 2004):

- Translating research into practical projects
- Managing programme implementation and coherence
- Delivering financial and other resources
- Effective processes for evaluation, learning and change.

The first and last of these challenges are particularly relevant to this discussion, so I shall concentrate on those.

Translating research into practical projects was an enormous problem for the programme. One should never underestimate the time it takes to put together an evidence-based bid, especially when there is no seed funding available for developing bids. Delivery agencies, such as police services, needed to bid for funds to undertake programmes initiatives, such as targeted policing. However, in developing those bids, the process of translating what research really means – what needs to happen on the ground – was difficult. In seeking to address these difficulties, the programme took two main approaches. Firstly, it commissioned researchers (and subsequently NACRO) to work with prospective bidders on developing evidence-based bids and to advise on the implementation of successful bids. This approach is in line with the organisational excellence model described above. Secondly, there was a high degree of specification from the centre. In the case of the reducing burglary initiative, this entailed the development of a detailed prospectus for delivery agencies to bid against, and subsequently the central development of the Crime Reduction Toolkit – a toolkit designed to support the translation of research into practice. The second approach, central specification approach, echoes the embedded research model referred to above, where practitioners do not need to engage directly with research findings because others have already translated these into practice tools. This centrally directed, embedded approach became the predominant approach to research translation within the Crime Reduction Programme.

Because of the enormous problem of translating research into practical projects, it is not surprising that there was generally underbidding for projects. At the end of the programme's initial three-year period, it had spent only 65% of the total funding available. Although there were many reasons for that, it was due in part to the difficulty of translating research findings into projects on the ground.

The challenge of developing effective processes for evaluation, learning and change was another difficult area for the programme. In particular, it proved difficult to balance 'in-programme' and 'from-programme' learning and in the event, the balance was very much in favour of 'from-programme' learning. The evaluation strategy was constructed on the basis that evaluators should be independent of the projects they evaluated. Consequently, communication between evaluators and the projects they were evaluating was limited in the interests of objectivity. This created problems for 'in-programme' learning.

A great deal of knowledge was nevertheless generated during the Crime Reduction Research Programme. The challenge was to accumulate and store that knowledge and get it to impact on mainstream practice. The Home Office is still trying to get to grips with developing an appropriate knowledge management system, so that the effort put into building up evaluation knowledge does not just dissipate within the system. This is not just knowledge about 'what works' but also 'know-how', which often gets lost when there is a high turnover of staff, as was the case in this programme.

Another learning challenge I want to mention is how to ensure that longer-term research questions are addressed. Most of the programme's projects were only funded for 12 months and many of these projects were only just finding their feet when their funding came to an end. As a result there were many stunted evaluations and lost opportunities for addressing important longer-

term research questions. Because of this, the review report recommends that action to research and develop the evidence base should be undertaken as a separate but related activity to the delivery of a policy programme.

Finally, I want to comment on the funding that was to have been allocated to innovative projects in order to build knowledge in areas where the existing evidence base was considered to be weak. In the event, very little funding was actually given to innovative projects. There is much that needs to be learned about how one ensures that innovation survives in a centrally-driven and top-down process of developing and delivering evidence-based policy.

Conclusions

What lessons can be drawn from these two reviews of social care and crime reduction? A whole-systems approach to research use makes sense. Individual researchers and practitioners do not operate in isolation; they are constrained by the systems within which they work. So, if we are going to encourage research use, we must consider what needs to happen across a whole system of activity. In planning activities, we also need to think about an appropriate balance between different models of research use, such as the three models outlined above. Activities underpinned by different models of research use are likely to be required at different times and in different places.

Seeking to direct whole-system change from the centre, as happened within the Crime Reduction Programme, poses major challenges. One should not underestimate both the investment and the time that this takes. A centralised, top-down approach to system change appears to over-emphasise the embedded research model, to the detriment of other models of research use. The embedded research model runs the danger of becoming too mechanistic if used wholesale. In thinking about system change, 're-engineering' is not the metaphor we need – we are not dealing with mechanical systems. Instead, we should be thinking about how one works with and adapts to changes in living systems.

References

Homel P, Nutley S, Webb B and Tilley N (2004) *Investing to deliver: reviewing the implementation of the UK Crime Reduction Programme,* Home Office Research Study 281, London: Home Office

Walter I, Nutley S, Percy-Smith J, McNeish D and Frost S (2004) *Improving the use of research in social care.* Knowledge Review 7. London: Social Care Institute for Excellence/Policy Press

Methodological issues in the evaluation of interventions to promote parenting and family support – the case of home visiting

Jane Barlow, Reader in Public Health, University of Warwick

The authors of the recent systematic review entitled *What works in parenting support: the international evidence* (Moran and Ghate 2004), were forced to conclude that we don't currently know what works in a UK context because there is insufficient rigorous evidence available from which to draw any firm conclusions.

As a researcher working in the field of parenting support, I found this conclusion to be fairly salutary, and having joined in the call for more rigorous evaluation of parenting and family support interventions, it might not have been unreasonable to hail the randomised controlled trial (RCT) that is the subject of this paper as just such a beacon of rigour. However, I am going to use the findings of this study, alongside those of other similar UK evaluations of home visiting programmes, to explore the problems of using this method to evaluate complex community-based interventions.

This paper addresses some of the methodological issues involved in the evaluation of interventions to promote parenting and family support, by describing an RCT we have been conducting to evaluate the effectiveness of an intensive home visiting programme. The results of that study are used to demonstrate some key methodological issues regarding the evaluation of parenting and family support programmes more generally.

Background to the research

While the use of structured and intensive home visiting programmes for pregnant and new mothers has been increasing in many countries over the past two decades, the UK has been cutting back on one of its main providers of home visiting – health visitors. The need for the further development and evaluation of home visiting in the UK has, however, recently been acknowledged by a number of important sources. Professor Acheson (1998), for example, clearly viewed home visiting programmes as an important factor in the fight to reduce inequalities in health, and at a recent public health conference, he lamented that he was disappointed not to see further development of regular visits by health visitors to mothers in the home.

The recently published Health Development Agency (HDA) review of reviews (Bull et al. 2004) showed that home visiting programmes can be effective in improving a range of important outcomes for mothers and babies. However, it concluded that there is a clear need for more well-designed studies to build the UK knowledge base regarding their effectiveness in this country. In particular, it cited the need for studies to address the methodological limitations identified in existing studies.

Funding bodies are also beginning to address this lack of evidence. There have been four large-scale evaluations of the effectiveness of home visiting programmes in the UK during the past few years:

- In 2000, the British Medical Journal published the findings of an HTA-funded randomised controlled trial of the effectiveness of community post-natal support workers (Morrell et al. 2000);
- The Joseph Rowntree Foundation funded an evaluation of Home Start, the national scheme providing volunteer based home support to vulnerable families (McAuley et al. 2004);
- The Health Technology Assessment (HTA) has published the findings of a second study, comparing supportive listening visits by specially trained health visitors with eight community-based interventions and a control group (Wiggins et al. 2004);
- Finally, we are about to publish the findings of an RCT of the effectiveness of an intensive home visiting programme, funded jointly by the Department of Health and the Nuffield Foundation.

All of these studies have two important things in common other than the fact that they are the first large-scale studies of home visiting programmes to be conducted in the UK:

- They all had to address some very difficult methodological issues;
- They all had problems demonstrating effectiveness.

Methodological issues

The following are just some of the many methodological problems of existing studies of home visiting programmes highlighted by both the HDA and the HTA reviews:

- No random allocation
- Small numbers – insufficient power
- Non-standardised measures
- Unblinded assessment of outcomes
- Multifaceted interventions – which part produced the outcome?
- Too many outcomes – Type 1 errors
- No detail about intervention
- No process outcomes assessed
- Substantial/differential attrition
- No theoretical framework
- Limited results reported
- Surveillance bias.

The list is rather long and includes all of the cardinal sins associated with the evaluation of interventions to promote parenting and family support more generally. Our four home visiting studies showed varying degrees of success in avoiding these. This was never going to be an easy task, however, because some of these methodological issues are rather difficult to avoid when evaluating complex, community-based interventions aimed at supporting parents. For example, while RCTs are currently accepted as the 'gold standard' in research, there are in reality many problems associated with the use of this methodology to evaluate family support programmes, including practical difficulties of randomising participants, the fact that preference may play a part in intervention effectiveness, and the possibility of 'contamination' in which the control group are affected by some aspect of the intervention following randomisation.

These are not just issues for the research community. When, for example, we interviewed women who had refused to take part in our home visiting study, we discovered that the process of randomisation had itself been an important reason for some women refusing to take part:

> *Not to say to me – we'll dangle this in front of you and you may or may not be able to have it. I mean, that's ridiculous … not say to a vulnerable person, we've got this super duper help that you **might** be able to have, but there's only a 50:50 chance that you can. I just thought, stuff that …*

Another woman felt that she couldn't receive such a personal service from someone that she had been randomised to, and whom she might not like.

> *… this is probably one of the most personal things that there is, and if you don't feel there is a rapport with the person then you wouldn't open up and there'd be nothing gained.*

Other methodological problems include:

- Difficulties in operationalising broad aims such as 'promoting better parenting';
- How to measure outcomes such as 'parenting capacity';
- How to assess whether the intervention has prevented outcomes such as abuse, given the issue of surveillance bias;
- The fact that consequences of home visiting may take time to be realised;
- Difficulties in knowing which part of the intervention contributed to which outcomes.

Despite these problems, we set about trying to conduct a rigorous evaluation of the effectiveness of a professionally delivered home visiting programme here in the UK. The service we developed was built on existing evidence about what works, in terms of content, frequency and duration.

The home visiting service

In January 2001, we established a new home visiting service at 40 GP practices across Oxfordshire and Buckinghamshire. The service comprised weekly visits to women identified by midwives as being at risk of poor parenting, including abuse and neglect. The visits began during

the second trimester of pregnancy and continued for 18 months. They were conducted by health visitors who had received training in the Family Partnership Model combined with strategies for enhancing parent–infant interaction.

When we were thinking about how to go about intervening with families, we were clear that the intervention had to meet a range of criteria. It had to be 'home grown' in the UK, and be capable of being facilitated by generalist NHS professionals such as health visitors. We needed a programme for which there was preliminary evidence of effectiveness for outcomes such as parental mental health, self-esteem, relationship with partner, and children's emotional and behavioural well-being. It also had to be modifiable to meet the needs of vulnerable families, with mechanisms that would be effective in terms of the primary prevention of abusive or neglectful parenting.

The Family Partnership Programme (or Parent Advisor Programme as it was called then) appeared to meet of all these criteria (Davis et al. 2002). The intervention is primarily cognitive-behavioural. It assumes that parents and children make sense of their worlds by developing constructs, and that any change requires exploration and modification of these constructs. An example of such a construct is where a mother believes that her baby is behaving in a particular way in order to annoy her, which may result in the baby being punished.

The intervention also focuses on relationship building. The home visitors were all trained in the use of effective communication skills and methods of working in partnership with families such that the family's agenda was given priority over the agenda of the professional. The home visitors were encouraged to work in a way that demonstrated factors that are important for relationship building, such as warmth, respect, genuinessness and empathy. They were also provided with training in four methods of improving mother–infant interaction:

- Infant massage (Onzawa et al. 2001)
- Baby dance (Maattanen 1999)
- Songs and music (Street 2002; PEEP 2000)
- A modified version of the Brazelton method (PIPPIN 2000).

The eight-day training programme was delivered using a training manual over eight consecutive weeks. In addition, home visitors received twice-monthly group-based supervision provided by two skilled psychotherapists trained in psychodynamics and with experience in infant mental health.

Evaluation

We used an RCT design to evaluate this new service. Our sampling frame consisted of pregnant mothers at 40 GP practices in Oxfordshire and Buckinghamshire, during 2001. Vulnerable mothers were identified and recruited using a two-stage recruitment process, described below. Consenting women were then randomly allocated either to the intervention group, who received weekly visits from a home visitor, or to the control group, who received standard services.

Recruitment involved a two-stage process. In stage one, midwives attached to the 40 participating GP practices screened all pregnant women at the booking-in visit to identify women meeting our entry criteria. Those women who agreed were referred to the trial researcher.

At stage two, women who were contactable were visited by a researcher to ensure that they were eligible to participate. Inclusion criteria were as follows:

- Homelessness/moved more than twice in last 12 months
- Severe debt/financial hardship
- Absence of support networks
- Aged 17 or under
- Mental health problems
- Substance misuse/addiction
- Major current parenting difficulties
- Child on Child Protection (CP) register
- Domestic violence
- Social worker involvement relating to children
- Non-specific serious concern of midwife.

Factors such as frequent moves in the last 12 months, being under age 17, mental health problems, substance misuse and addiction, a child on the CP register and domestic violence are all strong predictors of abusive and neglectful parenting. We excluded only women who were not able to understand spoken English or who did not meet at least one of the above criteria.

All eligible women were then given oral and written information about the study, and allowed two weeks in which to make a decision about participating. At the next visit, consenting women were randomised using computer-generated allocations with sequentially numbered, sealed opaque envelopes for consecutive eligible mothers.

Table 1 demonstrates the high-risk characteristics of the sample that were recruited. The average participant had five risk factors; an example case was a woman experiencing domestic violence, an unwanted pregnancy, no support, poverty, and repeated moves in the last 12 months.

Table 1. Sample characteristics

	Home Visiting		Control Group	
	%	number	%	number
8 or more risk factors Example case: drug abuse; domestic violence; alcohol problems; recent event; housing concerns; mental health problems; partner with mental health problems; social worker; child with behavioural problems; unhappy childhood; no support; poverty; criminal record	22.4%	15	14.1%	9
5 – 7 risk factors Example case: domestic violence; recent unwanted pregnancy; housing concerns; no support; poverty; moved more than twice in last two years	32.8%	22	32.8%	214
1 – 4 risk factors Example case: alcohol problems; mental health problems	4.8%	34	53.1%	30

Main outcome measures

The main outcomes, measured at a 12-month follow-up for both mother and infant, were:

Maternal

- Mental health (GHQ)
- Parenting competence (PSOC)
- Parenting stress (PSI)
- Parenting attitudes (AAPI).

Infant

- Home environment (HOME Inventory)
- Mental development (Bayley scales)
- Social and emotional development (BITSEA).

Mother–Infant

- Maternal sensitivity and infant co-operativeness (CARE-Index).

The Home Inventory, Bayley Scales and Care Index were all measured by independent assessors who were blind to the intervention groups. The remaining measures are based on the parent's report.

We used these proxy measures of abusive parenting as primary outcomes because there is increasing evidence to suggest that it is not possible to evaluate the effectiveness of home visiting in reducing the incidence of abuse. This is due to the fact that the extra surveillance guarantees that cases are identified in the home visiting arm, whereas cases in the control arm are not picked up until much later.

We did, however, collect some objective measures of the effectiveness of home visitors in identifying abuse as part of our economic analysis, in addition to the mother and infant's use of hospital and community-based services. The cost of the resources used was combined with our primary outcomes within a cost effectiveness analysis. A cost-benefit analysis using a discrete choice experiment was also undertaken.

Demonstrating effectiveness

The findings of this study raise some issues that are fundamental to the way in which we go about evaluating interventions to support families, and fundamental to the issue of evaluation itself. As I indicated at the beginning of this paper, the four recent studies that have evaluated the effectiveness of a home visiting programme in the UK have all shown problems demonstrating effectiveness using the primary outcome measures. I chose to use the term 'problems demonstrating effectiveness' because a lack of evidence of effectiveness does not mean that something is ineffective, and also because it is hard to believe that programmes such as Home- Start (i.e. one of the four home visiting programmes that was evaluated), aren't making a difference to families. This issue is very clearly demonstrated if we compare the outcomes from the two types of data (i.e. quantitative and qualitative) that were collected as part of our evaluation of the effectiveness of a home visiting programme.

Very briefly[1], the quantitative results at 12-months post-natal showed no significant differences between the two arms for any of the parent-report measures, and for only one of three independent measures, the CARE Index, which showed improved sensitivity and infant co-operativeness for the intervention group. The findings from the qualitative data, however, tell a somewhat different story, and the remainder of this section provides further detail about what this data showed.

The qualitative data

The qualitative data was obtained from in-depth interviews with participating women and in-depth interview and focus groups with the participating home visitors. The aim was not only to examine process issues, such as whether both service providers and recipients complied, and whether they were satisfied with the intervention, but also to capture user and provider perceptions concerning the impact of the intervention on participants' lives.

1 There is unfortunately not scope within the current paper to present some of the interesting data on the identification of child abuse, or the economic data on service use and cost which are available in the full report about the study (Barlow et al., 2005).

Home visitors' perceptions of the benefits of home visiting

The data from service providers showed quite clearly that all of the home visitors perceived there to have been important changes in their practice. They talked about using a preventive model, as opposed to the 'crisis management' that they felt characterised their standard practice. They described having more in-depth knowledge as opposed to 'skimming the surface'; and this permitted better monitoring of the lives of high-risk babies. They talked about working holistically; being more client-centred and less directive; and focusing more on developing relationships and on helping the mother to get to know her baby.

Perhaps most importantly, they described a range of benefits resulting from their intervention, some of which are quite important:

- Building confidence;
- Improving parents' ability to understand and appreciate their baby and its needs;
- Preventing family breakdown and improving mother-child relationships during child protection investigations;
- Enabling a mother to face her fears and deal successfully with psychological problems;
- Providing ongoing support to families while they receive temporary support from other agencies, thus providing continuity;
- Providing significant mental health support, which would normally require referral to a specialist service;
- Mediating on behalf of families who might normally be rejected by specialist services;
- Providing, for some women, the first stable relationship they may ever have experienced.

As one home visitor told us:

> *She managed to chuck her … partner out. He was violent towards her and he was getting the older sibling to beat up the study infant … the fact that there was a constant source of support there and I was listening to her side of the story gave her the confidence to do that.*

Another home visitor said:

> *… her mental health was very low. I guess about 10 days after the baby was born, and at one point she did say 'I can't care for this baby, I want this baby fostered'. She didn't really think that she could do it, and the relationship that she has with her son is just wonderful to see now, which is a sense of satisfaction. She is a really good mum, she is really responsive to him.*

The evidence from the home visitors resonated very strongly with the evidence provided by the participating women. The in-depth interviews with 20 of the participating women showed that, despite considerable hostility towards professionals at the outset, the majority of the women interviewed developed strong trusting relationships with their home visitor. Many of the women

said that it had taken them a long time to trust their home visitor, but they had then really opened up, sometimes for the first time in their life.

There was also considerable evidence about the effects of the partnership model of working and, in particular, the value that these women placed on being treated with respect and helped to feel in control.

Like the home visitors, the participating women described multiple benefits. Similar themes emerged from the two sets of data: more confidence, better mental health, less domestic violence and so forth. One woman told us:

> *I have learned quite a few things with [the Home Visitor] like doing baby dance … that I had never heard of before and she used to love it. We do the movements and we sit and you know she encourages you to turn the telly off and read to her, which I do, and some other things, play with her and interact with her, that I would never have thought of doing … she really likes books so I do make the effort to read to her most days … but again, that is not something I would have thought to do.*

Another woman said:

> *This study has made me feel so much more confident in myself and made me more assertive. I'm now able to cope with day-to-day activities and dramas – it has made me strong enough to see that I am worthy of having a happy life. I am really enjoying being a mother and if you were to see my baby you would see what a happy, loving little baby he is. I would like to take this opportunity to thank my health visitor for all her care and support over the last year.*

Reasons for discrepancies

There are several possible reasons for the discrepancies between the quantitative and qualitative data. For example, it may simply be that we measured different things with the two sets of data. Some of the things that the women and health visitors talked about, such as domestic violence, were not measured using standardised instruments. However, this was not the case for all outcomes. So while many women who were interviewed reported improved confidence and better mental health, these improvements were not apparent using our standardised measures.

Was it the case, then, that our interviewees were in some way unrepresentative, and the next 20 women would have been wholly negative about the impact of the intervention on their lives? While this seems unlikely, it may be the case that we would we have obtained more interesting data if we had used the findings from the quantitative study to decide who to interview. In other words, instead of interviewing the first 20 cases, we might have selected those families that showed changes on our outcome measures and those that showed no change.

Perhaps there were problems with the standardised measures such that this particular population were unable to be honest because of their preoccupation with child protection issues and fears that they might lose their baby. Or perhaps they lacked insight about their problems, or failed to understand the questions.

Many of these issues could be untangled with further time and resources. For example, we could use cognitive debriefing methods to find out what our interviewees understand and think about when they answer questions from standardised questionnaires. We could also do case studies in which we examine more closely the outcomes on the standardised measures for the women who were interviewed and see if they are consistent.

Perhaps most fundamentally, we need to know which of our two stories is more accurate, whether our intervention is effective or not. I think the answer is that both types of data are accurate, and that the answer to whether these interventions are effective is 'yes' and 'no'. They are effective for some of the people some of the time. The fact that they are not effective for everyone all of the time does not mean that they are ineffective.

The reason for this mixed result is that parenting and family support interventions are social and psychological interventions. They are delivered by people with a range of skills, to families with a range of backgrounds, with different levels of readiness for change and/or commitment to the process, and so forth – all of which determine whether an intervention is effective.

Conclusion

Where does this leave those of us whose lives are devoted to evaluating these interventions? I think it is time for a change of focus. We need to begin to focus more on context, mechanism and outcomes, or 'realistic evaluation' (Pawson and Tilley 1997). Fundamental to this approach is the recognition that, unlike many medical interventions, complex parenting and family support interventions do not work in and of themselves. Rather, it is the actions, motivations and desires of the participants and stakeholders that make them work.

Put briefly, this may mean that we invest more time in searching for answers concerning for whom an intervention works, and under what circumstances, rather than focusing on 'effectiveness' across groups in which we have tried to control for differences using randomisation, for example.

Discussion

The discussion focused initially on whether this study was under-powered. It was suggested that the effect of complex community-based interventions of this nature is likely to be small, and that a much larger study might be needed to detect the small changes that may have occurred.

There was also some discussion concerning the benefit of examining the data more closely to identify the characteristics of those families for whom the intervention may have been effective.

References

Acheson D (1998). *Independent Inquiry into Inequalities in Health Report*. London: The Stationery Office.

Barlow J, Davis H, Stewart-Brown S (2005). *The Oxfordshire Home Visiting Study: Final Report.* University of Warwick.

Bull J, McCormick G, Swann C, Mulvihill C (2004). *Ante- and post-natal home-visiting programmes: a review of reviews.* HDA.

Davis H, Day C, Bidmead C (2002). *Working in partnership with parents: the parent advisor approach.* The Psychological Corporation.

Maattanen K (1999). Dialoginen auvatanssi psyykkisen kuntouttamisen menetelmana lastenkodeissa (Dialogical baby dance as a method of therapy for foster homes). *Suomen Laakarilehti (Finnish Journal of Medical Doctors)* 99, 4022–4027.

McAuley C, Knapp M, Beecham J, McCurry N, Sleed M (2004). *Evaluating the outcomes and costs of Home-Start support to young families experiencing stress: a comparative cross nation study.* Joseph Rowntree Foundation.

Moran P, Ghate D (2004). *What works in parenting support: the international evidence.* Policy Research Bureau.

Morrell CJ, Siby H, Sewart P, Walters S, Morgan A (2000). Costs and effectiveness of community postnatal support workers: randomised controlled trial. *British Medical Journal* 321, 593–8.

Onzawa K, Glover V, Adams D, Modi N, Kumar RC (2001). Infant massage improves mother-infant interaction for mothers with postnatal depression. *Journal of Affective Disorders* 63, 201–207.

Pawson R, Tilley N (1997). *Realistic evaluation*. Sage Publications.

PEEP (2000). *Learning together with babies.* PEEP.

PIPPIN (2000). *Your baby as a person,* Training Notes. PIPPIN.

Street A (2002). *The use of music in PEEP*. PEEP.

Wiggins M, Oakley A, Roberts I, Turner H, Rajan L, Austerberry H, Mujica R, Mugford M (2004, in press). *Postnatal support for mothers living in disadvantaged areas: a randomised controlled trial and economic evaluation*. Health Technology Assessment.

Evaluating the impact of Playing for Success

Caroline Sharp, National Foundation for Educational Research

Playing for Success (PfS) was set up in 1997 to establish study support centres at professional sports clubs. Its main aims are to contribute to raising educational standards, especially in numeracy, literacy and ICT, and to improve pupils' motivation to learn. It is targeted at young people in Key Stages 2 and 3, particularly those who are identified as struggling a little with skills in these areas and who are often demotivated at school. Pupils attend PfS centres after school for around 20 hours over ten weeks. The initiative began by establishing centres in professional football clubs, but went on to encompass sports other than football (including rugby league, rugby union, cricket and basketball). The number of centres has grown from six in 1997/98 to over 100 clubs signed up and 83 centres open.

The NFER was responsible for the national evaluation of PfS for four consecutive years, from 1998 to 2002 (Sharp et al. 1999, 2001, 2002a, 2003a). The NFER has also just carried out a study looking at the longer-term impact on pupils' attainment (Sharp et al. forthcoming).

Aims and design

The two main aims of the four national evaluations were to provide an indication of the effectiveness of the initiative, and to identify and describe those features leading to success in terms of participation, gains in motivation, positive attitudes towards learning and enhanced learning outcomes.

The evaluation was always conceived as a national evaluation. The team focused on the achievement of the initiative's core purpose, rather than the delivery of the programme in individual centres, in order to measure the effectiveness of PfS as a national initiative. The Department for Education and Skills wanted both a process evaluation (focused on implementation) and an outcome evaluation (focused on impact). The DfES also developed a strategy for local evaluation and monitoring.

The evaluation gathered data from the key interest groups – pupils, parents, teachers, and centre staff – using questionnaires, tests, interviews and observation. The main data collection took place at the beginning and end of pupils' attendance at the centres. In the second and third years, the evaluation adopted an experimental design with a control group. As control group designs are of particular interest in relation to evidenced-informed practice, the following section provides more information on this part of the evaluation design.

Experimental design

The first year's evaluation indicated that young people attending PfS centres had enjoyed the experience and had made progress in tests of reading and mental arithmetic. But the obvious question was whether their progress was greater than would have been expected had they not attended. To some extent, the evaluation team could answer this by referring to the norms established in the test instruments. To provide a better estimate of the impact of the initiative, we suggested attempting a control group design.

There were a number of challenges in establishing a control group. First, the evaluation team had to find an approach that would keep any extra demands on centre managers and schools to a minimum. Second, adopting a control-group design held both ethical and practical challenges. Ethical challenges concerned the possible exclusion of a group of pupils from a potentially beneficial initiative. Not only would they be excluded from attending, but they would be required to complete tests that would not be of any immediate benefit to themselves or their teachers. Practical challenges concerned organising the control group at relatively short notice (a matter of months in advance) and finding a method that would provide robust data within the limited time and resources available.

In response to these challenges, the evaluation team visited the centre managers to discuss the issues. Some centres had a group of schools with which they were developing a longer-term partnership. These schools sent a small cohort of pupils in the target group to the centre each term. We therefore proposed that the schools should select twice their usual number of pupils (all of whom would meet the normal criteria for participation) and then use a randomised method of selection. For ease of application, we suggested that teachers should list pupils' names in alphabetical order and select every second pupil to form two groups. The first group (the treatment group) would attend the centre as usual. The centres were asked to make places available to the remaining pupils (the control group) during the following school term. This had the advantage of addressing the teachers' concerns about a group of pupils missing out on the experience. A disadvantage was that the control group would subsequently participate, so that the possibility of studying the longer-term effects of the initiative on these cohorts would be lost.

Five of the 12 centres participating in the national evaluation at the time agreed to participate in the control group study. Four of them used the method of random selection outlined above. The fifth centre took pupils from different schools each term. They offered to provide a control group comprising pupils from a matched school. The control group sample was relatively small, comprising 105 pupils in the second year. A similar method was used to establish a control group in the third year of the evaluation, comprising 130 pupils attending four centres. It was decided not to attempt to recruit a control group for the fourth year, because it was considered that the two previous years had provided sufficient evidence about the impact of the initiative and that the focus of the final year's evaluation was on whether standards had been maintained as the programme expanded. It was therefore felt that the effort put into recruiting control group participants and the burden on pupils, centres and schools could not be justified for another year.

Evaluation instruments

The first year's evaluation focused on the first six centres, and considered both the challenges of setting up the centres in professional football clubs and the outcomes achieved by pupils attending them. Outcomes were measured by a combination of commercially available tests (in mental arithmetic and reading) and instruments developed by the NFER (attitude scales and pupil, parent and school questionnaires).

Feedback from the centre managers participating in the first year's evaluation indicated that the commercial tests used (QCA mental arithmetic tests for KS2 and 3 and NFER-Nelson Group Reading Test, second edition, Forms X and Y) were too long and challenging for the pupils attending. The NFER team therefore developed new tests to assess pupils' numeracy and reading comprehension. The tests were designed to be appropriate for the age group (in keeping

with the National Curriculum), and contained items at different levels of difficulty. These tests were trialled and age standardised using a sample of almost 2,000 Year 6 and Year 9 pupils from the national sample of schools with an achievement profile similar to those sending pupils to PfS. Each test had two different versions of equivalent difficulty (for pre- and post-course administration) to avoid pupils performing better on the second occasion due to familiarity with the tests. Tests were returned to the NFER for marking. The NFER team also developed a checklist to measure pupils' ICT skills.

The instruments developed for the second year were also used in the two subsequent evaluation studies and are as follows:

- Pupil data form (attendance data linked with pupil characteristics, such as gender, date of birth and eligibility for free school meals);
- Centre manager questionnaire (centre location, capacity, facilities, staffing, school liaison and learning programme);
- School questionnaire (impact on pupils, organisation and liaison, suggested improvements);
- Pupil questionnaire (views on their participation and a set of attitudinal measures);
- Parent questionnaire (views on their child's participation in the initiative);
- Pupil self-assessment of ICT skills (computer basics, word processing, e-mail and internet);
- Tests of numeracy and reading comprehension.

The pupil and parent instruments were administered twice: at the beginning and end of pupils' participation. To keep the evaluation burden to a minimum, pupils completed two of the four instruments (either the questionnaire and the numeracy test, or the ICT self-assessment and the test of reading comprehension).

The quantitative approach was complemented by qualitative work, including visits to centres and schools to observe sessions and to interview pupils, teachers and centre staff. One question addressed in the second year of the evaluation was whether the initiative continued to have an impact on pupils after they had completed the course. The research team visited schools a term after their pupils had attended Playing for Success. Visits were made to nine schools to interview 39 pupils and their teachers.

Analysis of quantitative data

The data from the questionnaires and tests were subjected to statistical analysis to establish whether pupils had made progress, and whether the progress was likely to have been influenced by their participation in PfS. Results from the tests were age standardised (i.e. adjusted to reflect the score expected to be achieved by a pupil of a given age). Pupils' answers to individual items in the attitudinal questionnaire were combined into scales, using factor analysis.

The study used multilevel modelling to consider the evidence for the effects of PfS on young

people's progress, while controlling for other factors known to influence attainment. Multilevel modelling is a development of regression analysis that takes account of data grouped into similar clusters at different levels (see Goldstein 1987). For example, individual pupils are grouped into year groups or cohorts, and those cohorts are grouped within centres. There may be more in common between pupils within the same cohort than with other cohorts. Multilevel modelling allows statisticians to take account of this hierarchical structure of the data, and to produce more accurate predictions, as well as estimates of the differences between pupils, cohorts and centres.

For each outcome measure, all pupils with both pre- and post-course scores on that measure were included. Initial analyses indicated that there were different patterns of results for pupils in primary and secondary schools, so it was decided to treat these as separate groups. (It should be noted that centres ran separate sessions for primary and secondary pupils.) Finally, the results of pupils in the control group were included alongside pupils who had attended PfS. The following variables were included at the pupil level:

- Key stage
- Sex
- Entitlement for free school meals
- Stage of the code of practice for special educational needs
- Ethnic background
- Fluency in English
- Attendance at PfS centres/control group.

At the centre level, the model included:

- Centre attended
- Course length (number of hours).

In the second and third years of the evaluation study, comparisons were made with the control group data collected the same year. In the fourth year, comparisons were made between the progress achieved by pupils attending PfS centres and the control group's results in the previous year.

Evaluation results

The national evaluations focused on pupils attending PfS centres during the spring term in each year for four consecutive years. During that time, the evaluation included 57 centres and more than 5,150 pupils, 1,640 parents and 330 schools. Some centres and schools participated more than once.

Despite the programme's rapid expansion, the evaluation findings were remarkably consistent from year to year. There was also considerable consistency of impact for pupils from different backgrounds. On the whole, results were similar for pupils with different characteristics, including sex, ethnicity and eligibility for free school meals. There were some indications, however, that pupils with a statement of special needs did not make as much progress as others attending the centres. Further information on

this is provided in the main evaluation reports and in Sharp et al. (2003b).

The evaluation considered the degree of satisfaction with the programme as well as its impact on pupils.

Satisfaction with the initiative

It was clear from the evaluation results that PfS was very popular with all its client groups. For example, the centres achieved high levels of attendance (the majority of pupils attended for at least 80% of the available time), despite the fact that pupils were making a voluntary commitment to attend in their free time. Parents said they were pleased that their children had been selected to participate. Pupils, parents and schools all considered that the initiative had benefited pupils' learning and self-confidence. The overwhelming majority of schools participating in the evaluation indicated that they were keen to participate again.

Pupils' responses showed that they enjoyed the experience of attending the centres. For example, less than 5% of the sample rated the centre as 'boring' and an overwhelming majority (85% or more in each year of the evaluation) felt that the centre had been 'fun', 'interesting' and 'a good idea for me'. This was reinforced by answers to the open questions in the pupil questionnaires. The following were typical:

> *The centre helped me to do maths better and quicker and realise I am not stupid but a bit slow in working out the answers.*
>
> *I am now a very confident person. I can put up my hand and have a go without getting embarrassed.*

There was a further indication of the strength of pupils' allegiance to the centres. In the first year's evaluation, the pupils were asked to note anything they had not liked about the centre. Over a quarter of pupils chose not to answer, and over a third said that there was nothing they had disliked. In subsequent years, we rephrased the question, asking whether there was anything pupils felt could be improved. This provoked a protective response from some pupils, who made comments along the lines of '*our centre is good already, so please leave it alone*'. The only consistent suggestion for 'improvement' (from both pupils and parents) was that PfS should be expanded so that more pupils could attend for longer periods of time.

Changes in pupils' attitudes

The strong message from pupils, parents and schools was that PfS had impacted positively on pupils' attitudes, especially their self-confidence and their motivation to attempt more challenging work. However, the attitude scales failed to provide strong evidence of this effect. On the whole, the attitudes of pupils attending PfS (as measured by the 12 attitude scales) were more positive at the end than at the beginning of the course. In many cases, the gains were statistically significant when compared with the control group. However, to gain a better impression of the extent of the impact, we used effect size calculations (see Slavin and Fashola 1998; Schagen and Elliott 2004). The only results of 'educational significance' (i.e. those with an effect size of 0.25 or more) were in relation to independent study skills (for KS2 pupils in the year 4 evaluation and KS3 pupils in the year 3 and 4 evaluations), and self-image (KS2, year 4 evaluation).

One issue that should be mentioned is the relatively low reliability of attitude scales over time (when compared to other measures, such as tests of knowledge). Attitudes may be influenced by many factors, including how the individual is feeling on a particular day. Observed changes in pupils'

confidence and behaviour (e.g. greater confidence in class) were, however, consistently mentioned by all key stakeholders, including centre staff, teachers, parents and the pupils themselves.

Results in ICT, literacy and numeracy

With regard to the assessment outcomes, the evaluation studies showed statistically significant improvement in pupils' literacy, numeracy and ICT. Each year, and in each of the three skill areas, pupils' results were significantly higher, on average, at the end of their time at the centre.

ICT skills, measured by a self-report checklist, were obtained for three years (the ICT skills checklist was developed in the second year). The results showed statistically significant gains in pupils' ability to operate a computer, carry out word processing functions, send/receive e-mail and use the internet. Control group comparisons using effect size showed that the difference was of educational significance for both key stages for each of the three years in which this instrument was used, with an average effect size of around 0.8.

Use of age-standardised numeracy and literacy tests enabled the evaluation to adjust the scores to take account of improvement in post-course scores due to maturation (i.e. to compensate for the fact that pupils could be achieving higher scores on the second test simply because they were slightly older). It also enabled comparisons with a national sample of pupils.

Comparisons with a national distribution of scores showed that, on average, the pupils selected to attend PfS were achieving at a very low level compared with national norms in literacy and numeracy. By the end of the course, pupils' scores had risen closer to national norms and, in the case of numeracy scores, pupils in KS2 who attended PfS were achieving at just below the expected level for their age.

Comparisons with the control group showed that gains were greatest for numeracy. In numeracy, all groups of pupils who attended PfS out-performed the control group to a statistically significant extent. Pupils also made progress in reading comprehension during their time at the centres, but to a lesser extent.

As noted above, the analysis considered the progress of pupils in each key stage separately. In each of the three years when control-group comparisons were made, one of the two age groups out-performed the control group in reading comprehension, but it was not always the same one. In the 1999–2000 evaluation, pupils in KS3 made greater progress than the control group in reading comprehension. In 2000–01, pupils in KS2 out-performed the control group to a statistically significant extent, as did pupils in KS3 in 2001–02.

By comparing their progress on the two tests to national norms, it is possible to estimate pupils' progress in months (over and above the effects of maturation between the first and second test). This is shown in Table 1. Results for the first year's evaluation are excluded because it used different tests and did not include control group comparisons.

Table 1. Progress in months achieved by pupils attending Playing for Succes

	Reading comprehension		*Numeracy*	
Year of evaluation	**KS2**	**KS3**	**KS2**	**KS3**
2000	ns	6	21	8
2001	15	ns	18	14
2002	ns	8	17	24

Based on test results for 2,800 pupils

The table shows the progress in months for the groups of pupils who out-performed the control group to a statistically significant extent. Pupils attending PfS made up to 15 months' progress in reading comprehension, and 8–24 months' progress in numeracy.

The gains in numeracy – greater than gains in literacy – are worthy of further comment. It may be that pupils spent more time at the centres working on numeracy than on literacy skills. We know that centres tended to focus on mental arithmetic, numeracy, writing (composition) and reading skills. While improving mental arithmetic skills is likely to help pupils complete a numeracy test, there is not the same equivalence between improvements in writing and reading comprehension skills. Therefore the match between the content of the tests and the learning experiences offered by the centres is closer for numeracy than for reading comprehension.

It also seems likely that it is easier to help pupils to make progress in numeracy skills, especially within a relatively short period of time. This view is supported by previous research. For example, Professor MacBeath's (2001) school improvement study found that the amount of variance attributed to the influence of the school was much lower for reading (12%) than for mathematics (33%). To succeed in a test of reading comprehension, pupils need to call on a complex range of skills and conceptual understanding, including the ability to read text, understand and interpret meaning, put together ideas from different parts of the text, draw inferences, and use their general knowledge to make logical deductions. In comparison, numeracy is a relatively discrete skill, and it may therefore be easier for pupils to make rapid progress in numeracy once they have grasped the concepts involved.

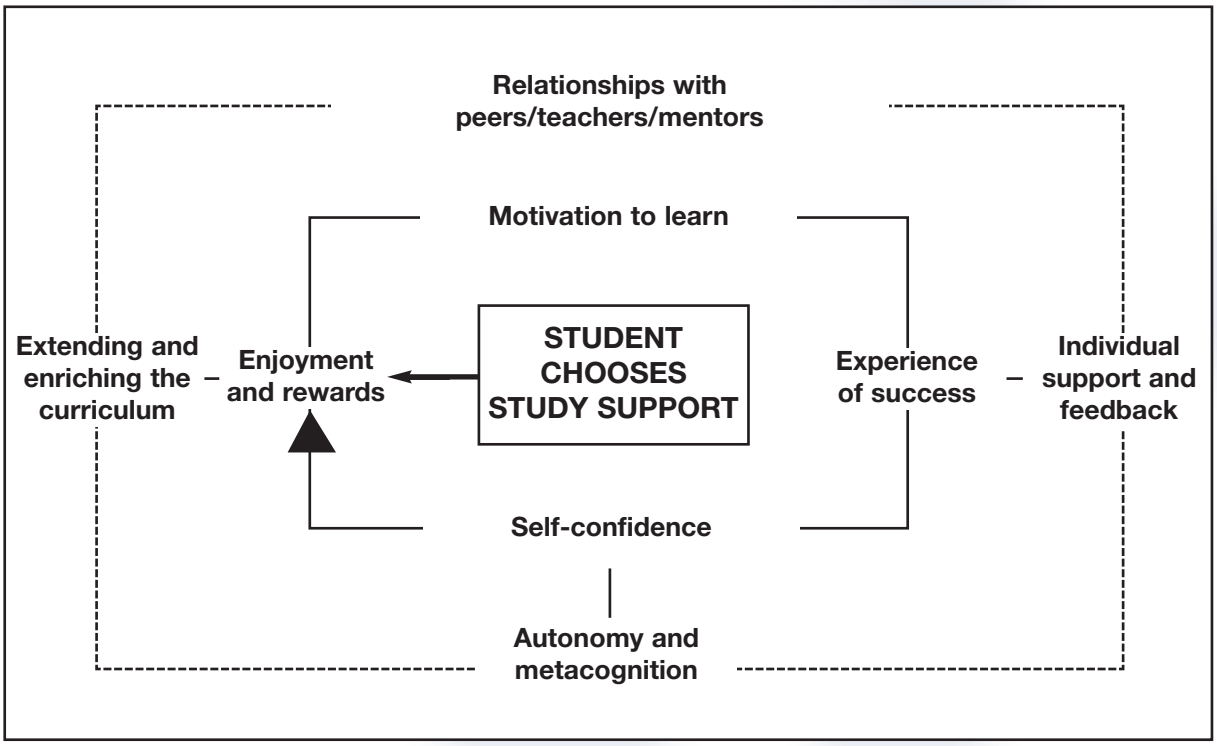
How does Playing for Success help under-achieving young people?

A question that has constantly been asked about the evaluation results is: 'Given that pupils only visited the centres for up to 20 hours, how has so much progress been possible in such a short time?' The results achieved by the control group are helpful in indicating what happens to under-achieving young people who do not benefit from a specific intervention to help them make progress. The scores recorded by control group pupils tended to remain unchanged during the evaluation period (around a term), whereas the expectation for most pupils would be for scores to increase as they matured. As time goes on, therefore, these under-achieving pupils are in danger of slipping further and further behind the achievement expected for pupils of their age. A further indication of this is the fact that the pre-course standardised (i.e. age-related) scores of pupils in KS3 tended to be much lower than those of pupils in KS2.

In the second year's evaluation, the research team carried out a qualitative study of pupils who had attended one of three PfS centres about a term before. This provided insights into the experience of lower-achieving pupils. The young people interviewed related that they had often got stuck on a particular concept at school. Failure to grasp basic concepts or acquire skills held them back, and made them stop trying. They lost confidence in their ability to learn. As a consequence, they tended to avoid work wherever possible, gave up easily and 'kept their heads down' in class rather than attempt to answer a question and risk getting it wrong.

To explain the relationships evident in the evaluation data, we suggested a model of how PfS encourages pupils to become more successful, independent learners (Figure 1). This was based on previous work considering the process of learning in study support in relation to the theory of 'self-regulated learning' (see Sharp et al. 2002b).

Figure 1. How Playing for Success influences learning



The model begins with a pupil's decision to attend Playing for Success. This represents a positive choice, for although some pupils were nominated by their teachers, the pupils decided whether or not to take up the opportunity. Choice in this matter is important, because it represents a positive decision to re-engage with learning.

Once at the centre, pupils experienced the immediate reward of visiting a prestigious venue with good computer facilities. The centres encouraged an atmosphere of learning in a fun and supportive environment, which added to pupils' initial enjoyment of the experience and made them want to return.

From their first session at the centres, young people were given help designed to meet their particular learning needs. Most centres used a target-setting process, whereby pupils were encouraged to identify the aspects of literacy and numeracy they found most difficult. This sense of working on the learner's agenda and setting realistic targets reinforced pupils' motivation to learn. Pupils also appreciated the support they received from centre staff and mentors, finding them helpful, patient and encouraging. Pupils described how staff motivated them to keep trying, rather than giving up if they found something difficult at first. As a Year 6 pupil said: '*All the staff were very kind. I thought they were dead helpful*.' This view was reinforced in the comments of parents, one of whom wrote:

The best thing was the ever-supplied help and encouragement from all the staff and helpers who were never too busy to listen, and that is very important to a child.

Pupils were further motivated by the fact that they were beginning to experience success in their learning. This was underpinned by the skilful matching of tasks to pupils, so that they were

constantly making progress and receiving feedback on their improvement. Centre staff ensured that pupils understood the concept in question, and pupils got the chance to practise and master the necessary skills. Pupils said they enjoyed the learning tasks and relished the opportunity to use up-to-date computer facilities. (This was mentioned frequently in pupils' responses to the questionnaire.) Staff also encouraged pupils to become more active, independent (autonomous) learners. As a result, they learned new skills and began to regain their confidence in their ability to succeed. They also learned 'metacognitive' strategies (mental approaches to help manage and integrate concepts), so that they were more able to monitor and fine-tune their own learning.

Greater self-confidence leads to greater enjoyment in learning. Pupils gained both extrinsic rewards (praise from staff, certificates and small prizes) and intrinsic rewards (such as the pleasure in succeeding at a difficult task). Once pupils' potential was unlocked in this way, they became more effective learners, making rapid progress in basic skills.

One of the pupils we spoke to explained how the centre had helped with a specific aspect of learning:

We did this question in maths at school. We did all this homework on it and everything, but I just didn't understand. Then I went to the centre and asked the centre manager. She explained how to do it in our sense of understanding and saying it. Then I understood, and now I know how to do it and I don't get low marks on it in my mental tests.

Evaluation challenges

Conducting a national evaluation of a national initiative holds many challenges, especially if the aim is to provide evidence on an initiative in its early stages of development. Playing for Success was a new concept, requiring the establishment of new partnerships, content and ways of working - all of which took time to evolve. Not surprisingly, centre managers were concerned that they might be judged prematurely, and worried that the evaluation might threaten the initiative's chance of success.

In the early days, there was concern among some centre managers that, while they could achieve the objective of motivating young people to re-engage with learning, they were less confident that this would result in demonstrable gains in attainment within the short time they had to work with the pupils. There was also the problem of using tests as a method of data collection. As centre managers pointed out, the pupils attending PfS were underachieving, and the last thing they wanted to do on their first visit to the centre was to sit a test. However, the DfES wanted robust evidence of impact on attainment, and testing appeared to be the best option to provide the evidence required. This continued to be a concern for some centre managers, but most found ways of incorporating the testing into their programme, so that it was not unnecessarily burdensome or demotivating to pupils. And, as noted above, their concerns about the focus on academic outcomes proved to be unfounded, as pupils made significant skill gains during their time at the centres.

A programme evaluation, by its nature, focuses on the core aspects of a given programme. In the case of Playing for Success, core requirements were clearly defined, but centres had the flexibility

to tailor provision to respond to local needs while meeting those requirements. This meant that there was sometimes a perceived difference in emphasis between what individual centres were providing and the aspects focused on in the national evaluation study.

The key lesson to emerge from this experience is that evaluators (and sponsors) need to consider very carefully the balance between national evaluation and local programme development. Building respect, trust and a good working relationship was a process that took time and effort on the part of the evaluators, centre managers, sponsors, the evaluation steering group, and the cadre of 'critical friends' who helped the programme to develop. In particular, meetings between the evaluators and centre managers were important, as were the individual feedback reports for each centre. (The results of these were not disclosed to the DfES by the evaluation team.)

Conclusion

PfS is an innovative programme that has used the medium and environment of sport to motivate underachieving young people. The national evaluation has demonstrated that pupils who attended made significant progress in a short time, particularly in ICT skills and in numeracy. The use of a control group model was important, because it demonstrated the extent of the impact on the pupils who attended. It also provided an insight into the consequences of not intervening to help underachievers to make progress. The combination of quantitative and qualitative methods enabled the evaluation to provide robust evidence of impact, together with an insight into how that impact was achieved. We hope that, by helping to tell the story of Playing for Success, we have contributed to the development of the programme and have highlighted some effective strategies that can be used by others seeking to support underachieving young people.

References

Goldstein H (1987). *Multilevel models in educational and social research*. Charles Griffin.

MacBeath J (2001). *Improving school effectiveness*, TOPIC, 25(1).

Schagen I, Elliot K (Eds) (2004). *But what does it mean? The use of effect sizes in educational research*. NFER.

Sharp C, Blackmore J, Kendall L, Greene K, Keys W, Macauley A, Schagen I, Yeshanew T (2003a). *Playing for Success: an evaluation of the fourth year (*DfES Research Report 402*)*. DfES.

Sharp C, Blackmore J, Kendall L, Schagen I, Mason K, O'Connor K (2002a). *Playing for Success: an evaluation of the third year* (DfES Research Report 337). DfES.

Sharp C, Kendall L, Bhabra S, Schagen I, Duff J (2001). *Playing for Success: an evaluation of the second year* (DfES Research Report 291). DfES.

Sharp C, Kendall L, Schagen I (2003b). *Different for girls? An exploration of the impact of Playing for Success, Educational Research*, 45(3) 309-24.

Sharp C, Mawson C, Pocklington K, Kendall L, Morrison J (1999). *Playing for Success: an evaluation of the first year* (DfEE Research Report 167). DfEE.

Sharp C, Pocklington K, Weindling D (2002b). *Study support and the development of the self-regulated learner*, Educational Research, 44(1) 29-41.

Sharp C, Schagen I, Scott E (forthcoming). *Playing for Success: the longer-term impact.* A multilevel analysis. DfES.

Slavin RE, Fashola OS (1998). *Show me the evidence! Proven and promising programs for America's schools*. Corwin Press.

Factors influencing the transfer of good practice

Michael Eraut, University of Sussex, on behalf of the research team from Sussex and Demos

The report of this research project, carried out by a team from the University of Sussex and Demos (see Fielding et al. 2005 for a list of team members) was published in January 2005. This paper gives a brief review of its findings, their practical implications, and the theories of practice development and professional learning that emerged from them. The last five years have seen several government initiatives to encourage and support the transfer of 'good practices' developed by schools and individual teachers. These policies have evolved over time; our research focused on those policies being implemented in Summer 2003.

The purpose of the research was to study and report on the factors facilitating or constraining the transfer of good practice between schools, at school or individual level. It concentrated on classroom practices and the organisational or management practices that supported them. Until then, policies had focused mainly on supporting the 'suppliers', those who had developed the practices. Our project, however, aimed to understand more fully:

- The challenges of the transfer of good practice from the standpoint of the receiver
- The nature of practice and its reception
- The challenges for the originating institution.

Methodology

Sources of data

The project entailed in-depth qualitative research with more than 120 individual practitioners in schools and clusters of schools, head teachers and senior staff, and external 'brokers'. The main sources of our evidence are summarised as follows:

- 10 Beacon and leading edge schools + 17 of their partner schools
- 2 specialist schools + 4 of their partner schools
- 9 schools that had discontinued transfer work
- 3 clusters of schools within Excellence in Cities or Education Action Zones
- Heads from schools in challenging circumstances
- 2 virtual Education Action Zones
- 13 Advanced Skills Teachers + 10 of their partners
- 12 Best Practice Research Scholarship scholars + 2 partner practitioners
- 12 additional head teachers (36 in all)

- 13 'brokers'
- 11 schools in the 'Leading Aspect' programme.

The report contains several methodological cautions, which we do not elaborate here. Schools in urban areas of high social deprivation are under-represented in our sample because, understandably, most of them refused or ignored requests to be involved in the research. The funding arrangements usually meant that 'lead schools' acted as 'gatekeepers' to their 'partner schools', which made it hard to access partnerships that may have failed or run into difficulties. Within schools, we were obliged to spend more time with senior managers than classroom teachers. Moreover, it would have taken several visits to establish relationships with teachers that might have engendered sufficient trust to engage in observation; and that was beyond the resources of the project. As a result, our report relies heavily on accounts rather than observations of teaching and practice sharing. This is highly problematic because:

- Claims about practice transfer are often subject to idealisation and simplification about process and impact.
- Large parts of what teachers do are not accessible through discourse, but will always remain tacit.

Observation was to have been the main research tool that could give us access to different representations of practice, yet it proved difficult to observe lessons as originally intended.

Preconditions for transfer of practice

Transfer processes may be initiated and/or implemented at the level of school, individual teacher or department/team, but usually involve all three to some extent. At every level, good pre-existing relationships are advantageous and accelerate the onset of some form of mutual exploration and/or negotiation. Schools with similar student populations and not competing for students were the most likely to embark on a transfer process, but there were exceptions. Implementation or serious piloting of a new practice usually required establishing mutual trust between the participating teachers but, again, there were also a few autonomous importers of ideas.

To understand why relationships are so important, it is useful to distinguish between an idea and a practice. Typically, a person 'picks up' an idea and then decides, either independently or in consultation with others, what, if anything, to do with it. It can be helpful to understand the context from which the idea came and how the idea was used in that context, but the idea cannot be implemented without building it into existing practice or constructing some new practice around that idea. The implementation process requires far more learning and creative work than just the finding and assimilation of the idea.

Practice, however, is a holistic concept, encompassing everything with which a practitioner is engaged during a particular episode or longer period of work. It is a complex chunk of action, incorporating many combinations of personal knowledge and skills, for a range of purposes, some well defined and others often implicit (Eraut 1994, 2004). Although many new variations may feature, the core of a practice has usually been developed and honed over a considerable period of time, and it operates at many different levels of consciousness. Many aspects of a practice are routinised and taken for granted, they seem to work and do not need to be explained. But even these aspects may be regularly adjusted, even daily, to meet the ever-

changing situations that characterise classroom life. Monitoring oneself and one's environment is always a critical dimension of practice, and one's understanding of that environment and other people's actions within it underpins the continuous fine-tuning and continuing development of one's professional practice.

Thus, teaching practices are the product of teachers developing their craft in particular contexts over long periods of time. Moreover, the interpersonal nature of most of their work requires that they commit themselves to it, so that their practice becomes part of their personal identity as well as their professional craft. This makes teachers vulnerable, because their practice is more open to public criticism than most other aspects of their lives. The combination of challenge, complexity, and vulnerability makes practice a very sensitive area. Some teachers have good experiences of trust relationships with other teachers that have strengthened their professional lives, but even they take time to get to know people before disclosing much about their practice. Other teachers have had little or no experiences of such trust relationships with colleagues, and are naturally suspicious of those who want to discuss practice in any detail. Indeed, in many professional contexts, practitioners develop, not necessarily consciously, a deceptive discourse about practice, which enables them to discuss their practice in a manner that develops affinity with colleagues and mutual emotional support, but remains sufficiently vague to protect themselves from criticism and protect their autonomy. Such discourses run counter to any research or evaluation of practice, or any attempts to describe practice in a way that might promote mutual learning with colleagues (Eraut 2000, 2004).

Even observers can discern only a small part of what is going on in a classroom; and their representation of what they observe will be fragile and partial unless they spend a considerable amount of time there. Teachers get to understand other teachers both by spending time together in classrooms and by engaging in joint activities. What they learn is much more than what is said, and more than they can explain. Hence, transfer of practice, rather than transfer of ideas about practice, requires not only observation but also some mutual engagement.

Our report devotes considerable attention to the nuances involved at all levels, when there are attempts to transfer practices between classrooms in different schools. It cites many examples of both positive and negative factors. We concluded that there are four sets of contextual conditions that are usually required for the successful transfer of practice:

- Good trust relationships between the individual teachers concerned, and often also between their schools or departments;
- Respect for the personal and professional identities of the teachers, and recognition of the intrusiveness and demanding nature of the transfer process;
- Recognition of the teachers' preparedness to risk their time and reputation, and sensitivity towards their vulnerability;
- Understanding the need for learner engagement and creating the time needed to construct new practices in 'receiving' classrooms.

The process of transfer

We found it useful to distinguish between three types of activity that contributed to the process of transfer at classroom level: experiential learning, reflective learning, and contextual support for learning.

Experiential learning

With experiential learning, the partner is directly involved with the practice being transferred from a more experienced person. Examples are described below.

- **Lesson observation.** Though originally conceived in terms of potential partner teachers observing model lessons to develop their practical understanding of an originating teacher's practice, we also found observation being used by originating teachers (especially Advanced Specialist Teachers), when visiting a partner's classroom, to enable them to discuss the next steps in their partner's practice development.
- **Joint planning.** Observation often led to sharing perspectives on the lesson(s) observed and joint planning of future lessons.
- **Practical workshops.** These were used in areas such as drama or thinking skills, and involved potential partner teachers actively participating as learners in activities like those used in the originating teacher's class. Partners gained sufficient understanding of these practices to be able to introduce similar practices in their own classrooms.
- **Co-teaching.** The most challenging activities were probably those where the originating and partner teachers worked alongside each other in a classroom, usually but not exclusively in the partner's school.

Reflective learning

In reflective learning activities, the practice provides a basis for reflection, but the partner does not directly experience the practice. Such activities include:

- **Meetings to discuss shared problems or issues for development.** For example 'We're trying to develop this aspect of our work. What do you do in your school?'
- **Seminars, theory/idea-based CPD training and education.** Such activities were normally a combination of ideas or theories underlying good practice and practical examples of how these ideas could be used in the classroom;
- **Discussing good practice with other teachers.** In formal situations (e.g. shared CPD) 'carousel' type arrangements allowed teachers to tell one another about aspects of their classroom practice they felt positive about. Significantly, exchanges like these were often accompanied by a sharing of resources, which allowed them to describe their practice more clearly, and provide practical help to other teachers;
- **Attending to affective dimensions of learning.** Some originators suggested that their work was as much about being supportive and enhancing colleagues' confidence, as the exchange of technical ideas.

Contextual support

Contextual support for learning refers to activities that play a key role in improving the conditions in which transfer may take place. We observed three kinds of such support:

- **Acquisition and development of teaching and learning resources.** This entailed advice on purchasing, the exchange of materials, or advice on prototype materials being developed by partners.
- **Technical support.** This was often found to be valuable, especially in ICT.
- **Developing policy.** At a more strategic level, teachers described enhancing and developing policy. This was significant in supporting and acknowledging practice internally, and in providing documentation for external inspection.

Our research suggests that schools engaged in transfer activities should attend to all three of these dimensions. However, we also observed exceptions. For example, teachers described one-off events that 'spoke to them' and from which they had made changes to their practice. In doing this, they used their own creativity and experience to understand what the ideas they heard would look like in practice, and found ways of developing practice in their existing environment. Conversely, in some negative examples, ideas and ways of working included in carefully planned ongoing activities were rejected by teachers, because they believed that these practices would be impractical in their own school's context (Doyle and Ponder 1977). It is also possible that the challenge was too great for them to contemplate.

Incorporating experiential, reflective and contextual supporting activities attends to different aspects of teachers' practical knowledge. As suggested above, experiential activities can begin to engage with this tacit knowledge by allowing teachers to work alongside one another and develop an understanding of good practice, in a practical and physical sense, rather than through formalised representations or descriptions of practice. Reflective activities not only enable teachers to develop knowledge of technical and formal aspects of good practice, but also attend to the purposes of any change to practice, or to the affective dimensions of changing practice. Opportunities are thus provided to review and celebrate progress, or for partners and originators to think about areas in which their own practice could be developed. In our fieldwork, teachers raised the importance of all of these types of activity and drew on a range of examples of how both had influenced their practice. Our research did not indicate that one of these approaches should be prioritised over the other.

'Transfer of practice' or 'joint practice development'?

Where teachers are developing new practices, it is rare for them to replicate the good practice of others. In our fieldwork, teachers were more likely to describe the extension and refinement of their existing repertoire of practices, through collaborative and affirming work with other teachers. Teachers saw themselves as having 'travelled' or 'grown' in their work. However, both originators and partners were reluctant to label themselves as 'experts'. Teachers' language suggested a developmental process that must meet receptive partners in order to take root, and must engage with some embedded values that appeal. The picture that emerges does not portray discrete 'blocks' of practice that can be passed from teacher to teacher.

This led us, in many cases, to question whether this joint work with other teachers should be labelled as 'practice transfer' or whether 'joint practice development' would provide a better description. This would both avoid denigrating the existing practice of teachers trying to learn new ways of working, and acknowledge the effort of those who are trying to support them, as well as recognising their having developed creative ways of working on the complex task of opening up and sharing practices with others.

In the case of individual learners, it was particularly apparent that new ways of working did not represent leaving behind old practice. More frequently, an innovation in the teacher's work was adapted to incorporate existing aspects of work, or allowed teachers to expand their repertoire of available approaches. Toward the end of our research, teachers in our case studies drew on a variety of sources in working out changes they aimed to make in their classrooms. The teachers' decisions as to which elements of any suggested innovation would be adopted was based on their judgement of the likely outcomes, predicted through a practical working knowledge of the students they work with.

The literature drew our attention to 'coaching' as a means of enabling teachers to develop their practice through the support of colleagues and external agencies. Many of the characteristics of this approach resonate with the features of joint practice development described above. Both exemplify an extended professional relationship between two practitioners involved in experimentation and critical reflection in the classroom. Both exemplify the asymmetry of expertise commonly encountered among those involved. The relationships between teachers engaged in joint development of practice were only rarely symmetrical. One teacher usually had some claim (frequently through external judgements) to be a better teacher and was seen as the 'originator' from whose expertise the practice of the 'partner' could benefit.

Much less frequently, we encountered teachers working with colleagues from other schools in the joint development of practice in symmetrical relationships. However, there were some important exceptions. One of the teachers we interviewed was engaged in a Leadership Incentive Grant cluster, where teachers investigated and experimented with different ways of working, and then shared challenges and progress with colleagues. He described the process of developing and changing practice as one of 'chipping away at things'. The outcomes of the work weren't specific to the introduction of a single innovation; instead, a range of ideas contributed to his existing practice. If we were to focus on joint practice development and enhance teachers' access to peer-to-peer networks, we might conclude that 'flat' relationships are likely to provide opportunities for a far wider spectrum of teachers to articulate and explore their own practice, and to support colleagues in the same process.

Another argument for supporting the idea of joint practice development is that we consistently found that, in asymmetric relationships, the originating teachers also learned a great deal, both through having to articulate their practice more clearly and from their engagement in joint problem-solving with partner teachers. Often the originating schools derived greater benefit from the arrangements than their partner schools! In all successful partnerships, one might expect to find that both partners:

- Learn a great deal and improve their practice
- Gain from observers' feedback

- Engage in joint planning and problem solving
- Share teaching and learning resources.

In asymmetric partnerships, however, the more experienced partner is more likely to assume the role of consultant or coach.

What is needed for the facilitation of joint practice development? The findings of our project suggest the following prerequisites:

- Understanding and accepting the challenges of partnership;
- Understanding and accepting the challenges, both cognitive and emotional, of developing new practices;
- Provision of personal and group support for those engaged in partnership work for developing and formatively evaluating new practices;
- Agreement between the partners and their schools on first exploring, then developing a plan for joint practice development;
- Adequate time and resources to give the chosen project(s) a reasonable chance of a good outcome.

In addition, the project devotes considerable space to the ways in which heads and senior staff provide support for the transfer processes described above. These include:

- Strategies for reaching agreement on student needs, so that the investment required for supporting transfer can be fully justified;
- Setting the tone of the school, so that teachers feel able to take risks;
- Distributing leadership, so that senior teachers can give practice development and its facilitation more attention;
- Building networks as a basis for finding appropriate and motivated partners;
- Creating time for teachers to engage in joint practice development.

The role of external brokers in developing partnerships, networks and consortia is also discussed, as well as their help in finding resources for joint work.

What do we mean by 'good practice', and how do we evaluate it?

Our earlier discussion emphasised the holistic nature of a teacher's practice; this suggests that the most sensible entity for discussing what we mean by good practice is the total practice of one teacher with one class. However, when we come to discuss the transfer of practice between two teachers, we do not expect any clones. The focus of the practice being transferred could be:

- A new addition to a teacher's repertoire;
- A principle (such as equality) or a general concept (such as constructivism) that is capable of being interpreted in a wide variety of ways;

- A set of skills developed over time, such as learning to learn, critical thinking or learning in groups;
- A cluster of activities and/or learning materials that constitute the main body of the curriculum for a particular subject.

How can we reconcile these two views of practice, the holistic, which cannot be separated from its personal creator, and the abstracted, transferable entity? From an evaluative perspective, it is the holistic practices of the two teachers that matter, because that is what the students experience. But we also know that new practices take time to develop, and that it will take at least one year, if not two years, for the receiving teacher to have integrated and fine-tuned a new practice of any substance and complexity. So, by the time it is appropriate to evaluate their newly created practice, all those concerned will have already made up their minds!

However, this is not the only difficulty. The first response of policy-makers to the question of credibility is to look for independent evidence of practice outcomes such as national examinations and standardised tests or inspection reports. The examinations and tests are probably the most important, because they have a strong influence on Ofsted reports and there is no reverse influence. People may debate whether the examinations and tests measure the intended range of outcomes, but that was not the focus of our project. What does concern us are two more technical questions:

- What learning is covered by these examinations?
- To what body of practice should the results be attributed?

The more appropriate value-added data make the answer to the first question very clear. The examinations and tests provide value added data that purports to measure all the learning in a particular subject that occurred in or out of school between two specific dates. Thus it refers to the outcomes of two, three or four years of practice in the subject, resulting from the work of all the teachers and others who contributed to that learning, including the students themselves. Hence value-added data cannot be attributed to the work of any individual teacher. What examinations and tests do provide (though many would argue over too narrow a range of accomplishment) is evidence for analysing students' needs. This, in turn, provides the justification for seeking to learn and create new practices that promise to meet those needs better.

What then is the role of research? Many researchers are interested in finding generalisable aspects of good practice, because they recognise that some aspects of practice will always be personal, and that holistic practice is not generalisable. They therefore seek to find features that make a difference to students' learning. This entails comparing otherwise similar practices with and without their focused feature(s). This can be very difficult to find, so they often try and design an experiment to focus specifically on this feature. The experimental and control groups then have to be sufficiently similar for the difference attributable to their key feature to outshine other differences between the two groups. The problem for most teachers engaged in transfer, however, is not only lack of access to such research, but that very few of the features that characterise the practice being transferred have been shown by research to add value in contexts such as their own.

How do teachers evaluate the effectiveness of those new aspects that they incorporate into their practice?

The practice that we saw being shared and transferred between schools was usually a bite-sized chunk rather than a large slice, too small for its contribution to students to be expected to show. So what kinds of evaluation are appropriate for judging the successful transfer of bite-sized chunks of practice? This depends on the difference between the new practice and the practice that preceded it. What precisely has changed in the classroom? If the content is new, then there will be no examples of previous student work available for comparison. So the implicit standard may be the quality of work of students in the originating class. This may not be a valid comparison for two reasons: differences in the student population, and the contrast between practices in their first try-out and in their later, mature form. We have to consider the delayed effect of introducing a new practice caused by the time taken by both teachers and students to get used to it and maximise its potential.

If the practice involves teaching similar content in a more challenging way, then the possible benefits may include any of the following changes:

- A larger number of students demonstrate their understanding of the topic;
- A significant number of students develop a deeper understanding of the topic;
- Students are beginning to develop more critical approaches to texts or experiments;
- Students are beginning to learn in a more productive or independent way.

In theory, the first two types of change can be evaluated by comparing outcomes for the topic with those from previous cohorts of students. However, this assumes that the new practice incorporates activities comparable to the previous practice in its later stages or final assessment; whereas, in many cases, the old pattern of assessment might be incompatible with the new practice. The second two types of change are more difficult to evaluate, because their success depends on longer-term changes in students' approaches to learning. It could trigger more general changes in both teaching and learning that have a long-term effect, or the impact might prove to be only temporary. The problems of measuring such changes might also be beyond the assessment expertise of most teachers.

The notion of 'impact' as a criterion of success is strangely ambiguous. We have heard the word criticised as technicist, yet it also captures an important motivational dimension of teaching and learning appreciated by those in the arts. Although originating teachers attribute students' motivation to the activities they are now sharing with other teachers, their infectious enthusiasm is also part of the story. Even when partner teachers cite improvements in test results, they usually attribute them to the motivating and challenging properties of the new practices being transferred, and their impact on students previously perceived as problematic. Indeed, students' motivation appears to be enhancing that of their teachers. That is what they mean by 'practices that work'.

We also know that teachers' enthusiasm can positively affect their students. Whichever sparks first, both are required to sustain successful learning over time. Good practice depends both on sustaining motivation and on a more analytic monitoring capability that spots areas of student inattention or difficulty and adopts a problem-solving approach to improving student learning. Thus, on the one hand, teachers may become increasingly motivated as they engage with new

practices and begin to realise their potential, and this motivation may itself have a positive impact on students' motivation. On the other hand, from a long-term teacher development perspective, one could argue that teachers' professional growth depends on their engaging with change under conditions where:

- They see the relevance of the change to their aspirations for improving the learning of the students concerned;
- The magnitude of the change(s) with which they are engaged is sufficient to sustain a continuing challenge to improve their practice, but not so large as to swamp them with more change than they can handle without reducing the quality of their current classroom practice.

A long-term teacher development perspective on the practice development process should also note that a key factor in teachers' learning is their meta-practice of improvement – the way they think about, evaluate or seek to improve their practice. This also includes:

- How they learn from experience and talking to other people;
- Their awareness of their strengths and weaknesses;
- Their ability and disposition to address aspects of their practice in need of improvement;
- Their disposition to expand their repertoire by seeking and trying out new practices and ideas;
- Their ability to manage their time to make room for a developmental dimension to their work.

There are clear advantages to engaging in mutual support and joint activities, in which two or more teachers are engaged with an identical or similar practice development challenge. Some of the new practices mentioned above demand continuity over three or more years. Thus, practice development for groups of teachers within a department may be more effective in the long run, even though it is sometimes tactically wise to start with a small group of volunteers and then seek to expand it. There is also a meta-practice at department level, which needs to include:

- How the department/team supports the practice development of individual teachers;
- How the department/team finds out about, assesses and develops new departmental practices;
- How the department/team connects with and contributes to practice evaluation and development at school level.

These approaches to long-term practice development require a school culture and management style of the kind described by some of the headteachers we interviewed. Engaging in the sharing and transfer of practice may be one of the best ways of developing such a culture, one that is easier to pace and develop incrementally than attempting a more radical upheaval. In this context, transfer of practice is seen as part of a wider school development strategy, rather than a goal in its own right. It can also be seen as just part of growing a sharing relationship within partnerships and consortia that have yet to develop their full potential. Once more, there is a need to consider a longer timescale than is currently common.

Decision-making during transfer

It is important to look at research and theorising outside education to understand that these problems are not specific to schooling, but are typical of most areas of human activity. The period after the Second World War was characterised by increasingly ambitious claims for social science research, which have gradually been replaced by more modest goals. These centred around a dominant paradigm that conceptualised decision-making as a choice between a clearly defined set of options on the basis of strong evidence. Thus, decision-making was defined as a purely analytical activity, even though the related process of problem-solving was often seen as involving creativity as well as analysis. This approach was particularly influential in business, where decision-making came to be viewed mainly in financial terms, because that was where the experts promised to provide the most reliable evidence and the most plausible reasoning. In the intervening 50 years, critics of the model have pointed out that this approach over-simplified complex situations, and that strong evidence was only rarely available on some important aspects of the decision context. The decision-making model itself, however, remained unchallenged until the 1980s, when the parallel development of computer-based 'expert systems' by cognitive scientists finally prompted a series of research studies to investigate how recognised experts actually made their decisions.

It soon became clear that experts did not use the paradigmatic model. Several alternative models were developed to interpret these new findings, some more relevant to particular contexts than others. However, there was general agreement that the salient features of most 'naturalistic' decision-making settings included the following:

- Problems are ill-structured;
- Information is incomplete, ambiguous, or changing;
- Goals are shifting, ill-defined or competing;
- Decisions occur in multiple event–feedback loops;
- Time constraints exist;
- Stakes are high.
- Many participants contribute to the decisions;
- The decision-maker must balance personal choice with organisational norms and goals (Orasanu and Connolly 1993).

Once these contextual parameters had been identified, it became easier to understand why the decision-making processes observed in natural settings were so different from those advocated by the old paradigm. In particular:

- Experts frequently generate and evaluate a single option rather than analyse multiple options concurrently;
- Experts are distinguished from novices mainly by their situation assessment abilities, not their general reasoning skills;

- Because most naturalistic decision problems are badly structured, decision-makers choose an option that is good enough, without continually striving for the best;
- Reasoning and acting are interleaved, rather than segregated (Weick 1983);
- Instead of analysing all facets of a situation, making a decision, and then acting, it appears that in complex realistic situations people think a little, act a little, and then evaluate the outcomes and think and act some more (Connolly and Wagner 1988).

The implications for knowledge use are that: (1) the relationship between knowledge and decision-making is rarely simple; (2) good decision-making is critically dependent on how the decision is framed by the decision-makers in the light of their situational understanding; and therefore (3) the balance is tilted more towards the personal knowledge of the decision-maker and less towards the codified knowledge management system than might be implied by classical decision-making theory (Eraut 2004). Politicians and business leaders have little difficulty in recognising this scenario.

Teachers have a similar problem, because the evaluations they are required to carry out for school accountability purposes do not contribute to the decisions they make when seeking to improve their practice. Thus, research into naturalistic decision-making is particularly relevant to teachers contemplating transfer. Their process is more likely to be a series of small decisions to find out more, or to engage in some direct experience to get the feel of the new practice, rather than a single irreversible decision to adopt another teacher's practice wholesale or to go for a pre-planned adaptation without testing things out first. Thus, the affective dimension of transfer revealed by the importance of relationships is accompanied by a gradual process of engagement with the new practice, through developing some understanding of what it is like to teach that way, how best to involve one's students, and what difference it might make to the motivation and achievement of one's class, particularly for those members of it who are a current source of concern.

The metaphor of 'courtship', often used to describe the relationship between teachers from originating and partner schools, could equally well be applied to the partner teachers' courtship of the new practice and even, perhaps, to the partner teachers having to renew their courtship of their classes.

If one accepts this rather different framing of the transfer opportunities, then the kinds of evidence that affect a receiving teacher's ongoing decision-making are likely to be of the following kind:

- The teacher begins to feel that she can learn the new practice;
- The teacher becomes confident that the originating teacher can offer support for this learning in a flexible way that suits her personal approach;
- The teacher feels that the new practice is compatible with the kind of relationship she seeks to develop with her students;
- The teacher notices that the new practice attracts a high level of student participation, first in the originating school, then in her own class;
- The teacher notices that students are motivated by the new practice;

- The teacher sees ways in which she could adapt the practice to fit the needs of her class more closely;
- The teacher sees evidence of students learning important aspects of the curriculum;
- The teacher sees how the practice might help her to respond to the needs of challenging students in her class;
- The teacher finds ways of fine-tuning the new practice to the needs of her class, as she becomes more used to it;
- The teacher feels that, once she has become accustomed to the new practice, it will not be significantly more demanding of her time and effort.

We would suggest that future evaluations of transfer work involving chunks of practice should be directed towards refining and improving these natural evaluation processes, so that better decisions are made and more is learned about the transfer process itself.

References

Connolly T, Wagner WG (1988). Decision cycles. In: RL Cardy, SM Puffer, MM Newman (Eds) *Advances in Information Processing in Organisations Vol 3*. JAI Press.

Doyle W, Ponder G (1977). The practicality ethic in teacher decision making, *Interchange*, 8(3) 1–12.

Eraut M (1994). *Developing Knowledge and Competence.* Falmer Press.

Eraut M (2000) Eraut, M. (2000) Non-formal Learning and Tacit Knowledge in Professional Work, *British Journal of Educational Psychology,* 70, 113-136.

Eraut M (2004). Informal learning in the workplace, *Studies in Continuing Education,* 26(2) 247–273.

Fielding M, Bragg S, Craig J, Cunningham I, Eraut M, Gillinson S, Horne M, Robinson C, Thorpe J (2005). *Factors Influencing the Transfer of Good Practice,* DfES Research Project Report, Department for Education and Skills.

Orasanu J, Connolly T (1993). The reinvention of decision-making. In: GA Klein et al. *Decision Making in Action: Models and Methods.* Ablex.

Weick K E (1983). Managerial thought in the context of action. In: S Srivastva (Ed) *The Executive Mind.* Jossey-Bass

Research in teaching and learning: The NRDC Effective Practice Studies

John Vorhaus, National Research and Development Centre for Adult Literacy and Numeracy

This paper describes five related projects designed to identify effective teaching and learning practice in the areas of reading, writing, numeracy, ICT and ESOL. The projects run from 2003 to 2006, and build on methodological insights from the influential American Institute of Research report, What Works Study for Adult ESL Literacy Students[1]. Methodologies and instruments are planned in common across all projects, with the partial exception of the ICT study, which places more emphasis on 'development' than the others, in response to the state of the field.

Effective practice in the teaching of reading to adult learners

Researchers: Greg Brooks, Colette Beazley, Maxine Burton, Pam Cole, Judy Davey, Richard Finnigan, Sam Roberts, Yvonne Spare, Marcin Szczerbinski, Jan Wright

This is a correlational study, assessing adult learners' attainment in and attitudes to reading, and observing the strategies their teachers use. The aim is to correlate learners' progress in reading and changes in their attitudes with their teachers' strategies, and thus acquire some understanding of what enables learners to make progress and/or develop more positive attitudes.

The project is intended to benefit both teachers (by providing some clear answers to the question 'what works?') and learners (by enabling teachers to adjust their teaching where necessary). Some potential findings may also benefit trainers (for example, if there is a clear message about whole-class v. individual work or some amalgam) and funders (if, for instance, there is a clear message about the benefit of sustained attendance).

Research methods

Data was gathered from 182 learners in 23 classes at 21 sites in 2003/04, and our targets for 2004/05 have been to add 320 learners in about 40 classes at more than 35 sites (with only a few sites and no learners repeated from 2003/04).

Background data on learners has been gathered via a profile, and data on attitudes and attainment obtained from learners at three points during an academic year. Four teaching sessions were observed in each class between the first and second assessments.

All data was first analysed descriptively; for example, in terms of gender, average levels and ranges of attitudes and attainment, general and specific teaching strategies observed, with what frequency. The four sets of data (attitudes, attainment, general strategies, specific strategies) were then correlated and the longitudinal trend of attitudes and attainment was analysed. Multilevel modelling has been used to analyse which strategies are most strongly predictive (in the statistical

[1] Condelli, L., et al., What Works Study for Adult ESL Literacy Students, American Institutus for Research, 2003.

sense) of changes in attitudes and attainment when less significant factors and background characteristics were partialled out.

In the reading study, we pre-tested 182 learners in 2003/04 and had complete data on 119 by the end of the year. This year we are aiming to pre-test up to 320 and to have an achieved sample with complete data over the two years of at least 250. The 2:1 ratio between target numbers at pre- and post-tests is based on the usual drop-out rate of 50% in adult basic skills provision (though in 2003/04 we achieved 65% retention in our project).

Each learner submits to about an hour of assessment at each stage. In addition we are gathering eight hours of observation data on each class, and then carrying out very detailed analyses, namely general teaching strategies for each lesson at group level, specific teaching strategies (analysed in minutes) for each lesson at individual learner level, correlations between progress in reading and changes in attitudes, and between each of those and both general and specific teaching strategies. When analysing down to individual learner level, we are using state-of-the-art multilevel modelling to separate learner-level from class-level (that is, teacher-level) effects.

Findings to date

From the learner profiles:

- 45% men, 55% women;
- 23% aged 16–29, 50% aged 30–49, 27% aged 50-plus;
- 10% English as additional language;
- 81% white ethnicity;
- 42% employed, 10% unemployed, 48% unwaged (in education, looking after home/family, sick/disabled, retired);
- 9% left full-time education before age 15, 5% after age 19;
- 52% said they had no educational qualifications;
- Exactly 50% said it was their first year on the course;
- 61% had taken a course within the last two years.

Tutors reported that 52 learners (29%) had dyslexia. Of these, 21 had been formally assessed and the rest were judged by their tutors to have dyslexia; no one was said to have been formally assessed as not having dyslexia.

Comparison of the profiles of the learners for whom we have full data ('returners') against those who dropped out shows a few differences – these will be taken into account in calculating results. Statistical tests of the attitudes questionnaire data for the three occasions show a small but significant increase in confidence. Analysis of the reading assessment data for the three occasions for the returners is ongoing.

We based the need for an achieved sample with complete data of 250 on Larry Condelli's ESOL

study in the USA[2]. He confirmed that this was adequate for a correlational study of this sort.

Even the sample of 119 with complete data is yielding interesting findings. It has been confirmed that better progress is correlated with more hours of attendance, and the attitudes questionnaire shows an improvement over the year in learners' self-confidence. Some strategies we had thought might feature rarely occur, and there is a very strong tendency for teachers to adopt one of two approaches – a whole-group section followed by individual work, or all individual work. Once we have two full years' data, we shall be able to make some very strong statements about average progress made, changes in attitudes, general and specific teaching strategies, and the correlations among all four, all cast against learners' background characteristics and (in the strongest analyses) partialling out the effect of both background factors and teaching strategies that rarely occur or make little difference.

Further work

The pre-, mid- and post-assessment and questionnaire data will be analysed, using multi-level statistical modelling (to separate individual- and class-level effects) on the data from the 119 returners. Progress in reading and changes in attitudes will then be correlated with both general and specific strategies, in order to investigate which strategies are associated with progress/change. We will also correlate progress in reading with changes in attitudes. Even if the average amount of progress/change is small, and some learners have lost ground, it will still be possible to detect which strategies are associated with progress/positive change and which with standing still/losing ground. Finally, subgroup analyses will be carried out, for example comparing mainly whole-class with mainly individual teaching, or comparison of the progress of male and female learners.

We hope to extend the range of provision in 2005 to include, for example, three prisons and a number of voluntary sector, private sector and workplace/work-based providers.

Effective approaches to the teaching and learning of writing: phase 1

Researchers: Sue Grief, Bill Meyer, Susanna Kelly, Luxshmi Soundranayagram

We present here the findings of three reviews designed to provide a basis for research to be undertaken in phase 2:

- A narrative review that looks at how studies that conceptualise writing and writing development contribute to an understanding of how adult learners develop as writers;
- A systematic review of primary research;
- A review of current practice.

[2] Op cit.

Background

In reviewing what was known about the teaching and learning of writing that might be relevant to adult literacy learners, it was important to include insights from research from a range of sectors. To make the task manageable, two separate literature reviews were undertaken:

- A narrative review drew on a wide range of sources and included studies written in the context of schools or higher education.
- A systematic review looked specifically at primary research in the field of adult literacy and addressed the question 'What factors in adult literacy programmes enable learners to develop effective writing skills?'

In addition, a review of current practice involved four focus groups with teachers and one with learners, as well as nine case studies across a range of provision in eight of the nine English regions, including one penal establishment. Observation of two classes in each organisation, interviews with teachers and group interviews with learners were undertaken by trained practitioner-researchers. The small scale of the review requires that any generalisations drawn from it must be treated with caution. The findings have informed the design of research tools for phase 2 of the project, which will provide more authoritative evidence. Finally, consultation meetings with practitioners, managers and researchers in the field of adult literacy provided invaluable feedback on the findings of the three reviews.

Key points

There is very little evidence about the factors in adult literacy programmes that enable learners to develop effective writing skills. Use of authentic literacy practices in the classroom was the only variable found to be significantly and positively related to changes in learners' literacy skills and practices. A number of variables may be significant in the teaching and learning of writing. In order to test whether these factors can be identified as predictors of learner progress, they have been developed as dimensions of teacher and learner practice for use in phase 2. The dimensions include the degree to which practice involves:

- Authenticity of materials and communication;
- Collaborative approaches to writing;
- Making the process of writing explicit to learners;
- Contextualisation of writing tasks;
- Relevance of teaching and materials to learners' lives.

Learners placed value on the mastery of spelling and punctuation, and identified these as the aspects of writing that caused them the most difficulty. The teachers valued the expression of ideas above, or in addition to, correctness in the surface aspects of writing. However, classroom observations indicated a heavy use of exercises that focused on the sub-skills of writing, which may reflect teachers' concern to respond to the needs of the learners.

The review of literature suggests that writing should be viewed as a process in which the writer interacts with what she has written – planning, physically writing and revising a text. It also suggests that writing cannot be separated from its social, cultural and political context, and that it is connected to the writer's sense of identity.

Phase 2

A full set of instruments and associated guidance papers were created and used during May – July 2004 at Finchdale College (a specialist college in Durham), at IMS (UK) Bradford, and in a Catterick College course for the Army. In each case, the courses were short and intensive, allowing the required minimum of 50 hours of tuition within the period to enable the instruments to be tested. Fifty-four learners undertook both pre- and post-test writing tasks across seven courses taught by nine different teachers; twenty learners were interviewed.

The fieldwork was undertaken in two sites by practitioner-researchers, and in the third by the project researcher. The two practitioner-researchers provided feedback on the instruments and practicalities of the fieldwork. The instruments have also been shared and discussed with all the researcher-practitioners working with the project and with Jane Mace, a consultant who has supported and advised the project from the start. In the light of the feedback received from all of these sources, and discussion with colleagues from other Effective Practice projects, the instruments and guidance have been revised and collated as a full researcher pack.

Thirteen practitioner researchers were recruited to work on this phase of the project. Six worked on the small review of practice undertaken in phase 1; all attended two days of training between April and September 2004.

Eleven new sites have been recruited and researchers have undertaken initial visits. It is hoped that this will provide in the region of 33 classes. The majority of these will be year-long courses starting in late September. Until the first class visits have been undertaken, the exact number of learners to undertake the pre-test writing tasks will not be known. Fieldwork at two further sites is being discussed. Our use of these depends on the level of recruitment to suitable classes and the willingness of teachers and learners to take part. Further sites were recruited in the period up to January 2005.

To date our sample, including the three sites used in the summer term, includes:

- 4 FE colleges
- 3 LEA adult and community
- 2 prisons
- 1 training provider
- 1 voluntary organisation (+1 possible)
- 1 Army
- 1 specialist college.

At the time of writing, we were negotiating with a further five sites and actively looking for at least another five. The original target was 450 learners, based on recruiting 25 sites, with an average of three classes per site and six learners per class. Because of the drop-out factor to which longer courses are inevitably subject, we aim to recruit some more intensive courses to redress the balance. However, there are limitations to this, because of the need for a representative sample of provision.

An ACCESS database – the Teaching And Learning Ipsative Assessment Database (TALIAD) – has been constructed to store information on learners, classes, assessments and observations, and to allow for analysis of class and learner progress against instructional strategy and emphasis measures, taking note of other potentially relevant independent variables.

Effective practice in inclusive adult numeracy teaching

Researchers: Diana Coben, Margeret Brown, Valerie Rhodes, Jon Swain, Debbie Holder, Sue Nieduszynska, Veronica Storey, Jackie Ashton, Cathy Magee

Context

There are numerous challenges in adult numeracy. There is a shortage of experienced teachers, teacher trainers and researchers. The stakes are high: there are an estimated seven million adults with literacy and/or numeracy problems. This is a sensitive issue in a technologically advanced democratic society. The issues are also complex – adults' difficulties may be in literacy, numeracy and/or language.

According to the Smith Report[3]:

> *… the adult numeracy strategy is a challenging and demanding one for teachers and learners alike. Progress could easily be undermined by:*

- *Uncertainties surrounding the teaching and assessment of mathematics in general and in particular the future of GCSE Mathematics and key skills;*
- *The limited pool of competent and confident teachers of mathematics and numeracy;*
- *The lack of employer engagement in raising the skill base of new employees.*

The Skills for Life Survey in England (DfES 2003) found that:

> *… one in four (25 per cent) respondents achieved Level 2 or above in the numeracy assessment, which means that around 23.8 million adults had numeracy scores at Level 1 or below. Nearly one in two (47 per cent) were classified at Entry Level 3 or below in the numeracy assessment, including one in five (21 per cent) at Entry level 2 or below. This means that 15 million adults in England had Entry 3 or lower level numeracy skills and that 6.8 million of these were classified at Entry Level 2 or below.*

The recent themed Inspectorate report (ALI/Ofsted 2003) largely treats numeracy, literacy and ESOL together. In all three areas, the report found a much lower proportion of good provision and significantly more unsatisfactory provision than in any other area of learning. It cited the need for a sharper focus on the quality of education and training. Specifically on numeracy, the report found that:

- Numeracy is taught less frequently than literacy, and there is less demand for numeracy, despite equivalent levels of need (or greater levels of need, according to the Skills for Life survey; see below);
- There is a need for greater expertise in teaching numeracy;
- Numeracy is too often taught by rote learning, rather than by developing understanding of numerical concepts.

[3] Smith, A., *Making Mathematics Count*, The Stationary Office, 2004

Methodology

The numeracy project uses both qualitative and quantitative methods. Data was collected in two phases, January to June 2004, and September 2004 to June 2005. Research methods include:

- Interviews with teachers and learners – individuals and groups;
- Attitude survey (learners);
- Materials development and testing (measures);
- Teachers' diaries and notes of sessions;
- Observations of teaching;
- Assessment of learners;
- Photographs and writing by learners.

Adult numeracy is acknowledged to be an under-researched area, and there have been few models of research to draw upon (Coben, Colwell et al. 2003). Consequently, the first phase of this project was used to pilot and refine techniques, in particular with the observation instruments. The assessments and attitude surveys were administered twice, once early in the course and again towards the end. Each course had between two and four observations, which focused on the relations and interactions between the teacher and learners.

Researchers made use of the Skills for Life observation instrument, and found it not altogether fit for purpose. For example, it has not been able to detect all important episodes and progress with learners with learning difficulties and disabilities.

Sample and sites

During the first phase, sites were located in the North West, South West and London area to give a broad geographical spread. One of the project's main objectives has been to record as many different forms of practice as possible in a variety of different settings.

Classes included Adult Basic Skills mathematics, Return to Employment, foundation ICT, and a vocational taster numeracy course in both day and evening classes. These encompassed a range of providers, including further education colleges, a neighbourhood college, a community group, the Army and a prison. Learners varied in age from 16 to over 70, and were working between pre-entry and level 2. They also varied in terms of ethnicity and gender, and included learners with learning difficulties.

Although 17 classes were recruited initially, only 15 were used in the final sample. Retention has been a problem, particularly in a workplace setting. The retention rate was 65%, although complete data sets (two attitude surveys, two assessments and background information) were not available for all these learners. Complete data sets were collected from 80 learners; 31 classes were observed; and 16 teachers and 53 learners were interviewed. Details of the data are given in the table on the following page.

Number of classes		Number of assessments		Number of attitude surveys		Number background information surveys
beginning	end	admin. 1	admin. 2	admin. 1	admin. 2	beginning
17	15	131	85	136	88	128

Emerging findings

- Taken as a whole, learners made significantly more progress on the second assessment than they did on the first.
- Of all the learner background variables (ethnicity, first language, age, qualifications in mathematics, numeracy course at work, age left school, gender, employment status, motivation for attending the course), only two were significant: ethnic group and first language.
- As there were only very small numbers in some ethnic groups, learners were put into three groups: white British, Bangladeshi, and Other (which included Indian, Pakistani, Caribbean, African and Asian). The 'Other' group made the most progress and the Bangladeshi group the least.
- Learners whose first language was not English made significantly more progress than those whose first language was English.
- A factor analysis showed a very high correlation between the pre- and post-test scores on the attitude survey, indicating that it was a reliable instrument. However, the scores on the attitude survey did not correlate with progress as measured on the assessments.
- More than a third of learners attended the course to prove something to themselves, get their qualification or gain confidence. Almost a third wanted to be helped with everyday activities, such as shopping (see Table 1). Slightly fewer learners hoped it would help them get a better job, and just over a quarter wanted to help their children with schoolwork.

Table 1. Learners' reasons for attending the course (n = 81)

Reason given by learner for attending the course	*Number*	*Percentage**
To prove something to myself	31	38
To get a qualification	29	36
To become more confident	27	33
To help with everyday things (e.g. shopping)	26	32
To get a better job	24	30
To help the children with their schoolwork	21	26
Other reason	15	19

*Learners could give more than one reason.

ICT effective practice study

Researchers: Harvey Mellar, Maria Kambouri, Sally Betts, Kit Logan, Viv Moriarty, Barbara Nance

The first stage of the ICT study demonstrated limitations in current practice in using ICT for adult literacy, numeracy and ESOL learners, even among those charged with taking the field forward (such as ILT champions in FE colleges). Our recommendations to practitioners arising from Stage 1 argued that tutors needed to address a range of issues:

- **Approaches to using technology:** Tutors need to consider why they are using ICT, and to match the way that they use the technology to their aims. Teaching also needs to address the changing nature of literacies in the digital age, as new relationships between media are created and new genres of writing are developing.
- **Teaching:** A wider range of technologies and ways of using them needs to be explored. Greater experimentation with teaching styles and forms of classroom management should be encouraged. Greater attention needs to be given to the range of learning styles and learner preferences, and the ways in which ICT might be used to address these.
- **Collaboration:** Work is needed to develop appropriate ways that learners can work together effectively using ICT.
- **Teaching ICT:** The question of how and when to teach ICT skills needs to be addressed more explicitly.

Stage two examines both the development of literacy and numeracy skills through the use of ICT, and the development of ICT skills. It aims:

- To develop effective ICT-based teaching strategies through a series of trials using ICT with adult literacy, numeracy and ESOL learners, with theoretically grounded ICT task designs targeted at specific learning objectives;
- By generalisation from these effective teaching strategies, to identify effective design guidelines for ICT-based teaching for adult literacy, numeracy and ESOL.

We have worked with the tutors to encourage the adoption of less prescriptive, more collaborative and more exploratory approaches to the use of ICT in teaching; and to address explicitly why they are using ICT in their teaching, and how ICT skills are developed alongside those of literacy, numeracy and ESOL. Initial indications are that the tutors are finding the new approaches useful, and that learners are positive about the new ways of learning.

In parallel to the research and development work, an MA module on ICT and adult literacy and numeracy has been designed through the work with the practitioner-researchers. This has been submitted for approval to the Institute of Education Continuing Professional Development Board, to form part of the Masters degree in Adult Basic Education.

Close collaboration has been established with other projects. Three of the practitioner-researchers and both development officers are tutors within the ABBSU-funded project 'Exploring e-learning for literacy, numeracy and ESOL teachers'. Another practitioner-researcher is working on the EU-funded M-Learn project, as part of her role within Cambridge Training and Development.

Methods

Working with a group of nine tutors, we developed teaching interventions that built on the recommendations of an earlier project, in which we carried out observational studies on current practice in the use of ICT in adult literacy, numeracy and ESOL. Involvement of tutors was to be a key element in deciding on the development of the ICT interventions, and in the research process, through reflection on their developing practice using ICT.

At monthly meetings with the tutors, we discussed progress and examined a range of theoretical perspectives on the use of ICT in adult learning. Between these meetings, two development officers visited the tutors in their classrooms and worked with them in developing their practice. Each week, the tutors completed an online reflective diary describing their progress with their classes and their understanding of the role of ICT in their learners' learning. Each term, the tutors drew up an intervention plan outlining their intended use of ICT. The plan included the rationale for the use of ICT, a list of learning objectives, and an assessment strategy to measure success in meeting these goals.

The projects

- **e-Portfolios.** Using OPEUS e-Portfolio, learners designed websites in which they stored and displayed their work. OPEUS was chosen for its ease of use, and its ability to display students' work instantly and easily reproduce students' paper-based work. Students found OPEUS challenging, but developed their understanding of ICT and became very confident users.
- **WebQuests.** Various forms of WebQuest were used successfully with level 2 ESOL students. A successful design for use with level 1 learners was developed based around local sites. This provided an element of familiarity, which helped to overcome the difficulties caused by the low level of language skills.
- **m-Learning.** ESOL and literacy learners used camera phones and handheld computers with mobile phone functionality and built-in cameras to send text, image and sound messages or e-mails to a website, to create a mobile photolog/weblog or moblog. The intervention was presented as a project during which learners acted as photo-journalists and covered a college open day.
- **Community provision.** The attraction of learning to use ICT brought in the learners for this community-based project at an under-used UK Online Learning Centre. The use of DfES learning materials in conjunction with the BBC Skills-Wise learning resources developed learner's literacy skills as they prepared for the level 1 and 2 national tests online.
- **Digital video.** The project aimed to consolidate students' language skills and develop ICT skills through working with a digital video camera. Learners found this to be a powerful way of working with language, and of collaborating in teams.
- **Tablets.** Learners used tablets in a classroom where access to IT would not otherwise have been possible, to plan and practise presentations as part of their assessed work. The project looked at the extent to which the use of portable devices can help increase collaborative learning and communication with minimum disruption to an existing class.

- **Mind mapping.** Students used mind mapping software (Inspirations) to focus their thinking, share ideas and help plan written work. Inspirations proved to be a useful tool, as students were able to put down their ideas and then structure them. It also facilitated discussion. These maps could be saved directly into Word for use as a basis for writing.
- **Cyberlab.** In this course, literacy and numeracy were embedded into topics including science, technology, history and culture, within a framework of ICT-enabled activities. The intervention included a mix of constructional, instructional and exploratory teaching and learning approaches. It included the use of Logo and robots when investigating algorithms, and digital cameras to take pictures of real life examples of fractals.
- **Family numeracy.** The intervention investigated how website resources could be used for assessment and practice of parents' maths skills. The workshop was tailored around needs identified by the assessments. A tablet computer enabled parents to model calculations. Strategies they could use with their children were discussed and practiced, using online children's games and school software.

Developing tutors' skills

The development officers worked with the tutors to identify appropriate pedagogical approaches for individual contexts. They were also able to focus aims and objectives as the projects developed. Tutors were familiar with the Basic Skills curriculum and were skilled at mapping competences in this domain, but became more able to link ICT skills to ESOL and Basic Skills as they developed their programmes.

One of the most important changes has been in pedagogy, which has moved away from a didactic style to enable students to become more independent learners. Very often, tutors were interested in promoting collaborative practices through the use of ICT. This way of working also had implications for teaching, which tutors became more aware of.

There was an awareness of 'softer' outcomes, such as increased confidence and self-esteem and the inclusion agenda. Perception of individual needs, combined with more systematic assessment skills, enabled a more powerful understanding of teaching and learning to be developed. Through the monthly sessions at the Institute of Education, and the support of expert tutors, the tutors became more practised in observing specific learning and ICT outcomes, and noting them.

Trials

Based on the interventions identified during the previous year, we are running a series of trials in literacy and ESOL between October 2004 and March 2005 with the tutors and with 'buddies' whom they are helping to implement similar schemes of work. The development officers will continue to support interventions, but their main focus will shift to observing the impact of these interventions on learners.

Learners were assessed in October and again in March, using a specially developed ICT test and a literacy or ESOL test. Observations of learners will take place on at least four occasions in each class. To observe tutor development and throw further light on learning, tutors will carry out critical incident interviews with their 'buddies'.

Impact

The ICT Effective Practice study has been influential in developing effective models of ICT use in adult literacy and numeracy practice. A number of the participating tutors went on to work as mentors on the LSDA e-learning project, disseminating their expertise to a much wider group of practitioners.

The study has also been influential in developing models for professional development, and has resulted in the development of an MA module in ICT and adult basic education.

Effective practice in the teaching of ESOL to adult learners

Researchers: Mike Baynham, Celia Roberts, Melanie Cooke, James Simpson, James McGoldrick, John Callaghan, Catherine Wallace, Martin Bygate.

The ESOL Effective Practice Project is a correlational study, combining quantitative and qualitative data collection. There were four guiding research questions for the project:

- What is the range of naturally occurring variation in the teaching of ESOL to adult learners?
- What correlations can be established between different pedagogical practices and learners' progress?
- What is the impact of course length and intensity on student achievement?
- What lessons can be drawn for effective pedagogical practice in adult ESOL?

The project is an investigation of the range of approaches to the teaching of ESOL to adult learners, with a particular emphasis on the teaching of spoken language as a distinctive characteristic of ESOL pedagogy. Wherever possible, it will establish correlations between the approaches identified and student progress and will make recommendations on effective practice.

Aims

- To establish the currently held principles of effective pedagogical practice in language teaching and learning, evaluating their appropriateness for adult ESOL provision in England;
- On the basis of this, to observe the range of pedagogical practices occurring naturally in classroom settings in the teaching of ESOL to adults in England, with an emphasis on identifying and documenting good practice from both theoretical and practice-oriented perspectives;
- To document the progress made by adult ESOL students in these exemplary classroom settings;
- To establish correlations between particular pedagogical practices and student progress;
- To provide proposals for addressing gaps in the pedagogical repertoire and/or available testing procedures;

- To draw implications for policy, provision and further professional development of ESOL practitioners.

Research methods

We use a multi-method approach, designed to collect quantitative data in order to establish correlations between the effectiveness of particular methods and teaching strategies in relation to student outcomes, as measured in before and after assessments; and qualitative data to provide in-depth accounts of the effective pedagogy identified. Data collected includes:

- Learner background questionnaires;
- Teacher background questionnaires;
- Pre- and post-observation cycle assessment scores;
- In-depth ethnographic-type interviews with learners using bilingual interviewers;
- In-depth ethnographic-type interviews with teachers;
- Observation field notes;
- Coding of specific teaching strategies;
- record of general teaching strategies;
- Record of opportunities for learner involvement.

Forty classes – an initial cohort of about 400 students – are being observed across sites that profile the demographic diversity of adult ESOL provision, both urban and rural, metropolitan and regional.

Preliminary observations

Early observations based on classroom observations, assessment results and interviews with learners and teachers include the following:

- Classes showing progress in post-assessment scores tend to have a high degree of learner involvement. Lessons in these classes consistently score highly on those general teaching strategies that encourage extended learner output, respond to learners' concerns as they arise, and provide opportunities for learners to work together. The importance of a high degree of learner involvement and the chance to work collaboratively in groups or pairs, in an environment where 'talk is work' (Baynham and Whitfield 2004) is reflected in learners' comments. Many ESOL adults have few opportunities to practise English outside their classrooms and place a high value on lessons that encourage talk. Learners report an increase in confidence gained after a period of formal classroom tuition, which they had failed to achieve prior to attending class;
- Learner involvement is affected not just by teaching strategies, but by the content of classroom materials. The ESOL core curriculum materials are used by teachers in a variety of ways and to differing degrees. Classes showing low assessment scores tend to be

those that are overly dependent on curriculum materials at the expense of materials from authentic sources or those that are learner generated. This suggests that unmediated use of the core curriculum materials does not inspire learner involvement. At the same time, content – especially talk – generated by students themselves, and coming from within the culture of the class itself seems to foster a greater degree of motivation than content that is 'brought in' by the teacher;

- As seen in the ESOL case studies (Cooke and Wallace 2004), learners generally work very hard to bring meaning to classroom material, however unpromising that may be, and to maximise their involvement in classroom processes. Higher assessment scores obtain where teachers also contribute to this process;
- In-depth interviews with teachers reveal that effective practice in ESOL is not only about specific or general teaching strategies, but also about teaching in learning conditions that are often far from ideal. Most teachers usually face one or more of: large classes, open entry policies, poor resources, multi-level learners, learners placed in inappropriate classes, or inappropriate learning environments, such as open-plan classrooms. Teachers also have to tackle a large amount of bureaucracy and must implement assessment regimes that they may feel are inappropriate for their learners or that do not correspond with their curriculum;
- Effective practice may well be about teacher qualities such as patience, caring or rigour as much as teaching strategies. Teachers and learners often talk about their classroom experiences in this way; they focus on the overall experience of how it feels to be learning in a group, rather than specific procedures or activities, although it may well be that teacher qualities are linked in some way with teaching strategies;
- Interviews with learners reveal the extent of their diversity in terms of their backgrounds, aspirations, level of education and the obstacles to learning many face. This is a challenge for teachers wishing to respond to individual needs at the same time as fostering the optimum conditions for learner involvement through group and pair collaboration. It is a major issue when describing effective practice. A class may include learners with a high level of education in their first language alongside others whose education has been severely disrupted; or learners who recently arrived in the UK alongside those who have been resident for many years but who have failed to acquire English. The question of 'what is effective practice' must also be 'effective for whom?'

References

ALI/Ofsted (2003). *Literacy, numeracy and English for speakers of other languages: A review of current practice in post-16 and adult provision.* Adult Learning Inspectorate, Ofsted and Her Majesty's Inspectorate.

Condelli, L. et al. *What Works Study for Adult ESL Literacy Students,* American Institutes for Research, 2003

Baxter M, Leddy E et al. (2005 forthcoming). *Measurement wasn't taught when they built the pyramids. The report of the NRDC teacher research project into the teaching and learning of common measures, especially at Entry Level.* National Research and Development Centre for Adult Literacy and Numeracy.

Coben D, Colwell D et al. (2003). *Adult Numeracy: Review of research and related literature.* National Research and Development Centre for Adult Literacy and Numeracy.

Coben D, Dawes M et al. (2005 forthcoming). *Financial Literacy Education and Skills for Life.* National Research and Development Centre for Adult Literacy and Numeracy.

DfES (2003). *The Skills for Life Survey: A national needs and impact survey of literacy, numeracy and ICT skills.* Department for Education and Skills. Accessed at: www.dfes.gov.uk/research/data/uploadfiles/RB490.pdf

Smith A (2004). *Making Mathematics Count: The Report of Professor Adrian Smith's Inquiry into Post-14 Mathematics Education.* The Stationery Office.

Swain J, Baker E et al. (2005 forthcoming). *Beyond the Daily Application: Making numeracy teaching meaningful to adult learners.* National Research and Development Centre for Adult Literacy and Numeracy.

NRDC reports are available on the NRDC website: www.nrdc.org.uk

Employer Training Pilots: a platform for progression

Jim Hillage, Institute for Employment Studies

The Employer Training Pilots (ETPs) were established in September 2002, to test the effectiveness of an offer of free or subsided training to employees without a level 2 qualification, with wage compensation (of various levels) to their employers for giving time off to train, plus access to information, advice and guidance. The pilots are administered by local Learning and Skills Councils (LSCs) in partnership with local Business Links and other agencies, through a network of local brokers or learning advisers and training providers. They were originally planned to run in six areas for one year, but were subsequently extended, first to cover 12 areas for an additional year and, from September 2004, to run for a total of three years in the Phase 1 areas, two years in the Phase 2 areas, and for one year in five new areas plus a regional pilot in the North East.

This paper summarises the findings to date from the evaluation of the pilots (Hillage et al. forthcoming) being conducted by the Institute for Employment Studies in partnership with the Institute for Fiscal Studies and MORI. The evaluation aims to assess the nature and extent of the take-up of the ETP offer, and to examine the key issues affecting how the pilots operate in practice, and the experiences of the employers and employees who take part.

The evaluation has two main strands:

- A quantitative assessment, through surveys, of potentially eligible employers and employees in pilot and control areas before the start and at the end of Phase 2, which will finish and be reported in the Spring of 2005;
- Surveys and interviews with the main stakeholders, providers, employers and learners involved in each of the 12 Phase 2 pilot areas, and analysis of management information data for the first two years of the pilots (September 2002 to August 2004).

This paper is based primarily on the latter element of the evaluation.

Findings to date

12,000 employers involved

At the end of the second year of the pilots, almost 12,000 workplaces were involved in the 12 Phase 1 and 2 pilots. Take-up has been stronger in the second year than in the first. Some 10% of eligible employers are involved – 12% in the Phase 1 pilots and 8.5% in Phase 2. Employer take-up levels vary significantly by pilot area. Seventy per cent of the establishments involved employ fewer than 50 people (25% employ 10 or under), 19% have between 50 and 249 employees and 11% employ at least 250. The pilots have attracted almost half the potentially eligible large employers in their areas.

Throughout, there has been strong participation of employers in health and social care – 28% come from that sector. Three-quarters of ETP employers are from the private sector.

Some are among the hardest to reach

Half of the employers involved in ETPs had a previous involvement with a business support agency (such as Business Link); a training provider (such as a further education college or private provider); or a national training initiative, such as Investors in People (IIP) or National Vocational Qualifications (NVQs). One-third had been involved with at least one agency, initiative or provider. The remaining 14% had no previous involvement, and could therefore be classified as among the hardest to reach sections of the employer community. One-third said they had not worked with a training provider and almost one-half had no prior involvement with NVQs.

ETP employers are more likely to have a positive approach to training than other employers. Most adopt a strategic approach to training, express positive attitudes to training and say that they provided at least some training to low-skilled employees.

Some employers would have provided the training anyway

We can infer from ETP employers' past training behaviour, and the fact that most agreed that they would have provided the training in any event, that there is a degree of 'dead-weight' in the pilots. Further evidence comes from comparing the training levels of low-skilled employees in the initial pilot areas with those in comparable non-ETP areas. This analysis suggests that most ETP employers were providing some form of qualification-based training to their low-skilled employees, although not necessarily in the form or to the extent offered in the pilots. A more conclusive measure of the net additional effect of the ETPs on workplace qualification-based training will be available in the quantitative assessment, due in Spring 2005.

80,000 learners signed up

At the end of August, 80,000 learners had signed up to the pilots, around 4% of the eligible total. Again, participation varies among the pilot areas. ETP learners tend to be female, aged between 26 and 45, and working full-time in personal service or elementary occupations. More than 10% are from minority ethnic groups – rising to more than one-quarter in some pilots. This is more than would be expected from the ethnic profile of the pilot areas.

Most ETP learners left school at or before the age of 16 without a positive view of their educational experience. A minority (around one-fifth) of the learners already have a level 2 equivalent qualification and are therefore theoretically ineligible to take part, although obviously they could still benefit from the training.

Only around one-fifth had not had any training in the past three years and, on the whole, learners held a positive view about learning in general and about the training and skill development opportunities in their workplace.

Most opt for NVQ training

Only 11% of learners are working towards a basic skill qualification, despite considerable effort by the pilots and others to attract such learners. Some pilots have been particularly successful at signing up basic skill learners, partly because of a high take-up among their minority ethnic communities for ESOL courses. Other pilots have struggled to convince employers or to get employees to identify basic skill needs. In some cases, basic skills needs have become apparent while a learner is working towards an NVQ, and pilots have adapted their provision to try to provide support.

In the first and second years of the ETPs, approaching 40% of learners working towards an NVQ were taking a course in health and social care or a related subject, although among other learners the range of courses taken has increased.

Employers are attracted by free, flexibly delivered, brokered training

The availability of free (or heavily subsidised) flexibly delivered training is the most attractive feature of the pilots. Hard-to-reach employers appear to be particularly attracted by the brokerage and help with identifying training needs. There is no apparent relationship between the level of compensation on offer and the level of take-up among employers or employees. Other factors seem more likely to explain the differential participation levels among the pilots.

Pilot organisation varies

The pilots vary in the way they are organised, including:

- Their approach to marketing the ETP programme;
- The size and experience of their team;
- The way they engage with employers; for example, through either independent brokers (working in Business Link or co-located with the LSC team) or 'recruiting' providers;
- How they are managed.

Training providers are an important source of employer recruitment

Employers most often get involved with the pilots either through responding to a local marketing campaign, or by being contracted by a training provider. The remainder mostly join up after being contacted by a Business Link or pilot broker.

There does appear to be some relationship between the way the pilots are organised and their ability to get employers or learners to sign up. For example, employer penetration rates are higher in pilots with recruiting providers, although completion rates tend to be higher in areas with independent brokers.

Learners have more choice over taking part than over course of study

In most cases, employees volunteer, or are 'volunteered', to take part in the training. They generally think they have a lot of choice over whether to take part at all, but less choice over what course they do. However, most said the training was relevant to their job.

Assess–Train–Assess operates, but only for a minority

While most learners had an initial skills assessment, only a small minority – around one in six – went through some form of initial assessment in the Assess–Train–Assess model, whereby skills needs are initially identified, existing skills accredited, and subsequent training adapted to fill the gaps.

Training and assessment is delivered in the workplace

Forty per cent of the employers involved in ETPs had their training needs assessed as part of their involvement (and one-third of those did it themselves).

ETP training tends to be organised to suit workplace operations. Two-thirds of learners work with private (non-college) providers. While those who have completed an NVQ take an average of seven to eight months (five to six for basic skills learners), a large number are still in learning after a year, which may eventually raise the average completion time.

We have estimated that it takes around 100 hours for a learner to complete an NVQ course, with roughly half that time spent in contact with the trainer; however, practice does vary by subject. A further quarter of the time was spent outside normal working hours (usually unpaid). The other quarter was spent on independent learning, or evidence gathering, during working hours.

Providers spend about one-third of their contact time (i.e. around 16–17 hours) training, around 40% assessing, and about one-quarter portfolio building. Learners spend nearly all their non-contact time collecting evidence and compiling their portfolio, and working through learning materials. Generally, most time is spent on assessment and evidence collation, rather than on formal training.

The minority who receive IAG like what they get

Just over one-third of learners received some form of information, advice and guidance (IAG), mostly from someone in their workplace or, less often, from their training provider. Learners found the information, advice and (if offered) guidance they received helpful, and were likely to be more satisfied than average with their ETP experience.

17,000 learners have completed

More than 17,000 learners have successfully completed their training including 14,400 NVQ learners. Just under half of those who started in the first nine months of the pilots have completed. Around 20% have left their course; the remaining learners, about 30%, are designated as 'still in learning'.

Most learners leave early because they have changed job or for personal reasons. However, 40% of those who left either found they did not have enough time to train, lost interest, or found the training too difficult – reasons implying that some might have benefited from further support.

Completion is affected by learner, provider and area factors

We have identified three sets of factors that appear to affect completion:

- **Learner-related factors:** Statistically, the group with the greatest chances of completion are learners who are older, male, do not have a disability, work flexibly (ie part-time or on a night-shift) have lower (but at least some) prior attainment and are not a member of a union. Obviously, some of these factors, such as union membership, are related to other characteristics, such as size of workplace or sector. However, we have looked at these factors separately and found no statistically significant relationship with completion;
- **Course-related factors:** For example, learners doing an NVQ in leisure and sport, hospitality or manufacturing had the highest chances of completion; while those on business administration, health and social care, or transportation-related courses had the lowest tendency to complete (however, some will still be in learning). We also found that the odds of completing a course were statistically lower for learners at colleges compared with those with non-college providers;
- **Area-related factors:** Some pilot areas have much higher completion rates than others, which seems to relate to their approach. For example, the pilot areas that concentrate most on initial assessment tend to have higher completion rates. The availability and extent of employer support also appear, at least anecdotally, to affect completion rates.

Employers and learners like their ETP experience

Employers and learners express high levels of satisfaction with their experience of the pilots. Ninety-three per cent of employers were either very or fairly satisfied. Similarly, 90% of learners were at least fairly satisfied. This is comparable to levels recorded in the LSC's national learner satisfaction survey.

Learners learn something new

Most learners think that they have learned something new that will help them do their immediate job better. Learners taking a health-related qualification were the most likely to say that they had learned a lot of new skills.

Learners see shorter-term benefits for their current job

As learners progress through their training, they tend to see fewer benefits per se, and those that they do see are more narrowly focused on their current jobs. In a second survey of the same ETP learners, after they had completed their courses, we found respondents were still very positive, but more likely than in the first survey to identify no benefits at all. In particular, they placed less emphasis on gains to do with having the skills to do a different job or with gaining a promotion or better pay.

Learners remain positive about the ETP and training in general once the training has finished, but tend to be less positive than when they started.

… but become more interested in further training

We have also found that learners are more inclined to go on to further study at the end of their course than at the beginning. The majority of those who wanted to study further were interested in a level 3 or higher qualification. Employees who become more positive during the course of their training are, not surprisingly, those most likely to want to progress to further learning. We found that learners who had received information, advice and guidance were also more likely to want to take their training further.

Employers benefit from better-motivated and skilled employees

All employers see a range of benefits from their participation in the pilots, the most important being giving employees more self-confidence and important business skills. As employers move through the programme, they appear to see the gains in more immediate terms and relate them to gains for their employees, rather than for the business as a whole.

On the other hand, employers become more positive in their attitudes to training following their ETP experience and say they are more likely to train less-skilled employees.

ETPs: a platform for progression

At the end of the second year, it is clear that the pilots have got off to a flying start, and that the offer is attractive to both employers and learners. Interestingly, some of the pilots that started more slowly in terms of getting in the numbers, and concentrated on offering a 'pure' approach, have been most successful in securing completions and therefore qualifications.

However, it is not just the offer itself that is attractive, it is the way it is being offered. The ETPs access learners through employers. Wage compensation does not seem as important as the provision of flexibly delivered, free or subsidised and, perhaps most importantly, brokered training in attracting employers. The broker plays a key role in taking the offer to employers and animating demand.

While the pilots have got the numbers in, there are still many more potentially eligible workplaces and learners to be accessed. Those yet to be involved will be harder to reach than those so far engaged.

Finally, we conclude that the pilots appear to offer a platform for progression:

- For learners to go on to level 3;
- For provision to become more flexible and workplace orientated;
- For employers to become more interested in training particularly for their lower-skilled employees;
- For policy to take the elements that work best and feed into the development of a national Employer Training Programme.

Reference

Hillage J, Loukas G, Newton B, Tamkin P. *Platform for Progression: Employer Training Pilots Year 2 Evaluation Report,* DfES Report ETP2.

The evaluation of Aimhigher: Excellence Challenge

Marian Morris, National Foundation for Educational Research

This paper focuses on the evaluation of Aimhigher: Excellence Challenge, a major government policy initiative that seeks to increase and widen participation in higher education among young people from disadvantaged areas. Launched in 2001 as the Excellence Challenge Programme, it was initially implemented in Phase 1 and 2 Excellence in Cities (EiC) areas and Education Action Zones (EAZs), focusing specifically on activities for young people aged 13–19. It has since been expanded to include all Phase 3 EiC areas and all Excellence Clusters. In August 2004, it was integrated with Partnerships for Progression (P4P) with the intention of providing a single coherent outreach programme. For the sake of consistency, Aimhigher: Excellence Challenge is used throughout this report, even though the initiative was known as Excellence Challenge when most of the research was carried out. (Quotes that included the title 'Excellence Challenge' have retained the original wording.)

Aimhigher: Excellence Challenge partnerships were formed primarily between schools, colleges and higher education institutions (HEIs), with support variously from Connexions, the local Learning and Skills Councils and local education authorities. Funds were delegated to local partnerships and HEIs to implement a range of locally agreed and targeted activities addressing perceived barriers to higher education. This dual strategy, of partnership directed activities combined with outreach activities initiated by HE institutions, presents specific challenges for the evaluation of the policy. Questions of attribution accompany any evaluation of a large-scale initiative: how does one identify the distinctive impact of Aimhigher: Excellence Challenge over and above other measures already in place to address disadvantage and increase motivation? In addition, the evaluation was given the task of identifying, as far as possible, the individual contribution of the separate funding strands, and to do so in a situation where there are many confounding factors.

This paper outlines the ways in which the evaluation is seeking to identify how successful Aimhigher: Excellence Challenge has been in meeting its aims. It highlights emerging findings about the activities that appear to have made a difference to young people's aspirations to higher education. It also provides a critical reflection on the methodology that has been adopted, and the tensions between the need to ensure a robust research strategy and the need to provide research evidence quickly in a fast-changing policy world.

Background

As Frost and Taylor (2001) outlined, most analysts have argued that there has been a 'qualitative shift' from 'elite, post-compulsory education towards a mass system', with total numbers in higher education in the UK increasing from around 300,000 in the 1960s to over 1.6 million by 1995. Indeed, the numbers entering higher education continued to increase throughout the 1990s, and total numbers in higher education are now in excess of 1.7 million (DfES 2004a). The Age Participation Index for young people under age 21 in Great Britain also rose from 19% in 1990/1 to 35% in 2001/02 (DfES 2004b). The Higher Education Initial Participation Rate (which replaces the previous Initial Entry Rate), was 44% in 2002/03. This indicates the proportion of 17

to 30-year-old England-domiciled first-time entrants to higher education, and the DfES's measure of progress toward its stated aim of 50% of young people having the opportunity to benefit from higher education by 2010.

Other significant changes have been noted, with a sharp increase in the numbers of part-time students (most of them mature students) and of female students. There has been an increase in overall participation among minority ethnic groups and, indeed, there is evidence of a proportionally higher rate of participation among young people from minority ethnic backgrounds than among young people from white backgrounds (Connor et al. 2004).

Despite these positive developments, many groups remain under-represented in higher education, including females from Bangladeshi backgrounds, young people with disabilities, those looked after by local authorities, and those who enter higher education via a vocational route. Most specifically, the strong negative link between socio-economic deprivation and participation remains evident. This was highlighted in Scotland by Johnston et al. (1999) and in England by the Higher Education Funding Council for England (HEFCE) (1997). While nearly 90% of young people from social classes I and II who have appropriate entry qualifications at age 21 achieve HE qualifications by age 30, the proportion of qualified young people from social classes III to V achieving such HE qualifications by the age of 30 is far lower. Here, the range is between 18% for social class III (manual) and V (unskilled) and 36% for social class III (non-manual) (Gilchrist et al. 2002).

Research carried out by UCAS (1999) found that, from 1994 to 1998, the increase in absolute numbers of potential students from social classes IV and V applying and being accepted to undergraduate higher education courses was:

> *… still extremely low compared with those from the social classes at the other end of the range. Although the proportions have changed over the five-year period, the initial numbers were so low that a great deal more effort is required to create a more equitable system which truly offers opportunities for all who are able to perform in and benefit from higher education.*

As Warwick (2001) argued, while:

> *… half the population is termed 'manual working class' … only a quarter of young entrants to higher education are from these classes.*

Indeed, while participation in higher education by social classes III (manual), IV and V increased from 11% in 1991 to 19% in 2001, participation among social classes I, II and III (non-manual) rose from 35% to 50% over the same period (DfES 2004c).

The challenge to widen participation in higher education continues, therefore. It is not unique to the UK – indeed, it has been recognised as a significant element in the promotion of lifelong learning across Europe. Of the four main strategies identified in member states (EEU 2000), two have been key to the most recent government policies for increasing participation in higher education among young people in England. Specifically, these have been:

- **increasing provision,** with additional places made available in further and higher education
- **extending access** to sectors of society that have previously been unrepresented.

The latter is the major premise behind Aimhigher: Excellence Challenge which seeks to improve access to higher education for young students from deprived areas and to reduce some of the current gaps in representation of different socio-economic groups.

The Aimhigher: Excellence Challenge initiative

Launched by the DfES in 2001, Aimhigher: Excellence Challenge was implemented with young people aged 13–19, in Phase 1 and 2 areas of Excellence in Cities (EiC) and in non-EiC Education Action Zones (EAZs). EiC was a key government policy initiative for redressing educational disadvantage and under-performance in schools in the most deprived urban areas. The government white paper The Future of Higher Education (DfES, 2003) heralded the policy's extension in existing Aimhigher: Excellence Challenge areas until 2006, and its expansion (again until 2006) to include all Phase 3 EiC areas and all Excellence Clusters. This expansion was accompanied by re-branding as Aimhigher. The Aimhigher logo was also to be used for all activities under Partnership for Progression, an initiative jointly funded, on a regional basis, by HEFCE and the Learning and Skills Council, to increase and widen participation in higher education (see Xavier and West 2003). This branding presaged the integration of Excellence Challenge and Partnerships for Progression into one coherent outreach programme from August 2004.

The integration of Aimhigher: Excellence Challenge and Partnerships for Progression was, in part, a recognition of the range of differently funded widening participation activities and partnerships already in operation across England, some of which predated Challenge or operated in areas outwith the policy. Recruitment of students from 'low participation neighbourhoods', for instance, had been included among in the performance indicators for higher education institutions since 2000, and many HEIs offered such widening participation programmes, sometimes in conjunction with local schools and colleges, in areas both within and outside the EiC programme or EAZs. In addition to HEFCE's existing special funding programme, several charitable bodies (such as the Sutton Trust, the Goldman Sachs Foundation and the John Lyons Charity) implemented their own widening participation projects. Moreover, extending access to further and higher education was an aspect of the wider social inclusion agenda that has prompted the focus on more flexible curriculum provision and enhanced support strategies for young people aged 14–19.

The existence of such a range of projects with similar or overlapping aims, as well as changes within Aimhigher: Excellence Challenge itself, have raised specific challenges for the national evaluation of the initiative and the extent to which it can address a key ministerial question: what works for whom and in what circumstances? Indeed, it is this question – albeit one that has become more pressing during the course of the evaluation than it had been at the outset – that has posed one of the most significant challenges to the evaluation team. However, before examining the research design adopted for the national evaluation of Aimhigher: Excellence Challenge, it is worth summarising the widening participation strategies that have been put in place.

The Aimhigher: Excellence Challenge strategies

It is widely acknowledged that many issues affect the decision to enter higher education. Research by Connor et al. (2001), for example, suggested that students from 'lower social class groups' – those whose parents were involved in manual, partly skilled and unskilled occupations – tended to take account of a wider range of issues than their 'higher social class counterparts' when making decisions about entry into higher education. Not least of these were issues related to current and potential employment and financial concerns. These factors were often seen by students as more discouraging than the lack of family experience of higher education.

In seeking to address these factors, Aimhigher: Excellence Challenge adopted a number of strategies, in which student-centred interventions – such as, for instance, summer schools, residential visits and Masterclasses for young people either designated as gifted and talented or as part of a widening participation cohort (DfES 2004d) – were combined with school/college-centred interventions (including partnerships between HEIs and schools and colleges). These activities provided young people with information and the academic and personal support they were thought to need to improve their performance and raise their aspirations. The strategies were based on a perception of 'what appeared to work' in previous widening participation and outreach work conducted by universities and other bodies.

In addition, a marketing drive and a range of marketing materials were introduced by DfES under the Aimhigher brand, both to raise awareness and to provide information. This campaign included the Aimhigher Roadshow and radio, television and magazine advertisements. Finally, to address some of the perceived financial barriers to participation, funds were made available to provide up to 26,000 Opportunity Bursaries over the period 2001/02 to 2003/04 (see DfES 2004e). These were open to young people obtaining places at funded HEIs (or at further education colleges offering higher education) who met the identified postcode and income criteria, and who came from families with little or no history of higher education.

While this was the broad picture of provision, the implementation of Aimhigher: Excellence Challenge at a local level varied with regard to the strategies used, the context in which they operated and the anticipated outcomes. Aimhigher: Excellence Challenge was broadly targeted at groups of young people who were under-represented in higher education, including those where there was a lack of family tradition of higher education, certain minority ethnic groups and young people with disabilities. However, the extent to which each of these groups was targeted at a local level varied between partnerships. Moreover, there was a fundamental difference between the prior experiences of young people in schools in EiC Phase 1 and 2 areas and EAZ areas. In the EiC areas, a range of centrally identified aspiration-raising activities and academic and personal support strategies had been funded for at least a year, if not longer, prior to the introduction of Aimhigher: Excellence Challenge; whereas, in EAZ areas, local strategies may have been in place to tackle such issues for some time, but with more limited funding and varying degrees of success.

Furthermore, at the outset there was only limited accord as to the exact definition of different categories of young people: there was little agreement, for example, among policy-makers and Partnership co-ordinators about what was meant by terms such as 'under-achievement'. This was also true for the range of local outcomes anticipated from the initiative, with significant variations between partnerships in terms of, for instance, the targets they set with respect to A-

level scores. How, therefore, have the evaluation team sought to take account of these differences?

The evaluation components

The research design developed for the evaluation was based on an integrated, mixed methodology strategy linked to the ongoing national evaluation of Excellence in Cities. Schools (and cohorts of pupils) had been engaged in the latter evaluation for a year prior to the launch of Aimhigher: Excellence Challenge. To reduce the potential burden on schools and to make cost-effective use of the existing evaluation strategy, it was decided, therefore, to augment existing EiC questionnaires (for pupils, teachers and senior managers) with appropriate items reflecting the range of expected activities and the anticipated impact of the Aimhigher: Excellence Challenge programme on attitudes to and aspirations towards higher education. These questionnaires were adapted for use in EAZ areas, omitting items that were solely related to EiC. Additional questionnaires were designed for use with Aimhigher: Excellence Challenge co-ordinators in FE colleges; teaching staff in school sixth forms and FE colleges (not part of the EiC initiative), HEIs, and Opportunity Bursary recipients and non-recipients.

The data from these quantitative research instruments was matched to the National Pupil Dataset and/or other databases, as appropriate, prior to use in a range of hierarchical statistical analyses. Alongside these, a range of qualitative research strategies were adopted, including annual interviews with Aimhigher: Excellence Challenge co-ordinators in EiC and EAZ areas, and ten area studies in which in-depth interviews explored the views of local actors and stakeholders (including young people) on the implementation, operation and impact of Aimhigher: Excellence Challenge. These interview programmes were augmented by analysis of documentary evidence obtained from partnerships' bids for funding and annual reports to the DfES, and local evaluation materials.

The strategies were designed, in part, to identify the changes to inputs and processes made in the course of establishing and augmenting links between, for example, schools, colleges and HEIs, and in developing outreach activities; and in part, to explore the apparent (and measurable) impact of the initiative on young people's aspirations, expectations, motivation, attainment and progression.

The wide scope of the data facilitates both cross-sectional analyses (comparing Year 11 cohorts in 2004, in 2003 and in 2002, for example); and longitudinal analyses (for example, following Year 9 pupils from 2002 through to Year 11 in 2004).

The data collected included background, attitudinal and performance data on (and from) around 10,000 young people from each year group for each of the three years of the survey, including data from a smaller but broadly matched group of young people from comparison schools. Input and process data was collected from teachers, tutors and senior managers from over 300 institutions.

The scale of the exercise enables a range of sophisticated statistical analyses to be undertaken, including multilevel modelling and logistic modelling. Among other things, these allow the research team to:

- Control for more observable differences between pupils than has been possible in previous research in this field;

- Isolate the apparent impact of different activities, taking account of background characteristics at pupil, school and regional levels;
- Control for unobservable differences that are constant over time.

The concurrent programme of qualitative data collection and analyses, and the further surveys of HEIs and Opportunity Bursary recipients and non-recipients, enable the findings from the statistical analyses of pupil and student data to be examined in context, while the survey findings have also informed the qualitative work. In this way, deeper understandings of the implementation, operation and impact of Aimhigher: Excellence Challenge have been developed than would have been gained from either the quantitative or qualitative programme alone.

However, while the evaluation has yielded many insights into the impact of Aimhigher: Excellence Challenge, the focus on input, process and outcome is, arguably, not quite the same as a focus on 'what works for whom and in what circumstances'.

The evaluation: a critical reflection

The concept of evaluation as a means of identifying 'what works for whom and in what circumstances' arises out of the paradigm of realistic evaluation, introduced by Pawson and Tilley (1997), and emerging from the realist tradition in the philosophy of science. This approach to evaluation has, as its basis, the proposition that 'causal outcomes follow from mechanisms acting in context'. In the context of Aimhigher: Excellence Challenge, this assumes, perforce, that an intervention (such as a residential visit to a university) would have the capacity to generate a specific outcome (such as a stated intention to go to university) when implemented with a particular targeted individual or group (such as a young person with no family history of higher education, or those from particular socio-economic backgrounds) in the appropriate circumstances.

However, realistic evaluation also presupposes that something is already known about the ways in which specific interventions (or mechanisms) might lead to different outcomes in different contexts. There was, at the outset, no singular theory about how each activity within the Aimhigher: Excellence Challenge programme would work to produce the desired outcome. There was, of course, an awareness of a lack of cultural capital among young people from working-class backgrounds, such that making a decision about further or higher education would be more akin to a 'chancy, uncertain process' than the 'natural, orderly, clear-cut' presentation of choice that it was likely to be for a middle-class student (Maguire et al. 1999). There was an acknowledgement that, as summarised by Archer and Leathwood (2003):

> *… working class participants and non-participants engage emotionally with negotiations around HE participation, which are grounded within genderised, racialized, classed and sexual identities.*

There was also an informed belief that young people from lower socio-economic backgrounds lacked information about higher education opportunities; that they might feel that the returns from higher education would be insufficient or too risky to be worth deferring current earning potential; and that they might perceive higher education as a threat to their identity ('university is not for the likes of us').

At the start, no specific link was made between these theoretical and conceptual frameworks and the activities and strategies implemented in local Aimhigher: Excellence Challenge partnerships. The wider policy was informed by UK and US research on factors that inhibited or encouraged participation in higher education among lower social classes (see Connor et al. 2001). It built on the suggestions that these studies made about strategies that might be necessary. These included, for example, communicating the benefits of higher education study; higher education champions working with young people from families with no history of higher education; and timely and relevant information on student finance and financial assistance for those in the greatest need.

The activities that were subsequently introduced were designed to:

- Raise awareness (through providing information via strategies such as the Aimhigher Roadshow);
- Raise aspirations (through experiences such as summer schools or other residential visits);
- Increase attainment (through supporting young people's academic work, in study skills groups or homework clubs, or through challenging them intellectually, for example, through Masterclasses);
- Address financial issues (through Opportunity Bursaries).

However, the introduction of such activities at a local level was based less on knowledge of the impact that particular activities had previously had, than on an awareness of types of outreach activities used previously to address the issues of under-representation and widening participation. In other words, the theoretical base from which hypotheses might arise as to what might work, for whom and in what circumstances, was not developed prior to the implementation of the initiative. The initial research design and data collection were not related to the existence of an 'exact form of hypothesis' (Pawson and Tilley 1997) arising specifically out of the Aimhigher: Excellence Challenge programme, but out of hypotheses related to an understanding of the barriers faced by young people and wider educational theory.

While the evaluation included a range of project-specific qualitative and quantitative data collection exercises, it was commissioned, as indicated above, to run in conjunction with the national evaluation of Excellence in Cities. The latter evaluation sought to identify the impact of each of the seven strands of EiC, and of EiC as a whole. It was not based on a tradition of realistic evaluation, even though it could be argued that there was a move towards this paradigm during the course of the evaluation. Certainly, in identifying 'what worked' in EiC, the evaluators sought to go beyond the identification of successful (or unsuccessful) strategies in order to identify with whom the interventions worked and, as far as possible, in what circumstances. The initial data collection for the EiC evaluation was informed by some theoretical understanding of young people's self-concepts, learning preferences and decision-making strategies, and by an understanding of the management of change. However, it was not informed by any educational theories related to the anticipated impact of, for instance, individually targeted interventions (such as learning mentors or the Gifted and Talented strand) or school-centred interventions (such as city learning centres or specialist schools) on specific groups of young people in particular learning contexts.

It could be argued, therefore, that some of the hypotheses now under investigation in the evaluation of EiC have emerged from the data, rather than informing its initial collection. This is arguably the situation for Aimhigher: Excellence Challenge. Moreover, the statistical techniques (which include multilevel modelling and propensity score matching) being used to evaluate the programme indicate significant associations rather than causation. The qualitative data being collected for the study enables some deeper insights about both the context in which Aimhigher: Excellence Challenge is operating and the attitudinal and attainment outcomes identified through statistical analysis. However, while annual interviews have been conducted with around half of the Aimhigher: Excellence Challenge coordinators, the collection of triangulated data from providers and participants is largely confined to selected case study areas. It could not be claimed, therefore, that the evaluation strategy facilitates collection of detailed data to allow the identification of 'mechanisms acting in context' in all possible local circumstances.

This final point is important. Realistic evaluation assumes a teaching and learning relationship between researchers and policy-makers, practitioners and participants. According to Pawson and Tilley (1997):

> *… the research act thus involves 'learning' the stakeholder's theories, formalizing them, 'teaching' them back to the informant, who is then in a position to comment upon, clarify and further refine the key ideas.*

This process, they continue, needs to be repeated over many evaluations, in order to feed into the development of an understanding about the relationship between intervention and outcomes. Although the evaluation of Aimhigher: Excellence Challenge was subsequently designated as one of its programme strands, this was not the case when the evaluation was commissioned, and did not overtly inform the design of the research. Moreover, while the research team provide feedback to schools and partnerships, and have conducted policy seminars for colleagues at DfES and disseminated project findings at workshops and conferences for Aimhigher: Excellence Challenge Partnerships, it would be disingenuous to claim that this constituted the teaching and learning relationship envisaged in realistic evaluation.

This may suggest, therefore, that the extent to which the evaluation can categorically identify 'what works for whom and in what circumstances' may be limited at this stage – at least as defined in terms of realistic evaluation. This is not to suggest that the evaluation lacks rigour or value or is not meeting its aims. Indeed, to date, it has contributed significantly to an understanding of 'what works' (in terms of the associations noted between activities and aspirations), as well as to an understanding of the types of young people with whom it works or has had little effect, and of some of the contexts in which interventions may be successful. These insights should contribute to the development of a better theoretical understanding of the relationship between interventions aimed at widening participation in higher education and improvements in attainment, motivation and aspiration. They should also contribute to the 'wider cycle of "enlightenment" between the research and policy fields' advocated by Pawson and Tilley.

Findings

What has the evaluation found to date? With whom has the policy had the biggest impact? And which activities appear to have contributed most to this impact?

- In 2003, 40% of the HEIs taking part in the study reported an increase in applications to university from young people from state schools since 1999;
- There were noted increases in applications from areas with low rates of participation (36%); from lower income groups (34%); from young people with no family background of higher education (34%); from disabled students (33%); and from young people from minority ethnic groups (30%) (Pennell et al. 2004; see also HESA 2004);
- Respondents were cautious in attributing these increases directly to widening participation activities – fewer than half (43%) were confident that this was the case;
- However, they acknowledged the role played by Aimhigher: Excellence Challenge in engendering and enhancing links with schools and FE colleges, and in expanding the number of widening participation activities in which they took part.

Despite the differences in context and implementation of Aimhigher: Excellence Challenge, the research suggests that there is a degree of overlap between the activities perceived to be motivating by the various players and participants and those that appear to be statistically associated with raising aspirations to higher education. HE staff and partnership co-ordinators tend to focus on the apparent part played by specific activities (particularly summer schools, residential experiences and visits to universities). However, analysis of student and pupil surveys suggests that the biggest impact may result from interaction with individuals (whether lecturers, undergraduates, teachers or other students) in HE-related contexts, and appropriate and timely information, advice and guidance.

Aimhigher: Excellence Challenge – what works?

In assessing any statistical association between activities and outcomes, the NFER evaluation team have used multilevel modelling and logistic regression techniques. The models facilitated the identification of associations between a specific input (such as summer schools) and outcome (such as an intention to enter higher education), while controlling for background variables at pupil and school level. Subsequently, the logistic analyses enabled the research team to explore the apparent impact of each activity on, for instance, the likelihood of particular individuals expressing a positive attitude to HE or an intention to enter HE.

Some common messages have already emerged through the various studies carried out with different cohorts of students. Once background characteristics (such as gender, ethnicity, prior attainment, home circumstances and parental education) are taken into account, the main predictors of taking up a university place among the older cohorts (who would have experienced Aimhigher: Excellence Challenge-funded activities only in post-16 education) appear to be a positive attitude to HE and a lack of disaffection with school or education pre-16. When young people's experiences in post-compulsory education are included in the statistical models, it

became clear that some of the educational interventions they had encountered during Years 12 and 13 (or equivalent) appeared to have contributed to their decisions to go to university. More specifically, these included:

- Talking to teachers or lecturers, higher education staff, or family and friends, about higher education (which increased the likelihood of a positive decision by exponentials of 2.48, 1.81 and 2.23, respectively);
- Undertaking a post-16 school or college course that they felt had provided helpful information about higher education (which increased the likelihood of a positive decision by an exponential of 2.11).

These inputs reflect some of the awareness-raising activities that have been implemented under Aimhigher: Excellence Challenge, but also reflect previous widening participation activities. What evidence is there of any specific Aimhigher: Excellence Challenge impact?

To date, it would appear that Aimhigher: Excellence Challenge may be promoting more positive attitudes to higher education. For some young people, it may have contributed to a decision to follow a university course, as the following examples from the 2002 Year 12 cohort illustrate (see Morris and Rutt 2003). Drawing on data from self-completion questionnaires administered first in 2001, when the young people were in Year 11 (and prior to the full implementation of Aimhigher: Excellence Challenge), and then in 2002, when they were in Year 12 (or equivalent), the probability of each young person expressing an intention to go on to higher education was calculated using logistic regression. This is a statistical technique similar to multiple regression analysis, but used where the outcome may be one of two possible outcomes, in this case the young person's expressed intention to go to university or not to go to university. The model's level of accuracy was 80%, in terms of prediction compared with observation. In total, 52% of all the young people expressed an intention of entering higher education.

The profile of traditional entrants to higher education is: non-disaffected, white high achievers studying at level 3 post-16, living in owner-occupied housing, with no money worries and at least one parent who had been through higher education. For young people who matched this profile, the base probability of expressing an intention to go to university was 35%, and 54% for those from minority ethnic backgrounds. However, where young people demonstrated a positive attitude to higher education, this probability rose to 63%, and to 79% among those from minority ethnic backgrounds.

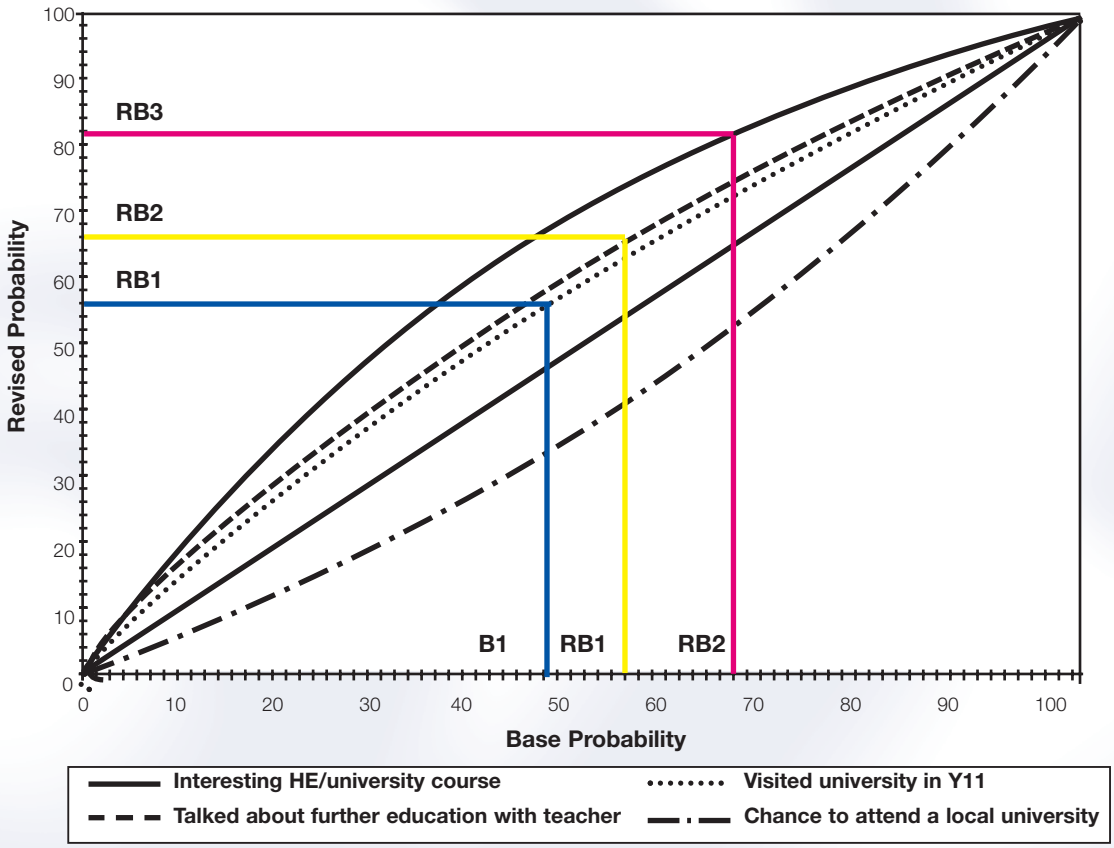
A similar pattern of increasing probability was observed for young people from non-traditional backgrounds. Compared with the traditional HE entrants in the survey, their base probability of aspiring to higher education was very low. High achievers studying at level 3 from families with no history of HE, who were living in rented accommodation and who expressed financial concerns, had only a 16% probability of indicating an aspiration to higher education (29% among students from minority ethnic group). However, probabilities rose to 38% (58% for students from minority ethnic group) among those with a positive attitude.

This suggests that young people's attitudes to higher education can have a very significant impact on the probability that they will aspire to higher education. To what extent can schools and other institutions affect these attitudes? Do awareness-raising and aspiration-raising exercises have any real impact on young people's attitudes to higher education?

To explore these questions, the research team used logistic modelling to examine the relationships between inputs and outcomes for various groups of young people. The first of these (see Figure 1) uses the example of a young person with the profile of a traditional entrant to HE:

- The probability of this student having a positive attitude to higher education is 49% (see B1 on the horizontal axis).
- A one-to-one discussion about further education with a teacher in the school in Year 11 enhanced this base probability to 55% (see RB1 on the *vertical* axis).
- The new probability of expressing a positive attitude is now 55%, shown as RB1 on the *horizontal* axis.
- Subsequently, a visit to a university increased the probability of a positive attitude to 65% (RB2).
- Finally, finding information about an interesting HE course increased this probability to over 80% (RB3).

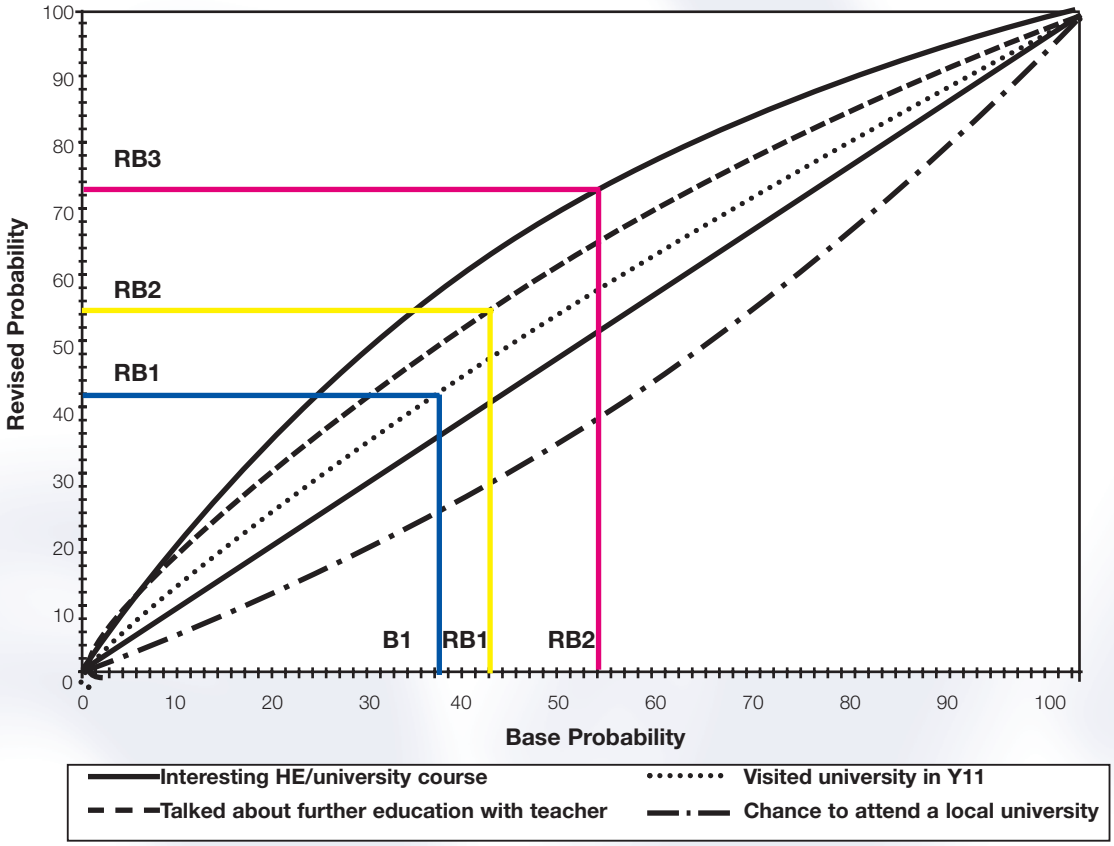
Figure 1. Probability of having a positive attitude to higher education: parent(s) completed higher education



Note: The estimate was obtained using logistic regression techniques. Note that these are estimates of probability and will have Standard Errors. The true probability may lie a few percentage points either side of those quoted.

Figure 2 illustrates the possible changes for a young person with the same profile as the typical entrant, but without a parent who had completed a higher education course. Here the base case probability was lower, at 37% (B1). Discussions with a teacher about further education raised this to 44% (RB1), while a visit to a university increased it to 55% (RB2). Finally, the possibility of pursuing an interesting course raised the probability of a positive attitude to HE to 73% (RB3). While this is still lower than that for the young person with a family history of higher education, it is significantly higher than it would have been without the interventions made by the school. Similar increases were identified for different groups of young people, including those with financial concerns and a level of disaffection with school pre-16.

Figure 2. Probability of having a positive attitude to higher education: no parental history of higher education



While it could not be claimed that all such activities occurred as a direct result of Aimhigher: Excellence Challenge, they are nonetheless the type of activities put in place at a local level. It is of particular interest that positive attitudes to higher education, post-16, have been influenced by pre-16 activities. For the Year 12 cohort in 2003, who had the opportunity for greater exposure to Aimhigher: Excellence Challenge activities pre- and post-16, the story is similar. As in the previous analysis, the apparent importance of pre-16 interventions in raising aspirations and motivating young people continued to be evident. These interventions included:

- Discussions with teachers about higher education (26% of the 2001 Year 11 cohort and 28% of the 2002 Year 11 cohort recorded such discussions);

- The opportunity to visit universities or HEIs (taken up by 20% of the 2001 Year 11 cohort and 21% of the 2002 Year 11 cohort);
- Lessons on transition skills, such as writing CVs and preparing job or course applications (available to 79% of the 2001 Year 11 cohort and 75% of the 2002 Year 11 cohort).

Additional factors that emerged in 2003, however, suggested the growing importance and influence of other activities that reflected the Aimhigher: Excellence Challenge strategies. In particular, discussions with undergraduate mentors (23% of the 2001 and 2002 Year 11 cohorts reported such discussions) and participation in summer schools (taken up by 12% of the Year 11 cohort in 2001 and 9% in 2002) were associated with an increased likelihood of stating an intention to enter higher education. Connexions Personal Advisers appeared to be significant both in helping young people develop a positive attitude to HE, and in promoting a desire to take up further study. One-quarter of the young people in each cohort said they had spoken to a Personal Adviser during Year 11, and around two-thirds of these indicated that such discussions had been useful.

For younger pupils (those in Year 9), a supportive climate in school, in which the value of going to university was discussed, and where progression to higher education was seen as normal among their peers, increased the probability that they would think about a university career by just under 5%. It was further increased (by just under 4%) where they had the opportunity to talk to undergraduates about life at university, and by nearly 30% where they had been encouraged to think that university was within their grasp.

It should be noted, however, that the young people in the survey who aspired to university were less likely than their peers to subscribe to a functional view of education, instead expressing a belief in study for its own sake. Those aspiring to university clearly appreciated the development of useful skills and knowledge during their time at school (this increased the likelihood of a positive aspiration by 3%), but appeared less concerned about its role in preparing them for adult and working life

These findings would suggest that some of the early Aimhigher: Excellence Challenge aspiration-raising and awareness-raising activities (particularly contact with university students) may have begun to have an impact, although the results were by no means universal. To what extent is there any evidence that these activities work with all young people in all circumstances?

Certainly, some differences emerged between EiC and EAZ areas in terms of the impact of the Aimhigher: Excellence Challenge programme. For instance, among post-16 EiC cohorts, the level of parental education was not a significant factor in promoting positive attitudes to HE. This is not to say that it was not important: the probability of a young person saying that they intended to go to university was greater among the young people who had at least one parent educated to degree level (over one-fifth of the cohort) than among those whose parents were not educated to this level. Rather, the range of strategies put in place by schools and colleges may be able to assist young people in developing such positive attitudes, whether or not their parents have been to university. Among young people from EAZ areas, however, the level of parental education still emerged as an important factor associated with the probability of a positive attitude to higher education.

This may suggest that, in non-EiC EAZ areas in particular, more work may still need to be done to raise awareness of higher education among non-traditional entrants and, perhaps, among their parents. Certainly, levels of parental involvement and interest were significantly associated with an increased probability of indicating an intention to go to university among the pre-16 cohorts. Young people whose parents expressed an interest in their continuing education, attended parents evenings at school and encouraged them to talk about university, had significantly higher probabilities (by some 12%) of wishing to go to university than their peers without such support. Young people who perceived parental support as primarily related to encouraging them to behave well in school, however, were generally less positive about continuing into HE (this reduced their probability of wishing to go to university by some 2%).

Of significance for the future development of Aimhigher is the emergence of a small group of young people on whom the scheme's activities appear to have had little impact as yet, in terms of a stated intention to enter higher education. Although sharing the same background characteristics, as well qualified and studying to the same level (post-16) as their peers, they appeared to differ only in the extent to which they expressed financial concerns and satisfaction with their post-16 courses. They were no less confident than their peers in their ability to cope with a university course, nor were they any less sure of their ability to fit into a university setting. However, they were more likely to have indicated a lack of motivation on their current course, to lack confidence in their potential examination results, to lack conviction that they would enjoy studying for a degree, and to feel that most people who went to university ended up in debt.

These views were not based on a lack of awareness of university life. Indeed, this group were more likely than their peers to have visited a university (29% compared with 20%), or to have spoken to a teacher (35% compared with 27%) or a member of university staff (17% compared with 11%) about higher education. Moreover, their lack of interest in university did not appear to have any strong social foundation. They were less likely than their peers who hoped to enter HE to think that their friends would think them a snob if they went to university (2% compared with 6%); and they were no different from other students in terms of the level of parental interest and support that they identified (including support for higher education). However, a significantly higher proportion (81% compared with 72% of other students) suggested that not having to worry about getting into debt would motivate them to consider university. More than half also suggested that obtaining a place at a local university (which, arguably, would reduce the costs) would also encourage them to think about higher education.

Key points

These findings highlight two key points for the future development of Aimhigher. Firstly, it suggests that there may be a core of young people for whom awareness and aspiration-raising activities are unlikely to prove motivating, unless accompanied by some alleviation of their fears about incurring debt. There are indications that Opportunity Bursaries may have contributed to some young people's decision to follow a degree course. Students from the 2003 Year 13 cohort who had successfully applied for an Opportunity Bursary, for example, were more likely than other young people to be taking up a university place. It should be recognised, however, that such young people were unlikely to have applied for a bursary without having at least a clear intention to go to university in the first place (Morris and Rutt, forthcoming).

There are also indications that the bursaries may have reduced the fear of debt. For instance, in a survey of Opportunity Bursary recipients and non-recipients (West et al. 2003), more than eight out of ten students said they agreed with the statement, 'I was worried about getting into debt'. However, significantly more non-recipients than recipients agreed with this statement. More non-recipients also agreed with the statements: 'I knew that I would have to work while at university' and 'I was worried about combining studying with a job'. On the other hand, more Opportunity Bursary recipients than non-recipients agreed with the statement, 'I was confident that the long-term financial benefits would outweigh the costs of doing the course'.

Secondly, these findings highlight the importance of good pre-16 careers education, information, advice and guidance. Disillusion with post-16 courses was not linked to specific post-16 destinations: these views were expressed equally by young people in school sixth forms and by those who had left 11–16 or 11–18 schools to attend further education or sixth form colleges. However, strong associations were identified during the research between positive attitudes to both learning and higher education, and the provision of good-quality information, advice and guidance. It would seem appropriate, therefore, that Aimhigher should advocate the provision of appropriate institutionally based advice and guidance systems for young people, and continue to promote effective interactions between Aimhigher partnerships and Connexions Services. Connexions Personal Advisers appeared to be significant both in helping young people develop a positive attitude to higher education, and in promoting a desire to take up such further study. Yet, only one-quarter of the young people in the post-16 surveys said they had spoken to a Personal Adviser during Year 11. (Around two-thirds of these indicated that they had found such discussions useful.)

Conclusion

The issue of widening participation in higher education is at the forefront of the government's higher education agenda, and creates an immediate need for up-to-date and accurate information about the impact and value of policy-related educational interventions. However, the rate at which research evidence can be compiled is subject to time constraints, both data-led (whether study-specific or from national datasets) and related to the length of time it takes for interventions to be implemented and become part of standard practice in the field. This inevitably leads to tensions. Researchers are rightly concerned to ensure that the evidence presented is robust, reliable and valid. While policy-makers would share this view, they are often faced with the need to provide immediate insights into whether or not policies are 'working'.

Nor are policies static. Aimhigher: Excellence Challenge, for instance, has not only had its name changed (to Aimhigher), but has seen the addition of new strands (the Student Associates Scheme) and the removal of others (Opportunity Bursaries have been replaced by grants for young people from low-income families). It has been extended over time, expanded to incorporate areas previously outwith the policy and, most recently, has been integrated with Partnerships for Progression, leading to the development of existing and the formation of new relationships amongst former Aimhigher: Excellence Challenge and Partnerships for Progression planners and practitioners. While this latter change took place in August 2004, after the main phase of data collection for the evaluation, partnerships were nonetheless gearing up for change for many months prior to this date.

The challenges that these changes have placed on an evaluation initially designed to assess the effectiveness of the Aimhigher: Excellence Challenge policy are many. Since the evaluation began in 2002, in order to meet the requirements of the changing policy, the research team has had to revisit, adapt and/or redesign a number of its planned strategies, adapt its research instruments and implement a range of different and exploratory statistical techniques to address the question of 'what works'.

The evaluation has met with some success, however. To date, it has highlighted aspects of Aimhigher: Excellence Challenge that have had an impact – that 'work' insofar as they appear to have engendered a positive attitude towards HE that might, in turn, lead to taking up a place at a higher education institution. It has also identified groups of suitably qualified young people for whom the barriers to higher education (including a lack of aspiration or belief in its longer-term value) have not yet been removed. It has demonstrated the value of juxtaposing different research methodologies in an effort to find out 'what works'. While interviewees often mention the impact of particularly memorable activities, because of their setting, their dynamics or their difference from normal curriculum inputs, the statistical analysis suggests that the more long-term impact may derive from regular exposure to discussions about higher education, in a climate conducive to the development of a belief that studying for a degree might be within their grasp.

The next stage might be to test out the findings in local Aimhigher partnerships – to develop and refine hypotheses about the impact of particular strategies (on particular groups of young people and in particular circumstances) that could be tested in local contexts.

References

Archer L, Leathwood C (2003). Identities, inequalities and higher education. In: Archer L, Hutchings M, Ross A (2003). *Higher education and social class: issues of exclusion and inclusion.* RoutledgeFalmer.

Connor H, Dewson S with Tyers C, Eccles J, Regan J, Aston J (2001). *Social class and higher education: issues affecting decisions on participation by lower social class groups* (DfEE Research Report 267). DfEE.

Connor H, Tyers C, Modood T, Hillage J (2004). *Why the difference: a closer look at higher education minority ethnic students and graduates* (DfES Research Report 552). DfES.

DfES (2003). *The future of higher education* (Cm. 5735). The Stationery Office.

DfES (2004a). *4.7: Post 16 learning: higher education student population.* Accessed 14 October 2004 at: http://www.dfes.gov.uk/trends/upload/xls/4_7t.xls

DfES (2004b). *Welcome to trends in education and skills.* Accessed 14 October 2004 at: http://www.dfes.gov.uk/trends.

DfES (2004c). 4.6: *Participation in higher education trends.* Accessed 21 September, 2004 at: http://www.dfes.gov.uk/trends/index.cfm?fuseaction=home.showIndicator&cid=4&iid=23.

DfES (2004d). *Aimhigher … programme website.* Accessed 4 October 2004 at: http://www.dfes.gov.uk/aimhigherprogramme/index.cfm?i_pageId=1&s_pageType=level3.

DfES (2004e). *Aimhigher … programme website.* Accessed 14 October 2004 at: http://www.dfes.gov.uk/aimhigherprogramme/index.cfm?i_pageId=4&s_pageType=level3

EEU (2000). *Lifelong learning: the contribution of education systems in the member states of the European Union.* Eurydice European Unit.

Frost N, Taylor R (2001). Patterns of change in the university: the impact of 'lifelong learning' and the 'world of work'. *Studies in the Education of Adults,* 33(1) 49–59.

Gilchrist R, Phillips D, Ross A (2002). Participation and potential participation in UK higher education. In: Archer L, Hutchings M, Ross A (2002). *Higher education and social class.* RoutledgeFalmer.

HESA (2004). Data accessed on 14 October, 2004 at: http://www.hesa.ac.uk/pi/summary_0203.htm

HEFCE (1997). *The influence of neighbourhood type on participation in higher education.* Higher Education Funding Council for England.

Johnston V, Robb G, Abdalla I (1999). Participation in higher education in Scotland: a geographic and social analysis. *Higher Education Quarterly* 53(4) 369–94.

Maguire M, Ball AJ, Macrae S (1999). Promotion, persuasion and class-taste: marketing (in) the UK post-compulsory sector. *British Journal of Sociology of Education* 20(3) 291–308.

Morris M, Rutt S (2003). *Aspirations to higher education: a baseline analysis.* Accessed 15 October 2004 at: http://www.nfer.ac.uk/research/documents/EIC/11-2003.doc.

Morris M, Rutt S (forthcoming). *Aspirations to higher education: one year on.* NFER

Pawson R, Tilley M (1997). *Realistic evaluation.* Sage Publications.

Pennell H, West A, Hind A (2004). *Evaluation of Aimhigher: survey of higher education providers 2003.* DfES.

UCAS (1999). *Widening participation* (Statistical Bulletin). Universities and Colleges Admissions Service.

Warwick D (2001). *Integrating approaches to widening participation. Update on Inclusion 3*, 5–7.

West A, Hind A, Xavier R, Jupp J (2003). *Evaluation of Aimhigher: survey of opportunity bursary applicants 2001/02: Preliminary Findings* (DfES Research Report 497). DfES.

Xavier R, West A (2003). *Excellence challenge funding and widening participation.* Unpublished report.

The impact of research evidence on running a university

Peter Scott, Kingston University

The subject of this paper, in essence the relevance of higher education research to policy and practice, has always been contentious. For example, a major UNESCO conference on this theme was held at the University of Tokyo six years ago and, although there were some interesting contributions (subsequently published as a book), no resolution of the underlying tension was offered (Teichler and Sadlak 2000). The topic can be approached from two angles: that of a higher education manager, and that of a higher education researcher. For a higher education manager, the key questions are:

- What kind of research would help college or university managers to do their jobs more effectively?
- Does it exist?
- If does exist, is it packaged in ways that a busy manager can access easily?
- If it does not, how can this deficit in our research knowledge be remedied?

As a higher education researcher, with my colleagues at the University of Leeds, David Smith, Catherine Bargh and Jean Bocock, I carried out research into (among other things) the governance of universities and the changing role of vice chancellors and chief executives, with support from the Leverhulme Trust. Because at that time I had no intention of becoming a vice chancellor, my perspective was that of the pure researcher (not of the manager dabbling in research).

This paper is informed by these two perspectives, both of them practitioner perspectives.

It might have been expected that the views from these two different perspectives would lead to different conclusions – with the researchers insisting on the quality and relevance of their research (although accepting that larger and better-funded projects would be desirable) and the managers bemoaning its absence. But these two perspectives lead to rather similar conclusions. Both suggest that there is a limited amount of high-quality research that is relevant to the 'running' of universities under current conditions – even in key policy and practice domains such as access and participation (Osborne, Gallacher and Crossan 2004). In fact, it is possible to go further: there has been a tendency for the already limited high-quality research to decline. This is surprising, given the central place occupied by universities in the knowledge economy and the corresponding degree of turbulence (and change) they are experiencing. Nor is this dearth simply a question of the parlous state of higher education research in terms of the traditional academic apparatus of research council grants, well-established research centres, and cohorts of PhD and post-doctoral students. Much of the potentially most valuable research in higher education is applied – either because it has been commissioned by policy agencies, or because it is undertaken in-house (in the form of institutional research). But applied research in this field is also scattered and fragmented.

Scope of higher education research

The first part of this presentation considers the research that is most relevant to running a university (even if some – or many – higher education managers might be slow to recognise its value). Five major categories can be identified:

- Better conceptualisation;
- Independent policy research;
- Research on leadership, governance and management;
- Research into the university's core businesses;
- Institutional research.

Conceptualisations

Much better conceptualisations are needed of both the (multiple) contexts in which modern universities have to operate. This is much more than an assessment of the increasing number of their stakeholders, and of how systems, sectors and individual institutions respond to these new environments – their strategic directions. In other words, there needs to be a better understanding of how higher education institutions relate to wider formations such as 'knowledge society', 'audit society' or 'risk society', or to phenomena such as globalisation or marketisation. Put most simply and starkly, there is a need for more and better theory. Although not all vice chancellors would agree, better theory is needed because 'strategising' is now such a key part of university management. There is a real risk that, unless institutional leaders are better at engaging – creatively and intellectually – with the contexts and frameworks that are illuminated by theory, 'strategising' will become a reductionist, tick-box, toolkit kind of activity.

Independent policy research

High-quality, sophisticated and, above all, independent policy research is more than evaluation studies, however excellent, on the one hand, or oppositional/ideological critiques of policy, on the other – although both have their place. Instead, sophisticated analyses are needed of the interplay of politics, public opinion, bureaucracy and professional expertise in the making and development of higher education policy. In other words, more of a classical political science perspective is required. Too many higher education leaders display naiveté, sometimes tinged with arrogance and condescension, about policy-making. They tend to dismiss it as mere 'politics' – and, therefore, as crude, superficial and unsophisticated, when it is none of these things. This perspective is very important, of course, because higher education institutions are themselves political organisations. Policy-making is a key activity within colleges and universities – not simply an activity undertaken by the DfES, the Scottish Executive and the National Assembly of Wales and their attendant funding councils.

Leadership, governance and management

Research on leadership, governance and management (probably in reverse order) comprises the third category. Here it is important to see the big picture. Researchers and managers, in their different ways, have to comprehend whole institutional cultures that embrace this trinity of leadership, governance and management, and much more. So organisation theory should be a key strand within higher education research. This is important for several reasons:

- Universities are large and complex organisations with multiple missions. As a result, traditional conceptions of collegiality are coming under increasing challenge.
- There is no longer such a strong consensus about the core purposes of a university (and even core values).
- Many, perhaps most, institutions are suffering from 'mission overload' – or, a more accurate and neutral description, 'mission stretch'. Roles and activities that as recently as ten years ago would have been regarded as peripheral have become part of the core functions of higher education institutions.

The proper intellectual tools, which can be provided only by high-quality research, are needed to understand what kind of organisations universities have become.

Core business

Research into the core business of higher education encompasses:

- Learning and teaching (now, inevitably, with a strong dose of ICT-delivered learning);
- The evolution of disciplinary and professional cultures, including intellectual history and an exploration of professionalism, as well as the life stories and identities that constitute, and are constituted by, professionalism;
- The development of 'knowledge production' – the research engine that powers our modern world and runs fastest in universities – with its crucial links via politics and the media into contemporary consciousness.

Some may argue such research themes are not directly relevant to 'running a university'. But it is impossible to run a university without engaging with its core values and responsibilities. Even in the private sector, sensitivity to values is increasingly seen as a key input into successful management. Nor is this simply a matter of political contrivance, of apparent empathy with former academic colleagues, to ensure buy-in to institutional strategies. It is not even a matter of well-informed and better-designed strategising. It goes to the heart of what higher education is about and for. It also raises the all-important question of who is responsible for running universities – which are, by their very nature, highly distributed authority systems. It is, perhaps, a mistake to concentrate too much on senior managers to the exclusion of the wider academic community.

Institutional research

What is usually called 'institutional research' – systematic evaluation, 'internal' research and self-reflection – is necessary in any self-critical, or 'reflexive', institution. Perhaps universities are not especially reflexive institutions, however reflexive their teaching and research practices, and however many their 'reflective practitioners'. But they should be, especially in the much more open and less predictable environments in which they now find themselves. However, even if the special and historic responsibilities of universities for critical enquiry and rational discourse are put to one side – and even if their urgent need to understand and engage with these turbulent environments is discounted – it is difficult to deny that any organisation of the size and complexity of a modern university needs an 'institutional research' capability.

The available research

A review of the available research that is relevant to 'running a university' supports the assessment that there is too little. But the weakness of higher education research is part of a wider picture. There are two important deficits, not simply in higher education research, but in all social science research today, and perhaps beyond the social sciences: a 'theory gap' and an 'empiricism gap'.

The theory gap

There is a 'theory gap' or, to put it more starkly, a lack of research that engages the really big questions. This has always been a particular issue with regard to England (Scotland is less affected) because of the English attachment to pragmatism and a reluctance to think 'great thoughts' for fear of being labelled an intellectual. However, the 'theory gap' is more serious than this. One of the consequences of the attacks on critical social science has been that the conceptual maps it employs are out of date. Many were generated a generation or more ago, in the 'golden age' of post-war social science (and often in the United States). This is very obvious in higher education research, where linear models of a transition from elite to mass to universal higher education, under controlled conditions of deliberate stratification, are still very influential. To the extent that these conceptual maps have been displaced, it is by the journalistic products of think-tanks, on the one hand, or the mist of post-modernism, on the other.

The empiricism gap

The second significant deficit is an 'empiricism gap' – a lack of sustained, substantial and well-designed empirical studies that thoroughly test important research questions and, in the process, generate genuinely new data. Such studies never really got off the ground in the first place in higher education research, and they are increasingly rare for two reasons. They are expensive – perhaps too expensive – and they are too extended, too long-term. The latter is unwelcome to policy-makers and commissioners, who need quick results and have short attention-spans (inherent in policy-making under contemporary conditions). It is also unwelcome to researchers, who need to sustain a high 'run rate' to compete successfully in the Research Assessment

Exercise (RAE). As a result, there is much more short-order, 'fast-food' research – and a much greater reliance on other people's data. (There is, however, a limit to how much can be squeezed out of data from the Higher Education Statistics Agency or the Universities & Colleges Admissions Service).

These general deficits, which are part of a wider phenomenon in the research system, have had a particular impact on higher education research. Higher education research has never managed to acquire a strong enough theoretical base. Nor has it ever enjoyed a sufficiently robust infrastructure, in terms of chairs, research centres, research programmes, and so on. Of course, it is possible to see these weaknesses as potential strengths, because it can be argued that the rather limited institutionalisation of higher education research means that it can embrace new, more open and entrepreneurial research paradigms more easily than better-established fields struggling to preserve old, and ultimately unsustainable, research economies. This is not an altogether convincing argument. Is it good in the context of the next RAE for higher education research to be subordinated to educational research, which in turn is contained within a larger grouping dominated by psychology? And what about the obvious difficulties encountered in placing subject-specific higher education research within the RAE matrix?

The available research can be mapped against the desirable research, using the same five headings.

Better theory

In this area, important contributions have been made by Ron Barnett (Barnett 1990, 1994, 1997, 1999) and others in terms of developing a more systematic philosophy of higher education. This does establish interesting links to new paradigms of learning and to the evolving taxonomy of academic disciplines. But there has been very little research that attempts to place universities within larger scenarios of social and cultural change. Much of the literature on the economics of higher education – and the economic consequences of higher education for individuals, communities and nations – is a mix of (rather naive) econometric analysis and (occasionally rather politicised) advocacy. For example, theories of globalisation generated in very different contexts tend to be imported, without any serious critical adaptation, into the domain of higher education.

Policy studies

In the area that can be broadly labelled 'policy studies', there are too many examples of studies that are too 'immediate' and, therefore, likely to be ephemeral. For example, studies of the impact of successive RAEs on the concentration of research, especially on a regional basis, or of the likely costs and benefits of top-up fees and alternative funding systems, belong to particular political 'moments'. Valuable work has been done by the Higher Education Policy Institute, directed by Bahram Bekhradnia, and by 'Evidence', the Leeds-based consultancy run by Jonathan Adams – but both these examples are of policy analysis rather than policy research.

There is a lack of significant research on how policy-making systems in higher education are evolving, either inside universities or within the system as a whole. For example, the complex but

highly significant changes in the steering of the English higher education sector have gone largely unexplored. In both normative and organisational terms, the University Grants Committee, which kept ministers and their officials firmly at bay, was an entirely different kind of intermediary agency from the Higher Education Funding Council, which was prepared to engage in a joint white paper/Higher Education Bill implementation project with the DfES. How did we get from there to here?

There is also a serious lack of adequate policy research on the intended, unintended and unknowable consequences of policies. For example, it is not yet clear whether middle-ranked universities are more likely to curb growth when the yield-per-student improves as differential (and higher) tuition fees are introduced, or whether they will recruit up to the top of their contract ranges to insure against any RAE disappointment. This makes it difficult, perhaps impossible, to make accurate predictions about the future supply of student places. There is an equal dearth of reliable studies about the impact of higher fees on student demand. There is an urgent need for 'real-time' research into the possible effects of the introduction of a quasi-market in higher education.

Leadership, governance and management

There are more examples of high-quality research in this area, but even here there are problems if the measure is to be the impact of research evidence on running a university. Leadership studies (broadly defined) are beginning to have an impact. Through the Top Management Programme, these ideas have influenced the next generation of university leaders and are no longer confined within the esoteric (even cultish) world of international business schools. Leadership studies have, to some extent been domesticated in the world of higher education – although through the medium of case studies, simulations and action research, rather than through more rigorous academic research. In the wider context of organisational culture, there is a dearth of good higher education-related work. There is an extensive, and interesting, literature on leadership of which Robin Middlehurst's work is the best example (Middlehurst 1993) – and also, from a more critical perspective, on 'managerialism' from researchers such as Rosemary Deem (Deem 2001). Much has also been written about the 'new public management', although in terms of things being done to universities, rather than higher education itself developing new managerial practices and organisational cultures. However, too much of this research is still overly dependent on concepts and case studies from outside higher education.

Core business

Here, there is sometimes a worrying sense of disengagement between research and practitioner communities and the 'management class'. There is a substantial volume of research into learning and teaching in higher education. Indeed, the majority of higher education research is in this area, rather than the area of theoretical or policy studies. Much of it is practitioner led, which is an asset, but it also suffers from a number of weaknesses. One is an occasionally protective (even exclusionary) attitude on the part of researcher-practitioners; another is the comparative lack of interest in such work among many senior managers. There is also a very impressive literature on what is sometimes, far too narrowly, described as 'science studies'. Stephen Fuller's work at

Warwick is a good UK example, although a lot of the best work has been done in the United States (Fuller 1999, 2000). Once again, however, the connections between this impressive body of work and the development of research strategies in individual British universities, or the recent review of the RAE, are tenuous in the extreme. Finally, of course, detailed changes in learning and teaching and in the organisation of research coalesce into the wider evolution of academic and scientific culture. Such changes should engage the attention of university leaders, even if they are bored by the details, because this is directly relevant to the strategic opportunities facing their institutions, individually and collectively.

Institutional research

An impressive amount of institutional research is being undertaken, even if the label is less familiar in Britain than in the United States, or even the rest of Europe. This is confirmed by a quick scan of relevant journals. In addition, it is important to remember that institutional research is not likely to be published in conventional formats. There are also many issues. One is striking the right balance between academic esteem and institutional status. Some institutional research is produced in what are effectively academic support departments, rather than academic departments, and much of the rest is undertaken by researchers in other disciplines. There may be a tension between research that is highly rated by peers and research that is most relevant to institutional needs. Another issue is the effectiveness of the feedback loop. It is often difficult to generalise from specific institutional research projects, and, where findings are more generalisable, it is sometimes at the expense of their immediate relevance. Moreover, many institutional researchers – in university planning units, for example – do not really see themselves as researchers. A fourth issue is how to translate and generalise the findings of institutional research so that they can have an impact beyond the institutions in which they were generated.

Conclusion

It is difficult to feel optimistic about the current state of higher education research, whether in toto or in the context of 'running' a university. The blame must be shared equally between the research community and higher education leaders. The primary goal of the former is to enhance its academic standing, for which it can hardly be blamed, given the insistent (and increasing) pressures of the RAE, because under-performance is severely punished. The latter do not yet have the habit of looking to high-quality research for help in improving institutional performance. The overall conclusion must be that more and better higher education research is badly needed.

Here, higher education leaders have a special responsibility – to protect and enhance the (inadequate) research capacity that already exists in the system; and to be knowledgeable about, and to use, research findings. Two obstacles must be overcome to achieve this. Research into higher education is often stereotyped as a 'soft', and therefore low-prestige, field of enquiry. It is seen as a sub-set of educational research in which, according to the Economic and Social Research Council, there is a general lack of research capacity (and educational research, of course, is one of the least respected among the disrespected social sciences). That has to change. The second obstacle is that too few higher education leaders regard running a university as an intellectual challenge as well as a managerial one.

But higher education researchers must also play their part by raising their aspirations – by grasping grand(er) theory, and being active contributors to the development of meta-discourses of globalisation, risk and knowledge society; by penetrating beneath leadership, governance and management issues into the substratum of organisational culture; and by tackling issues of intellectual and scientific change. These higher aspirations are just as important in the context of institutional research as among academic researchers in RAE top-rated departments. Paradoxically, it is by refusing to be simply 'short-order' researchers and consultants and by raising their intellectual sights that higher education researchers can make themselves more relevant to policy-makers and ensure that their research has a greater impact on 'running a university'.

References

Barnett R (1990) *The Idea of Higher Education,* Buckingham: Open University Press.

Barnett R (1994) *The Limits of Competence: Knowledge, Higher Education and Society,* Buckingham: Open University Press.

Barnett R (1997) *Higher Education: A Critical Business,* Buckingham: Open University Press.

Barnett R (1999) *Realizing the University in an Age of Supercomplexity,* Buckingham: Open University Press.

Deem R (2001) *'Globalisation, New Managerialism, Academic Capitalism and Entrepreneurialism in Universities: Is the Local Dimension Still Important*, Compartive Education 37 (1), pp 7-20.

Fuller S (1999) *'The Re-Enchantment of Science: A Fit End to the Science Wars?'*, Batty Memorial Lecture, McGill University, Montreal.

Fuller S (2000) *The Governance of Science: Ideology and the Future of the Open Society,* Buckingham: Open University Press.

Middlehurst R (1993) *Leading Academics,* Buckingham: Open University Press.

Osborne M, J Gallacher and B Crossan (2004) *Researching Widening Access to Lifelong Learning: Issues and Approaches in International Research,* London: Rouledge Falmer.

Teichler U and J Sadlak (2000) *Higher Education Research: Its Relationship to Policy and Practice,* Oxford: Pergamon.

Success for All in England:

Implementation and outcomes of a comprehensive literacy reform for primary schools

Robert E Slavin, Johns Hopkins University, and
Judith Wordsworth, Success for All Foundation, UK

Success for All (SFA) is a comprehensive programme for reading instruction designed primarily for schools serving many children who are in poverty. The intention of Success for All is to use strategies known from research to contribute to children's learning, to attempt to ensure that children begin with success in the earliest school years and then build on this success throughout their time in primary school.

The programme was initially developed in the US, at Johns Hopkins University. Since first SFA school was established in 1987, the programme has gradually expanded to about 1,500 schools in 48 of the 50 states, and has been adapted for use in Canada, Mexico, Australia, and Israel.

In 1997, David Hopkins, then at the University of Nottingham, introduced Success for All to a small group of schools in Nottingham, and led a small study of the programme's effects on student reading achievement. The model has expanded to a total (in 2002–03) of 27 schools, located in London, Nottingham, Hull, Derbyshire, Essex, and Leeds.

The importance of Success for All in the English context is in providing an alternative reform model for schools that serve many children in poverty. The National Literacy Strategy (NLS), which began in 1998, has brought about substantial improvements in the reading performance of English pupils; but in recent years the gains have levelled off, as measured by Key Stage 2 assessments (Earl et al. 2003). Many schools in high-poverty areas remain far below the norm in reading. The adaptation of SFA in England is completely aligned with the NLS, but provides a great deal of professional development, research-based materials, and other supports to help schools accelerate their progress toward national norms. It can be thought of as an intensification of NLS, not an alternative to it, for schools in high-poverty areas that are struggling to reach national standards.

This paper describes Success for All, briefly summarises the US and international research on its outcomes, and presents research on Success for All in England.

Programme components

Although the materials used in England have been extensively revised to adapt to NLS standards, as well as to the language and cultural context of England, the components are essentially the same as they are in the US. These are described in the following sections.

Reception

The current version of the SFA programme for reception is called KinderCorner. This programme focuses on providing a balanced and developmentally appropriate learning experience for young children. The curriculum emphasises the development and use of language. It provides a balance of academic readiness, particularly phonemic and alphabet awareness. Readiness activities include a programme called Story Telling and Retelling (STaR), in which students retell stories read by the teachers. More formal pre-reading activities begin during the second term of reception, incorporating a beginning reading programme called KinderRoots, described in the following section.

Beginning reading

Success for All uses a reading curriculum based on research and effective practices in beginning reading (e.g. Adams 1990; National Reading Panel 2000), and on effective use of co-operative learning (Slavin 1995; Stevens et al. 1987).

> Reading teachers at every grade level begin the reading time by reading children's literature to students, and engaging them in a discussion to enhance their understanding of the story, listening and speaking vocabulary, and knowledge of story structure. In Reception and Year 1, there is also a strong emphasis on phonemic awareness activities that help develop auditory discrimination and support the development of reading readiness strategies. Students are taught alphabet and sound blending in a programme called *FastTrack Phonics.*

KinderRoots is typically introduced in the second semester of the Reception year. In Year 1, this beginning reading programme is called *Reading Roots*. It uses as its base a series of phonetically regular but meaningful and interesting minibooks, and emphasises repeated oral reading to partners as well as to the teacher. The minibooks begin with a set of 'shared stories', in which part of a story is written in small type (to be read by the teacher) and part in large type (to be read by the students). The student portion uses a phonetically controlled vocabulary. Taken together, the teacher and student portions create interesting, worthwhile stories. Over time, the teacher portion diminishes and the student portion lengthens, until students are reading the entire book. This scaffolding allows students to read interesting literature when they have only a few letter sounds.

Success For All uses a synthetic approach to the teaching of phonics, blending and segmenting. These skills are introduced in an active, engaging set of activities beginning with oral language, and moving into written symbols. Individual sounds are integrated into a context of words, sentences and stories. Instruction is provided in story structure, specific comprehension skills, metacognitive strategies for self-assessment and self-correction, and integration of reading and writing. Specific adaptations are made for English-language learners being taught in English.

When students reach the Year 2 reading level, they use a programme called *Wings*, an adaptation of Co-operative Integrated Reading and Composition (CIRC) (Stevens et al. 1987). Wings uses co-operative learning activities built around story structure, prediction, summarisation, vocabulary building, decoding practice, and story-related writing. Students engage in partner reading and structured discussion of novels, poetry or non-fiction texts, and work in teams toward mastery of the vocabulary and content of the story. Story-related writing is also shared within teams. Co-operative learning both increases students' motivation and engages students in cognitive activities

known to contribute to reading comprehension, such as elaboration, summarisation and rephrasing (see Slavin 1995). Research on CIRC has found that it significantly increases students' reading comprehension and language skills (Stevens et al. 1987).

In addition to these story-related activities, teachers provide direct instruction in reading comprehension skills, and students practice these skills in their teams. Classroom libraries with real books at students' reading levels are provided for each teacher, and students read books of their choice for homework for 20 minutes each night. Home readings are shared via presentations, summaries, puppet shows, and other formats once a week during 'book club' sessions.

Materials to support *Reading Wings* through Year 6 are built around children's literature, poetry and non-fiction texts. Beginning in the second semester of programme implementation, SFA schools implement a writing programme based primarily on co operative learning principles (see Stevens et al. 1987).

Students in Years 1 to 6 are regrouped by reading performance level. The students are assigned to heterogeneous, age-grouped classes most of the day, but during a regular 90-minute literacy period they are regrouped by performance levels into reading classes of students at the same level. For example, a literacy class might contain Year 1, Year 2, and Year 3 pupils all reading at the same level. The literacy classes are often smaller than homerooms, because tutors and other certified staff (such as special needs support staff, teachers of English as an Additional Language, and teaching assistants) teach reading during this common reading period. Regrouping allows teachers to teach the whole literacy class without having to break the class into reading groups. This greatly reduces the time spent in 'independent work' and increases the time for direct instruction, eliminating workbooks, repetition, or other follow-up activities needed in classes with a wide range of reading ability groups. The regrouping is a form of the Joplin Plan, which has been found to increase reading achievement in the primary grades (Slavin 1987; Gutiérrez and Slavin 1992).

Eight-week reading assessments

At eight-week intervals, reading teachers assess student progress through the reading programme. The results of the assessments are used to determine who is to receive tutoring; to change students' reading groups; to suggest other adaptations in students' programmes; and to identify students who need other types of assistance, such as family interventions or screening for vision and hearing problems. The assessments are curriculum-based measures that include teacher observations and judgments as well as more formal measures of reading comprehension.

Reading tutors

One of the most important elements of Success for All is the use of tutors to promote students' success in reading. One-to-one tutoring is the most effective form of instruction known (see Wasik and Slavin 1993). The tutors are either certified teachers or well-qualified paraprofessionals. Tutors work one-on-one with students who are having difficulties keeping up with their reading groups. The tutoring occurs in 20-minute sessions during times other than reading or math periods.

In general, tutors support students' success in the regular reading curriculum, rather than using separate materials. For example, the tutor will work with a student on the same story and concepts that are being read and taught in the regular reading class. However, tutors seek to identify learning problems and use different strategies to teach the same skills. They also teach metacognitive skills beyond those taught in the classroom programme. Schools may have as many as six or more teachers or assistants serving as tutors depending on school size, need for tutoring and other factors.

During daily 90-minute reading periods, certified tutors serve as additional literacy teachers to reduce the literacy class size. Literacy teachers and tutors use brief forms to communicate about students' specific problems and needs, and meet at regular times to co-ordinate their approaches with individual children.

Initial decisions about literacy group placement and the need for tutoring are based on informal reading inventories that the tutors give to each child. Subsequent reading group placements and tutoring assignments are made using the curriculum-based assessments described above. Year 1 pupils are given priority for tutoring, on the assumption that the tutors' primary function is to help all students to succeed in reading the first time, before they fail and need remedial help.

Family support team

Parents are an essential part of the formula for success in Success for All. A Family Support Team in each school makes families feel comfortable in the school and become active supporters of their child's education, as well as providing specific services. The Family Support Team consists of a parent liaison, assistant headteacher (if any), counsellor (if any), facilitator, and any other appropriate staff.

The Family Support Team first works toward establishing and maintaining good relations with parents and increasing parental involvement in the schools. Team members may carry out 'welcome' visits for new families. They organise many attractive programmes in the school, such as parenting skills workshops. Most schools use the Raising Readers programme, in which parents are given strategies to use in reading with their own children. Family Support Teams also help teachers implement a social skills curriculum, Getting Along Together, which emphasises peaceful solutions to interpersonal problems.

The Family Support Team also intervenes to solve problems. For example, they may contact parents whose children are frequently absent to see what resources can be provided to help the family in getting their child to school. Family support staff, teachers, and parents work together to solve school behavior problems. Family support staff also provide assistance when students seem to be working at less than their full potential because of problems at home. Families of students who lack adequate sleep or nutrition, need glasses, are not attending school regularly, or are exhibiting serious behavioural problems, may receive family support assistance.

The Family Support Team is well integrated into the school's academic programme. It receives referrals from teachers and tutors regarding children who are not making adequate academic progress, and thereby constitutes an additional stage of intervention beyond that provided by the classroom teacher or tutor. The Family Support Team also encourages and trains the parents to fulfil numerous volunteer roles within the school, from providing a listening ear to emerging readers, to helping in the school cafeteria.

Programme facilitator

A programme facilitator at each school helps to oversee the operation of the SFA model. Facilitators help plan the programme, work with the head teacher on scheduling, and visit classes and tutorial sessions frequently to assist teachers and tutors with individual problems. They work directly with teachers on implementation of the curriculum, classroom management and other issues; help teachers and tutors deal with behavioural or other problems; and co-ordinate Family Support Team activities with those of the instructional staff.

Teachers and teacher training

Teachers and tutors receive detailed manuals supplemented by three days of in-service training at the beginning of the school year. Throughout the year, school and classroom follow-up visits are made by trainers, and additional in-service presentations are made by the facilitators and other project staff on such topics as classroom management, instructional pace, and co-operative learning. Facilitators also organise many informal sessions to allow teachers to share problems and solutions, suggest changes, and discuss individual children. The staff development model used in Success for All emphasises relatively brief initial training with extensive classroom follow-up, coaching and group discussion.

Special education

Every effort is made to deal with students' learning problems within the context of the regular classroom, as supplemented by tutors. Tutors evaluate students' strengths and weaknesses and develop strategies to teach in the most effective way. In some schools, special education teachers work as tutors and literacy teachers with students identified as learning disabled, as well as with other students experiencing learning problems who are at risk for special education placement. One major goal of Success for All is to keep students with learning problems out of special education if at all possible, and to serve any students who do qualify for special education in a way that does not disrupt their mainstream classroom experience (see Slavin 1996).

US research on Success for All

Early research

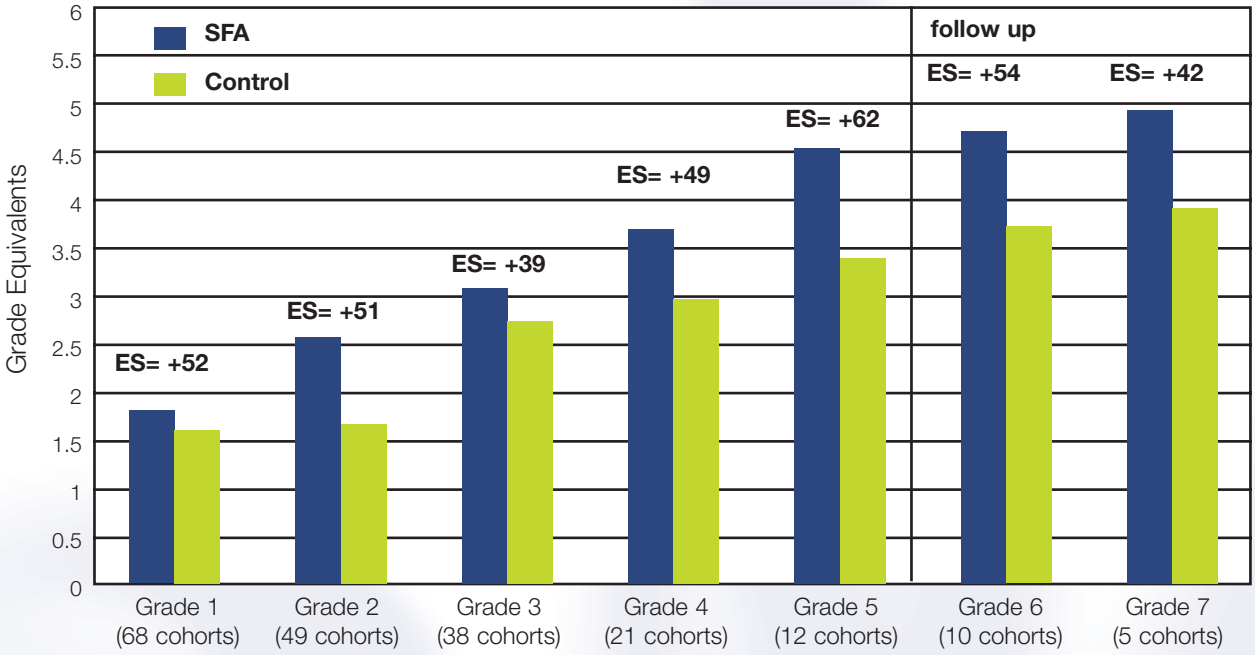
The early research on Success for All used a consistent paradigm. In each case, children were pre-tested (usually on the Peabody Picture Vocabulary Test) on entry to kindergarten or first grade, and then followed over time with individually administered reading tests given to all children each spring. These were typically scales from the Woodcock Reading Mastery Test and the Durrell Oral Reading Test.

From the first studies, it was clear that Success for All was making a substantial difference. Longitudinal studies of the first five schools in Baltimore found that these schools gained substantially more than matched controls, with effect sizes averaging around 50% of a standard deviation for students in general and more than a full standard deviation (ES=+1.00) for students

who began in the lowest 25% of their grades (Slavin et al. 1990, 1996; Madden et al. 1992, 1993). This paradigm was ultimately followed in schools in 11 districts around the US, and the results continued to strongly support the programme's impact (see Dianda and Flaherty 1995; Livingston and Flaherty 1997; Nunnery et al. 1997).

Figure 1 summarises the results from studies of various duration, from one to six years. The figure shows that by the end of fifth grade, students in SFA schools were performing about a full grade equivalent higher than matched control schools on individually administered tests. In addition to effects on achievement, studies found substantial impact on assignments to special education (Slavin 1996) and other outcomes (Slavin and Madden 1996, 2001).

Figure 1. Comparison of Success for All and control schools in mean reading grade equivalents and effect sizes 1988–99



Note: Effect size (ES) is the proportion of a standard deviation by which Success for All students exceeded controls. Includes approximately 6000 children in Success for All or control schools since the first grade.

Later research

After the many studies establishing the basic effects of Success for All, research attention has shifted in different directions. One line of research has focused on effects for English language learners, evaluating both a Spanish bilingual and an English language development adaptation (see Slavin and Madden 1999; Slavin and Cheung, in press). Both adaptations have been found to be effective. Research correlating quality and completeness of implementation with student outcomes has been a focus (Nunnery et al. 1997; Ross et al. 1995). A longitudinal follow-up of students who had been in the original Baltimore schools found that by eighth grade, these students were still performing significantly better on standardised reading measures than former control students, and were substantially less likely to have been retained in grade or assigned to special education (Borman and Hewes 2003).

Because of demands from policy audiences, some attention has shifted to studies that take data from routine state assessments. Formal studies in Texas (Hurley et al. 2001) and California (Slavin et al. 2002) have found substantially higher gains for SFA students than for the state as a whole. Similar analyses have found the same patterns in nearly every state with more than ten SFA schools. Such comparisons are less scientific than the longitudinal studies, but they inform policy-makers and educators how the programme performs on the assessments for which they are held accountable.

As research on comprehensive reform programmes and on reading programmes has taken on greater political and practical importance, a number of reviews of the research have appeared. The American Institute of Research (Herman, 1999) rated comprehensive reform models and found Success for All to be one of two elementary programmes with the strongest evidence of effectiveness. This conclusion was echoed in a report for the Thomas Fordham Foundation (Traub 1999). A recent meta-analysis (Borman et al. in press) identified 41 experimental control comparisons done to evaluate Success for All, of which 25 were done by third parties. This was the largest number of such studies for any comprehensive reform model. Borman et al. listed SFA as one of three programmes with strongest evidence of effectiveness. Finally, Pearson and Stahl (2002) gave SFA the highest rating for evidence of effectiveness among all core reading programmes.

Research on Success for All in England

As in the US, the implementation of Success for All in England has emphasised evaluation of the programme's outcomes. Three very different evaluations have been carried out. The first (Hopkins et al. 1999) focused on the first year in the first five schools to use the programme in England, all located in a disadvantaged neighbourhood in Nottingham. The second study (Tymms and Merrell 2000, 2001) evaluated four schools over a two-year period (see also Hopkins et al. 2002). Finally, an evaluation of gains from 2000 to 2002 on Key Stage 1 and 2 passing rates is presented for all of the schools that had begun SFA by Autumn 2000.

Hopkins et al. (1999)

David Hopkins, Mick Youngman, Alma Harris and Judith Wordsworth, all then at the University of Nottingham, carried out a preliminary evaluation of Success for All in its five pilot sites in Nottingham. In Autumn 1998, children in Years 1–6 were assessed on the Early Reading Progress test (Youngman and Parkins 1998). They were then post-tested in Spring 1999. Expected gains were computed by subtracting each year group's score from that of the next higher group. In Years 1–3, SFA students showed substantially greater gain than expected. Years 4 and 5 also exceeded expectations, although to a lesser degree, but Year 6 students approached the maximum test score and therefore showed a ceiling effect. Qualitative evidence noted positive changes in student behavior, motivation and attitudes. Great variation in quality of implementation was noted, however, and this variation corresponded with differences in student outcomes.

Tymms and Merrell (2001)

Peter Tymms and Christine Merrell, of Durham University, conducted a second study of Success for All. This study focused on data from the Performance Indicators in Primary Schools (PIPS), as well as SATs for Key Stage 1 and 2.

Four SFA schools were involved in this study: two primary schools, an infant school, and the junior school attended by children from the infant school. This means that in any particular grade year, only three schools were represented.

PIPS assessments were administered to all students in all grades as a pre-test in Autumn 1999, and then as a post-test in Spring 2000 and Spring 2001. Residualised scores (controlling for pre-tests) at the end of the second year showed positive trends in reading for Years 1, 2 and 6, but negative trends in Reception and Years 3 and 4. Maths scores, of interest only in assessing possible carry-over effects of the SFA reading programme, were similarly equivocal, as were science scores in Year 6.

Key Stage 1 data favored the SFA groups in reading, writing and spelling (but not maths). However, KS2 data were similar in experimental and control groups.

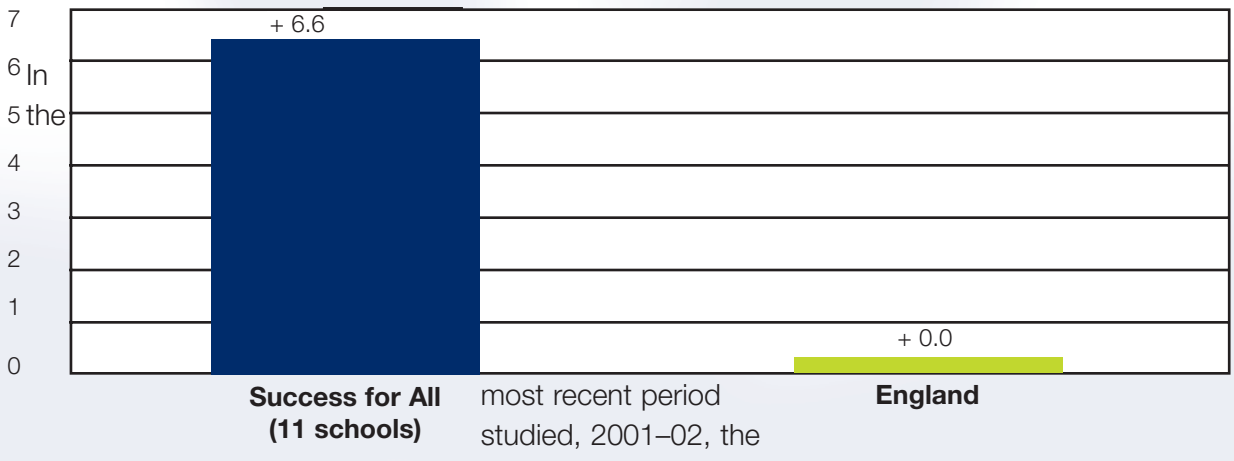
Reading attitude measures favored the SFA schools, but there were no differences in math attitudes or overall school attitudes.

In sum, the Tymms and Merrell (2001) evaluation found positive reading effects for Years 1 and 2, both on PIPS and on KS1 passing rates, but effects were mixed for other year groups.

National Key Stage 1 and Key Stage 2 gains

The third evaluation tracked gains in KS2 passing rates for all schools implementing SFA in England. Figure 2 summarises the gains on KS2 for all 11 schools that had begun Success for All by 2000. As the figure shows, SFA schools averaged a gain of 6.6 percentage points on KS2 from 2000 to 2002, while English schools as a whole were unchanged.

Figure 2. Success for All in England: gains in percent scoring at level 4 or above, Key Stage 2, Year 6, 2000–02



In the most recent period studied, 2001–02, the

16 schools that had begun SFA by 2001 gained 2.5 percentage points on KS2, while other schools were unchanged. On Key Stage 1, the SFA schools gained 9.4 percentage points, with ten of the schools averaging more than the national average of 82% passing. This is important considering that these schools serve some of the most disadvantaged areas in England.

Conclusion

Success for All has made a good start in England. The programme has demonstrated that it can operate effectively in the English context, and each year increasing numbers of schools adopt it. Three preliminary studies show variable but promising impacts on student reading performance, particularly at Key Stage 1 and Key Stage 2, the most important indicators of reading success in the UK.

However, the research evidence available to date must be considered preliminary. The first two studies involved too few schools and students for reliable estimates of programme impact. The KS1 and KS2 gains for 11 schools in 2000–02 are more robust and meaningful, but the gains for the SFA schools can be compared only to those for the nation as a whole, not to control groups designated in advance.

Success for All has shown enough promise in implementation and outcome to merit a larger-scale evaluation, hopefully by third party. A study involving at least ten experimental and ten control schools, preferably assigned at random to conditions, would scientifically establish the effects of the programme both on individually administered measures and on KS1 and KS2 assessments.

The literacy levels of disadvantaged students in England remain too low to permit complacency. In a study of the NLS, Earl et al. (2003) noted that for NLS to move to the next level of impact, teacher and school capacity issues must be addressed, especially in disadvantaged schools. Success for All and other robust, replicable models capable of working at a large scale may offer England's most disadvantaged schools an opportunity to close the gap, to achieve the literacy levels envisioned for all schools in the NLS. It is time to move to the next stage to test this possibility.

Acknowledgements

Implementation and evaluation of Success for All in the United Kingdom has been funded primarily by the Fischer Family Trust and the Department for Education and Skills. Major contributors to the research have included David Hopkins, Alma Harris, Mick Youngman, Michael Sinanan, Peter Tymms and Christine Merrell.

References and bibliography

Adams MJ (1990). *Beginning to read: Thinking and learning about print.* MIT Press.

Borman G, Hewes G (2003). Long-term effects and cost effectiveness of Success for All. *Educational Evaluation and Policy Analysis,* 24(2), 243–266.

Borman GD, Hewes G, Overman LT, Brown S (in press). Comprehensive school reform and student achievement: A meta-analysis. *Review of Educational Research.*

Dianda M, Flaherty J (1995). *Effects of Success for All on the reading achievement of first graders in California bilingual programs.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Earl L, Watson N, Levin B, Leithwood K, Fullan M, Torrance N (2003). *Watching and learning 3: Final report of the external evaluation of England's National Literacy and National Numeracy Strategies.* University of Toronto, Ontario Institute for Studies in Education.

Gutiérrez R, Slavin RE (1992). Achievement effects of the nongraded elementary school. A best-evidence synthesis. *Review of Educational Research,* 62, 333–376.

Herman R (1999). *An educator's guide to schoolwide reform.* Educational Research Service.

Hopkins D, Harris A, Sinanan M (2002). *Evaluation of SFA-UK: Final report.* University of Nottingham, School of Education.

Harris, A., Hopkins D., Youngman, M., & Wordsworth, J. (2001). The implementation and impact of Success for All in English schools. In R.E. Slavin (Ed.), *Success for All: Research and reform in elementary education.* Mahwah, NJ: Erlbaum.

Hurley EA, Chamberlain A, Slavin RE, Madden NA (2001). Effects of Success for All on TAAS Reading: A Texas statewide evaluation. *Phi Delta Kappan,* 82(10), 750–756.

Livingston M, Flaherty J (1997). *Effects of Success for All on reading achievement in California schools.* WestEd.

Madden NA, Slavin, RE, Karweit NL, Dolan LJ, Wasik BA (1993). Success for All: Longitudinal effects of a restructuring program for inner-city elementary schools. *American Educational Research Journal, 30,* 123–148.

National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* National Institute of Child Health and Human Development.

Nunnery J, Slavin R. Madden N, Ross S, Smith L, Hunter P, Stubbs J (1997). *Effects of full and partial implementations of Success for All on student reading achievement in English and Spanish.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Pearson PD, Stahl S (2002). *Choosing a reading program: A consumer's guide. (Technical report).* University of California.

Ross SM, Smith LJ, Casey J, Slavin RE (1995). Increasing the academic success of disadvantaged children: An examination of alternative early intervention programs. *American Educational Research Journal,* 32, 773–800.

Slavin RE (1987). Ability grouping and student achievement in elementary schools: A best-evidence synthesis. *Review of Educational Research,* 57, 347–350.

Slavin RE (1995). *Co-operative learning: Theory, research, and practice* (2nd edn). Allyn and Bacon.

Slavin RE (1996). Neverstreaming: Preventing learning disabilities. *Educational Leadership,* 53(5), 4–7.

Slavin RE, Cheung A (in press). Effective early reading programs for English Language Learners. *The Handbook of Research on the Education of Young Children.*

Slavin, R.E., Madden, N.A., Karweit, N.L., Livermon, B.J., & Dolan, L. (1990). Success for All: First-year outcomes of a comprehensive plan for reforming urban education. *American Educational Research Journal,* 27, 255-278.

Slavin RE, Madden NA (1996). *Success for All/Roots and Wings: 1996 summary of research on achievement outcomes.* Johns Hopkins University, Center for Research on the Education of Students Placed at Risk.

Slavin RE, Madden NA (Eds.) (2001). *One million children: Success for All*. Corwin.

Slavin RE, Madden NA (1999). Effects of bilingual and English as a second language adaptations of Success for All on the reading achievement of students acquiring English. *Journal of Education for Students Placed at Risk* 4(4), 393–416.

Slavin RE, Madden NA (2000). Research on achievement outcomes of Success for All: A summary and response to critics. *Phi Delta Kappan,* 82(1), 38–40, 59–66.

Slavin RE, Madden NA (2003, April). *Scaling up Success for All: Lessons for policy and practice.* Paper presented at the annual meeting of the American Educational Research Assocation, Chicago.

Slavin RE, Madden NA, Dolan LJ, Wasik BA (1996). *Every child, every school: Success for All.* Corwin.

Slavin RE, Madden NA, Liang C (2002). *Effects of Success for All on SAT-9 reading: A California statewide evaluation.* Johns Hopkins University, Center for Research on the Education of Students Placed at Risk.

Stevens RJ, Madden NA, Slavin RE, Farnish AM (1987). Co-operative Integrated Reading and Composition: Two field experiments. *Reading Research Quarterly,* 22, 433–454.

Traub J (1999). *Better by design? A consumer's guide to schoolwide reform.* Thomas Fordham Foundation.

Tymms P, Merrell C (2000). *Success for All evaluation report.* University of Durham, Curriculum Research and Management Centre.

Tymms P, Merrell C (2001). *Success for All quantitative evaluation report: Second year.* University of Durham, Curriculum Evaluation and Management Centre.

Wasik BA, Slavin RE (1993). Preventing early reading failure with one-to-one tutoring: A review of five programs. *Reading Research Quarterly*, 28, 178–200.

Youngman M, Parkins E (1998). *Evaluating early reading progress using the ERP.* University of Nottingham, School of Education.

Time to be bold: mainstreaming models of educational research that respond to user needs

Dr Marilyn Leask, Teacher Training Agency

This paper summarises recent developments in educational research, and challenges policy-makers, educational researchers and teacher educators to collaborate on new ways of undertaking and communicating research to meet the needs of potential users.

Debates in the education sector in England about how research can usefully inform classroom practice can be tracked back to the 1970s, if not further. Demands from policy-makers for research evidence to inform policy-making seem to be associated with the Labour government elected in 1997. However, teachers, teacher educators and policy-makers who wish to use research evidence as a foundation for their work are finding that the evidence base is incoherent and inaccessible.

In 2002, the Organisation for Economic Co-operation and Development (OECD) reviewed educational research in England and produced recommendations for improvement. The earlier Hillage report (1998) and the DfES briefing note prepared for the OECD (DfES 2002) also provide an evidence base to underpin development. These three reports set out key weaknesses in research communication and utilisation in England. While some of these have been tackled in the intervening years, others are being neglected.

This paper describes a number of developments that are delivering improvement in the access to and coherence of research. However, many of these initiatives are not widely known among users of research, and so the initiatives are not embedded in the education system. Many of these developments and innovations are fragile and will need sustained support and development by research users if new practices are to become widespread. There are major obstacles in moving to evidence-informed policy and practice in education – obstacles that can be overcome only by joint effort from all stakeholders.

There are five key constraints on the communication and utilisation of research for policy and practice:

- Traditional models of research and research reporting are not designed to communicate to users, nor to provide a foundation for users to change practice;
- Research methods commonly used yield that results lack generalisability. Sample sizes are too small to generate a base for developing national policy;
- Access to accumulated knowledge of our society is limited;
- Models of commissioning research are outdated;
- Processes for identifying useful research questions for teachers need to be extended.

There are other constraints, but this paper focuses on these five, which particularly affect the remit of the Teacher Training Agency (TTA) in improving the quality of teacher education. The discussion that follows describes models of practice that could help overcome these constraints.

Background: the evidence base for the need for change

In 2002, the OECD examined in England:

> *… the contribution of educational R&D to the knowledge base of education from which teachers and policy-makers can draw … (OCED 2002).*

The OECD report indicates that a number of countries are concerned about the lack of impact of educational research on practice. France, Denmark, New Zealand, Scotland, Switzerland and Wales are mentioned. US concerns about the lack of useful educational research are similar to those in the UK (see Skeeter 2004; Slavin 2004).

The DfES briefing document produced for the OECD examiners (DfES 2002) outlines the criticisms identified in the Hillage Report (1998) about the lack of usefulness of educational research to practitioners and policy-makers in England. The DfES document also provides information about national educational initiatives developed to address these concerns. These include:

- The National Educational Research Forum (NERF), which established the Educational Research Funders' Forum – a group of all major funders of educational research working to develop increased coherence in educational research programmes;
- The Evidence-based Policy and Practice Information Centre (EPPI Centre) based at the Institute of Education, University of London, which was contracted to develop methodologies for the syntheses of research findings;
- The National Teacher Research Panel, providing a bridge between policy-makers, headteachers and teachers;
- Best practice research scholarships – national funding for teacher–researcher projects (which was later devolved to schools);
- The Research Liaison Group, working across government agencies, towards synergy in national research programmes and approaches.

A wide range of initiatives are referenced in the briefing document as well. NERF has subsequently developed:

- The NERF bulletin *Evidence for teaching and learning*, summarising key research findings;
- A cross-agency group working to develop the inter-operability of research databases (the National Educational Evidence Portal project);
- A cross-agency group working to develop a national strategy to use research synthesis to inform policy-makers and practitioners about the evidence base for practice (National Strategy for Systematic Reviews in Education group).

The Hillage report (1998) had already identified issues of scale and national models rewarding researchers, as factors limiting the usefulness of research:

> *Research that addressed issues relevant to policy and practice was too small scale, incapable of generating findings that are generalisable, insufficiently based on existing knowledge and inaccessible. Pressure on researchers to produce empirical findings in published journals of international repute reflected different priorities to those which 'users' in the system need to inform policy and practice. (DfES 2002)*

While the OECD report acknowledged that the above initiatives were addressing many of the issues raised, it also identified what seem to be enduring problems:

> *Hillage … concluded that the connections among research, policy quality and practice were weak; that research was too supplier-driven; that an emphasis on short-term evaluation, at the expense of exploration and development, led research to follow rather than lead policy; that studies examining practice were small-scale and unable to generate findings that could be generalised; that research findings were disseminated ad hoc; and that policy-makers and practitioners lacked the capacity to use research when it was available.* (OECD 2002)

The OECD review, the DfES briefing document and the Hillage report made recommendations for the next stages of development.

However, the actions needed to improve the role of research in improving quality in teacher training require difficult collaborations between stakeholders to achieve change.

Such collaborations are difficult because they require culture change. They require policy-makers, researchers and other users to work together on a national solution that is beyond the remit of any group. And change is difficult because, to address the shortcomings, all stakeholders would need to change practice.

The DfES report listed the following as key issues that need to be addressed:

- *Continuing to develop and make more transparent the criteria for judging quality across the range of methodologies in educational research;*
- *Generating more high quality evidence capable of having an impact on policy and/or practice;*
- *Providing more development opportunities in research methods;*
- *Improving the access to currently available 'best' evidence;*
- *Establishing the support and involvement of many, if not most educational researchers to contribute to systematic reviews;*
- *Securing long term resources for systematic reviewing;*
- *Improving the capacity of policy-makers to access and use research;*
- *Developing greater demand for, understanding of and opportunities to participate in research amongst practitioners;*

- Supporting the development of greater collaboration between higher education, local education authorities and schools on research which will contribute towards genuine user engagement.

  This list involves culture changes at every level which are beginning to occur but have further to go. Policy-makers need to 'value' the role of evidence. Teachers need to look beyond their own schools for evidence. Funders need to make user engagement and planning for dissemination and impact requirements of research funding. Researchers need to be rewarded for appropriate achievements relating to impact in assessments of research. (DfES 2002)

How much progress has been made?

Among the OECD recommendations were an acknowledgement that, as a first strategy, the DfES Centres of Excellence approach was a positive step. Other suggestions included:

- *A second important strategy is to adequately fund a small number of very ambitious, but carefully designed, studies to look closely at a difficult problem of practice ... For example, one study might examine ways of improving professional development by focusing on effective use of formative assessments in elementary or middle schools (see Part VII) [this suggestion was based on the outcomes of the long-standing Assessment Reform Groups' systematic reviews.];*
- *A third strategy is to create networks of researchers and practitioners around core problems of practice. Such networks can take a variety of forms, including a virtual structure using the Internet, satellite conferencing for formal or informal meetings.*

(OECD 2002)

The OECD report stated that current structures would not help policy-makers and practitioners progress the issues that need to be resolved:

> *90% of educational R&D is carried out in university departments of education. While at least 100 separate institutions conduct education research, 80% of the funding from government, Research Councils and charities goes to 22 university departments or schools of education.* (OECD 2002)

The reviewers went on to make the point that the current single university-oriented approach to research, focused on individual expertise, would not support the changes needed:

> *… the university is not conducive enough to large-scale problem-oriented work involving people working in teams, often in the field and at a considerable distance from the universities. (OECD 2002)*

Neither were RAE processes seen to support the changes needed. The HEFCE funding was described as providing:

> *… incentives that drive the allocation of HEFCE funds … to push HEFCE research away from having a practical bent.*

While these various reports state the key issues, they do not give specific advice about how to move forward.

A neutral forum is needed, where the different parties can come together to work on solutions. It may be that NERF could provide such a forum to take forward, actively, solutions to the problems that have been clearly identified. No other organisation appears to have the independence and national authority to bring groups together to address these issues. Commitment to embedding change is needed over a decade to get change embedded in policy-making and practice.

Constraints and new models

This section gives more detail on the five key constraints listed at the beginning of this paper. Existing models demonstrating more effective ways of working that could be applied more widely across the education sector are also described.

***Constraint 1*: Traditional models of research and research reporting are not designed to communicate to users, nor to provide a foundation for users to change practice.**

In education, the term 'research' is used for what other sectors would expect to be 'R&D'. Why? In other sectors – health and industry, for example – research and development are linked. Blue skies research is not excluded, but research is seen to have the goal of improving performance of some aspect of the sector. The point is that linking research with development is normal in some sectors, but not in education. Is it, then, surprising that users' needs are not met? The concept of R&D implies a link with practice and performance, whereas the term 'research' does not.

Models of research training in education, through the higher degree route, themselves fall short of providing a model where research influences practice, and thus they perpetuate existing outmoded practice.

There are, of course, major research studies that have broken this pattern and link educational research directly to the development of new practices, but the prevalent model of educational research does not link 'R' with 'D'.

Specific school-related R&D that comes to mind includes:

- The cognitive acceleration R&D work of Adey and Shayer (Adey and Shayer 1994, 1998; Adhami et al 1998);
- The peer tutoring R&D work on Talking Together by Mercer, Dawes and Wegerif (Dawes et al 1999, Mercer et al 2000);
- The Assessment For Learning R&D work (e.g. Black et al 2003) and the Assessment Reform Group (ARG) funded by Nuffield.

David Hopkins, at an ARG Seminar at the Nuffield Foundation (November 2004) said the ARG provides an excellent model of how theory and practice can interrelate. But we need such models in every major area of pedagogy and every subject, and we need research that answers, systematically, quite fine-grained questions that teachers want answered.

***Constraint 2*: Common research methods yield results lacking generalisability, and sample sizes are too small to generate a base for developing national policies.**

The OECD pointed out that our system stimulates small-scale research with non-generalisable findings. This represents a huge waste of resources when there is no attempt to accumulate the knowledge gained from these studies. Apart from publication kudos, what is the point of all this effort?

Slavin (2004) compares developments in medicine, where there is a developing evidence base, with those in education:

> *The most important reason for the extraordinary advances in medicine, agriculture, and other fields is the acceptance by practitioners of evidence as the basis for practice.*

He adds, as a second reason, the importance of practitioners accepting the need to undertake large-scale studies from which findings can be generalised.

Whether you agree with Slavin or not, the fact remains that the funds (including staff research time) currently expended on educational research turn up a plethora of small studies – 5,000 per year are added to the British Education Index. How is a teacher, a teacher educator or a policy-maker to access the research findings relevant to them? How are they to know which studies to take seriously and which to ignore? No policy-maker and no practitioner can claim to keep up to date with this volume of new work.

Because of the small scale of the studies, few provide an evidence base strong enough to warrant changes in practice, so the current system of research and publication seems to minimise the usefulness of much research.

Leading professionals in education do accept that practice should be underpinned by research. There are networks of teacher-researchers and research-active schools, often supported by research-active teacher educators, who are often tutors on higher degree programmes where teachers undertake small-scale research.

How can the stakeholders in educational research shift the balance of educational research from numerous small-scale studies, to studies with large sample sizes and more chance of generalisable findings?

Perhaps models of collaborative research that give cost-effective access to large datasets could be developed. Outcomes from existing datasets also need to be maximised. Educational researchers who currently work on small-scale projects could be encouraged to collaborate to undertake large-scale studies, which might be more likely to yield generalisable findings. There are cost-effective means of generating large sample sizes – developing research networks of researchers, schools and HEIs, for example.

Changing the expectations of journal editorial boards about what is worth publishing is part of a necessary practice shift. For example, there could be an expectation that published articles would normally be based on substantial studies. Or is the academic journal not relevant to the world where research is accessible to the user?

When MORI want to sample the views of the population, they turn to a regular group of respondents who comprise a representative sample across the country. This work is becoming increasingly accurate in reflecting the will of the people, for example at election times. Is there a model of value to the education sector here?

Could stakeholders in educational research establish cost-effective models that would harness the energies currently expended in small-scale research into larger, significant studies, with sample sizes sufficient for users to have confidence that changes in practice or policy are warranted? In many cases, perhaps, the studies may also generate results that provide confidence that continuing to do the same is appropriate.

Can we imagine establishing a national educational research network, and perhaps subject networks, linking a select number of LEAs/HEIs/schools, able to collect data about practice at the fine-grained level in particular? Perhaps it might focus on the questions teachers say they want answered as evidenced, for example, by requests on TRIPS, the DfES teacher research site. There might be designated research schools. Within a relatively small network – which would be structured to give representative samples of schools, pupils, teacher training institutions and LEAs – sample sizes of over 1,000 teachers, 1,000 teacher trainees and 1,000 pupils in any year group could be easily found. Quality assurance and management of the project would need to be done by a central core, and network members would be involved in developing research design, data collection, analysis and reporting as and when appropriate. LEA-level models similar to this have been developed and evaluated in the past (Hopkins 1989).

***Constraint 3*: limited access to the accumulated knowledge of our society.**

The DfES briefing paper for the OECD commented that research was often 'insufficiently based on existing knowledge and inaccessible' (DfES 2002).

How do researchers and users access the accumulated knowledge they expect to exist about educational practice? With extreme difficulty. Three ways of improving access would be:

- Better web access;
- Synthesis of existing research studies;
- Identification of the truly important areas for research and focusing significant resources on these, so that findings are considered to provide sound grounds for changing practice (see Constraint 4 below).

Better web access. The internet has the potential to improve access to evidence to underpin practice, but web-based support for researchers and users of educational research is as yet unfocused. There are many sites holding valuable resources. This is clearly a role for a national strategic initiative, and we note NERF's initiative to explore the possibilities of establishing a research portal for educational researchers and users of research. For this NERF initiative to be successful, financial and general support from the various agencies will be essential.

Synthesis of existing research studies. How can we accumulate knowledge – access and build effectively on what has gone before? There are options from current practice that could be developed; for example, meta-analysis of practitioner case studies and research synthesis approaches.

Meta-analysis of practitioner case studies – the Stenhouse library approach. Lawrence Stenhouse, a professor at the University of East Anglia was well known for his work in the 1970s in developing teacher-researcher models and the idea of analysing across cases. Teacher-practitioner research is often criticised for not being reliable enough to be taken seriously, but in large cross-case analysis (where the cases use carefully chosen methods), we could reasonably expect to reveal soundly based insights into effective practice across contexts.

What would have to happen for a nationally recognised library of case studies to be developed and made available on the web and for meta-analyses to be undertaken? Not too much perhaps:

- The will on the part of key people with access to funds; we already have examples, but work to date does not seem to have reached a critical mass;
- A website with appropriate meta-tagging system, with reports structured systematically so they could be cross-searched;
- A quality assurance scheme to ensure that standards of data collection, evidence and analysis, were appropriate for inclusion (perhaps Masters thesis standard would be a good starting point).

Many of these resources are already available and small-scale versions of what can be achieved exist on diverse websites.

Research synthesis approaches: Various funders are involved in developing systematic models for research synthesis – to gather the accumulated knowledge that we have. In this country, these are undertaken mainly, but not exclusively, through the EPPI-Centre. The EPPI-Centre, however, has established a recognisable model with strict quality criteria for systematic reviews of the evidence base.

Oddly enough, there is some tension in the education research community around the notion of research synthesis, but policy-makers and researchers have always relied on a form of research synthesis to provide a 'state of the art' statement before new work begins. However, the methodology of the traditional literature review model is demonstrably unreliable compared with new models that include explicit explanations of how conclusions are drawn and the strength of the evidence base for the conclusions.

Models for synthesising research findings are developing in different disciplines including an initiative in education (described above). Teams of educators are becoming experienced in the application of different models, but consistent commitment across the stakeholder group is needed to make the vision of a library of synthesised research guiding policy and practice a reality. The Campbell Collaboration (www.campbellcollaboration.org), the Cochrane Collaboration (www.cochrane.org) and the EPPI-Centre (http://eppi.ioe.ac.uk) provide examples of work in this area.

The Cochrane Collaboration, for example, has thousands of health professionals from 89 countries working on reviews. For at least some health professionals in England, involvement in reviews is counted towards continuing professional development, which is required to remain on the national register. Therefore, there is an incentive for professionals to keep up to date and to be involved in reviews.

***Constraint 4*: Outdated models of commissioning research**

The constraints identified above have implications for the research commissioning process. Those commissioning research could:

- Build into research contracts funding and processes to ensure impact – perhaps in the form of an additional contract where findings warrant this;
- Ensure that outcomes from research are usable by the audience, including production of audio-visual materials that demonstrate the new practice;
- Ensure that the outcomes are easily accessible 'anywhere any time' and remain accessible while they are relevant;
- From the outset, involve those who will be considering and possibly implementing any changes warranted;
- To maximise value for money of research funds invested, increase coherence between funders of research;
- Know enough about the levers for change in the context relevant to the research to ensure findings are taken seriously by policy-makers and practitioners if change is warranted.

The Social Research Association (www.the-sra.org.uk/commissioning.pdf) has published guidelines for commissioning research that provide useful advice for policy-makers and researchers.

***Constraint 5*: Identifying research questions useful to users**

Who has the money for research? Who decides the questions? These are two major constraints on educational research meeting user needs.

It should be of no surprise that the needs of teachers and teacher trainers for useful research are not met. National funding models and major commissioners of research do not usually target the teacher as user. Nationally, research funds are accessed principally through the Research Assessment Exercise, educational charities, government agencies, and the Economic and Social Research Council. The questions to be answered will be determined by the charity, by the researcher, or by government priorities. These are rarely at the fine-grained level of what the teacher needs to know in order to do things differently in the classroom in a particular situation. For example, as a science teacher, I might want to know the effectiveness of quite specific actions I can take in science – ways of explaining, particular types of homework, particular types of questioning specific to the context of the subject. These are research questions of a very fine-grained type that are unlikely to be answered through current commissioning processes. They are questions that might be addressed through MA theses and teacher-practitioner research.

In England, there is no systematic programme of research into pedagogy. Traditionally, government agencies commission research to meet policy objectives that are usually at a higher level than pedagogy in individual subjects. The subject of ICT is in a slightly different position. Becta gathers all the research articles related to ICT together and is in the fortunate position to fund considerable research into ICT pedagogy. Other subject areas, such as Citizenship and PE, have had the benefit of major investment from government agencies.

NERF has a brief to bring coherence to educational research, but the work that is required to meet the needs of the users, as described above, goes beyond the current remit of NERF. It would require an attempt on the part of all agencies to have a systematic programme of research in different subject and cross-curricular areas. And, of course, before any research like that should be done, reviews of the current state of knowledge would have to be undertaken. The task is substantial. But it is a task worth tackling if we consider evidence-informed policy and practice to be a goal worth striving for.

Conclusions

> *Evidence-based policies have great potential to transform the practice of education, as well as research in education. Evidence-based policies could finally set education on the path toward progressive improvement that most successful parts of our economy and society embarked upon a century ago. With a robust R&D enterprise and government policies demanding solid evidence of effectiveness behind programs and practices in our schools, we could see genuine, generational progress instead of the usual pendulum swings of opinion and fashion.*
>
> (*Slavin 2004*)

Those of us whose needs for research are not satisfied must ask ourselves these questions:

- What have we done to enable the research community to meet our needs?
- What could be done to ensure that the research we need is available when we need it?
- What is our role in developing clearer and shared understandings of the roles that teachers, teacher educators, LEA staff, policy-makers, researchers and different agencies may play as agents for change in the education system where research findings warrant change?

Communication between teachers, policy-makers and educational researchers about the use of research and evidence in education is poor. To use a modern metaphor, communication seems to be operating in a form of Morse code – and few people understand Morse code nowadays. Even the shipping industry has dropped it as a medium of communication. With some stakeholders in educational research, connections might be at the equivalent of a 56K modem. What we need are broadband connections between stakeholders. Each of the stakeholders in educational research needs to consider what the challenge of moving to broadband communication with other stakeholders means for them.

Note: *Dr Marilyn Leask has responsibility at the Teacher Training Agency for national strategies supporting the development, dissemination and implementation of evidence-based practice in Initial Teacher Training. She writes here in a personal capacity, drawing on her background as an educational researcher and teacher trainer, as well as on her experiences in a government agency, in policy development and in collaborative work across government agencies.*

References and bibliography

Adey P and Shayer M (1994). Improving learning through cognitive intervention. London: Routledge. See also King's College London, Cognitive Acceleration through Science Education (CASE) and Cognitive Acceleration through Maths Education (CAME) www.kcl.ac.uk/education/case.html and www.kcl.ac.uk/education/research/mathsed.html.

Adey P and Shayer M (1994). *Really Raising Standards*. London: Routledge.

Adhami M, Johnson D.C. and Shayer M., (1998). *Thinking Maths: The Programme for Accelerated Learning in Mathematics.* Oxford: Heinemann Educational Books.

Bassey M (November, 2003). Retirement Speech (as Academic Secretary of BERA) at the Conference Dinner in the Museum of Scotland, *Research Intelligence:* 85, 8–9.

Black P, Harrison C, Lee C, Marshall B and Wiliam D (2003). *Assessment for Learning: Putting it into practice.* Maidenhead: Open University Press.
This work has been taken up by various governments. See Learning and Teaching Scotland for a Scottish perspective (http://www.ltscotland.org.uk/assess/); http://cms.curriculum.edu.au/assessment/default.asp for an Australian perspective; and http://www.standards.dfes.gov.uk/keystage3/respub/afl_ws Learning Intelligence for an English perspective.

Bell M, Cordingley P (2000). *MA and PhD Dissemination: Off the Shelf.* Consultation Report. TTA.

Dadds M, Kynch C (2003). *The Impact of RAE 3B rating on educational research in Teacher Education Departments: summary report.* St Martin College, Lancaster.

Dawes L, Mercer N and Wegerif R (2000). *Thinking Together: A Programme of Activities for Developing Thinking Skills at KS2.* Birmingham: Questions Publishing.
See also the Thinking Together website on http://www.thinkingtogether.org.uk/.

DfES (2002). Research and development in England. Background report prepared for the OECD review. Accessed at: www.dfes.gov.uk/research/downloads/OECDreview.doc

Dyson A, Desforges C (2002). *Building research capacity: some possible lines of action.* National Educational Research Forum.

Farrant J, Billing D, Temple P (2003). *Operational Review of the Research Assessment Exercise 2001: Report to the Joint Funding Bodies.* Research assessment review. UNIVERSITAS Higher Education Management Consultants.

Fielding M, Eraut M, Thorp J, Cunningham I, Jones C, Craig J, Horne.M (2001). *Factors influencing the transfer of good practice: literature review.* University of Sussex and DEMOS.

Furlong J (2003). From the President. Research Intelligence, 85, 1–3. http://www.bera.org.uk.

Hargreaves D (1996). *Teaching as a research-based profession: possibilities and prospects.* The Teacher Training Agency Annual Lecture.

Hillage J, Pearson R, Anderson A, Tamkin P (1998). *Excellence in Research on Schools.* DfEE.

Hopkins D (1989). *Evaluation for School Improvement.* Milton Keynes, Open University Press.

KPMG (2002). *Supply Constraints in Initial Teacher Training.* Final Report. TTA

Kushner S (2003). Dislocating Educational Research from Teacher training: a response to Michael Bassey. *Research Intelligence,* 85, 24–26.

Mercer N, Wegerif R and Dawes L (1999). Childrens' talk and the development of reasoning in the classroom. *British Educational Research Journal*, 25, 1,95-111. See also the Thinking Together website on http://www.thinkingtogether.org.uk/.

NERF (2000). *Building Research Capacity* – sub-group report. National Educational Research Forum.

Oates T (2003). *The effects of the Research Assessment Exercise: barriers to establishing systematic review in research into education and training.* QCA.

OECD/CERI (2002). *Educational Research and Development in England.* OECD Examiners Report. OECD/CERI.

Roberts G (2003). *The Roberts Report: Review of research assessment: report to the UK funding bodies.* SHEFC, HEFCE, HEFCW DELNI http://www.ra-review.ac.uk/reports/roberts.asp

Skeeter C (2004). From the Vice President, *Teaching and Teacher Education,* Fall 2004 pp1,6.

Slavin R (2004). *Evidence-based reform in Education: Promises and Pitfalls,* address to the British Educational Research Association, Manchester, September 2004.

Smith E, Gorard S, Furlong J (2003). *Factors supporting high quality in teacher training: a scoping analysis.* University of Cardiff and Teacher Training Agency.

Teaching and Learning Research Programme (2003). *TLRP's future provision for capacity building: a consultation (mimeo)*. University of Cambridge, Economic and Social Research Council.

Thornton M (2003). The Education RAE 2001: are there lessons 3bs can learn from 5/5*s? *Research Intelligence,* 85, 10–15.

Tunzelmann N, Mbula K (2003). *Changes in Research Assessment Practices in Other Countries since 1999.* HEFCE.

UCET (2003). *Review of Research Assessment Exercise: Observations from UCET.* Universities Council for the Education of Teachers.

Wooding S, Grant J (2003). *Assessing Research: The Researchers' View.* HEFCE.

Workshop: enhancing adoptive parenting: a randomised controlled trial

Chair: Sharon Witherspoon, Nuffield Foundation

Dr. Alan Rushton and Dr. Elizabeth Monck, Thomas Coram Research Unit

Studies of children placed from care beyond infancy into adoptive homes have shown that many of the children have behavioural, emotional, social and educational problems and are likely to present significant challenges to the adoptive families. Government is eager to promote more children being placed from care into permanent family homes and it is important to investigate what kinds of interventions are relevant in supporting these placements and which prove to be cost–effective. Little evaluative research has so far been conducted in this field.

The presentation described the setting up of this randomised controlled trial to compare a behavioural and an educational approach in helping adoptive parents. The DfES and the Nuffield Foundation are jointly funding this three-year project. As it has only just begun, no outcome data has so far been collected.

The realities of 'What Works?' research was described. The rationale and theoretical basis of the contrasting interventions were presented and the ways in which they have been tailored to the needs of adoptive families. Ethical and practical issues in setting up the study were outlined, how the families are recruited, the randomisation process, the selection and training of the parent advisors and the choice of measures to record changes in parenting and in the children's behaviour

Workshop: what works in supporting parents?

Chair: Elaine Farmer, University of Bristol

The SPOKES project

Stephen Scott, Kings College London

Stephen Scott reported on the design and findings of his study, with Professor Kathy Sylva, on the SPOKES project, a community intervention programme in parenting and reading. This study was funded by DfES under the Supporting Parents initiative.

For the programme, 1,033 Reception and Year 1 children (aged 5 and 6) were screened for difficulties using the Strengths and Difficulties Questionnaire; from 171 eligible children, 113 entered the programme and 103 completed it. The intervention comprised a 12-week parenting programme based on the Webster-Stratton model, together with a Reading Readiness programme for parents to encourage children's reading. Considerable efforts were made to engage parents, retain them in the programme, and maintain programme integrity.

The outcomes after intervention were:

- Reduction in antisocial behaviour
- Decrease in hyperactivity
- Increase in reading age
- Improvement in parenting style.

Emotional problems, however, did not show improvement.

Stephen Scott and his colleagues are now conducting a five-year follow-up of the intervention group.

Trial of an adoptive parenting programme

Alan Rushton, Kings College London, and
Elizabeth Monck, Thomas Coram Research Unit

Alan Rushton (filling in for Professor David Quinton of the University of Bristol, who was unwell) described a randomised controlled trial of a programme to enhance adoptive parenting. Late-placed adopted children (i.e. placed at ages 3–8 years) are screened for difficulties, using the Strength and Difficulties Questionnaire with adoptive parents and social workers. Those who agree to be included are randomly allocated either to a routine service group, or to a ten-week

programme based on a cognitive behavioural or educational/illuminative approach. Progress in parenting and of the study child was measured at two subsequent intervals.

The cognitive behavioural approach, staffed by trained parent advisers, trains new parents to use behaviour management techniques, and focuses on the most pressing challenges the child is presenting. The educational/illuminative approach builds on practice wisdom to explore the particular child's past adversities, and to enhance the new parents' understanding of how these are reflected in the child's current behaviour. It enables parents to gain a better understanding of their own feelings and reactions to the child.

Both programmes are home-based, individualised and delivered on the basis of manuals specially developed for the study. They aim to engage new parents positively, valuing their strengths.

At the time of writing, 80 children from 13 local authorities had been identified as suitable for inclusion in the programme. The reasons that some new parents decline to take part were discussed. It is expected that whatever the specific outcome of the study when it is completed, the intervention manuals will prove useful to practitioners.

Workshop: Study support

Chair: AV Kirwan, Programme Director, Quality in Study Support

Evaluating the impact of Playing for Success

Caroline Sharp, NFER

The full conference paper – is included elsewhere in this publication

Study support is a much wider notion than study skills and homework clubs. It encompasses all planned learning activities that students undertake voluntarily outside the classroom, including sport, art, music and drama, as well as hobby clubs and mentoring programmes.

The extension and enhancement of study support provision, in schools, libraries and elsewhere, has been part of a sustained DfES initiative since 1997. Significant funding for this provision has been provided to schools and LEAs through the Standards Fund and via the New Opportunities Fund.

The R-EWL project: study support and disaffected pupils at Key Stage 3

Tim Barnes and Dr Ruth Rogers, Canterbury Christ Church University College

The Re-engaging with Learning (R-EWL) action research project takes the studies of Sharp and MacBeath et al. as demonstrating that study support can be effective in raising attainment and improving pupil attitudes and motivation. It seeks to understand what schools need to do in order to re-engage disaffected pupils with learning. Thirteen schools in diverse geographical settings across five LEAs are taking part. Each school identified a group of disengaged and/or disaffected Year 8 pupils (the 'labelled group'). Baseline measures have been taken of the whole year group regarding attitudes to learning and to school, reported behaviour in school, and in participation in different types of study support activities.

Initial analysis of the data indicates that:

- There are no great differences between the 'labelled group' and the rest of each cohort in attitudes to learning and to school;
- Nor are there great differences between the labelled group and the cohort in the study support activities they participate in;

- The labelled group are different in their reported behaviour; they acknowledge that they are 'off task' more, but they consider being 'on task' as important as do the rest of the cohort.

In each school, two staff members were asked to work with a 'focus group' – a subgroup of 6–10 students from the labelled group. The focus group were to work together, using a digital camera to record the positive and negative places for learning in the school. The way in which schools had undertaken and then capitalised on this dialogue with pupils varied greatly. Some had used it to:

- Identify gaps in study support provision, and create new activity clubs to build teamwork skills and develop habits of good timekeeping;
- Expand the dialogue with pupils about learning, and create new opportunities for peer education;
- Improve the fabric of the buildings.

A few schools failed to do anything.

Further analysis is required to understand the differences in schools' responses, and to discover whether membership of a focus group is correlated with changes in attitude or behaviour.

Discussion

Questions and discussion focused on two areas:

- The reasons why the context of study support may have an impact on pupils' learning;
- The implications for school structures, teachers' role, and functions of increased study support provision.

Links with workforce reform and the personalised learning agenda were also noted.

Workshop: the school as a learning community

Chair: Lesley Saunders, General Teaching Council

Factors influencing the transfer of good practice

Michael Eraut, University of Sussex and representative from DEMOS

The full conference paper – is included elsewhere in this publication

The idea of schools as professional learning communities is now central to the DfES's core principles for raising standards in teaching and learning, and to the National College for School Leadership's revised National Standards for Head Teachers. The General Teaching Council (which co-funded the Effective Professional Learning Communities project, described below) locates the momentum for positive change and improvement firmly with the capacity of the profession to renew itself through professional exchange and learning.

Both the learning communities project, and the study on transfer of good practice, are essentially concerned with 'scalability': with whether and how developments and innovations in teaching and learning can spread laterally, as distinct from vertically, through the agency of teachers and other professionals in schools. They are also concerned with 'what works': that is, what works for whom, under what circumstances, and with what spin-offs and unintended consequences. Perhaps not surprisingly, a major theme to emerge from both studies has been the importance of professional trust and personal relationships.

Both projects also deal with the kind of big 'baggy' questions that are of deep concern to policy-makers and professional practitioners, but from which researchers quite often demur because these powerful ideas carry problems with them. For example, there is often a conflation of normative definitions with descriptive characteristics, and hortatory rhetoric with observable phenomena. It has therefore been part of the research task in each case to:

- Undertake critical reviews of the literature and substantial conceptual clarification
- Design careful and attentive empirical data collection and analysis that seeks, among other things, to establish a credible 'causation trail'.

Creating and sustaining effective professional learning communities

Agnes McMahon, University of Bristol, and Louise Stoll and Ray Bolam, University of Bath

This paper (presented by Agnes McMahon, Louise Stoll and Sally Thomas) provides evidence about the processes and factors associated with the creation, sustenance and impact of

'effective professional learning communities'. The new 'collaboration' and 'transformation' agendas are being developed because the limits of centralist models of educational change have been reached. However, we know relatively little about how, under what conditions and with what consequences, professional knowledge and practice transfer between individuals and institutions – if indeed they do so.

Discussion

The following questions, which were put to the Effective Professional Learning Communities (EPLCs) research team, can be seen as issues that policy-makers and other stakeholders should ask of future research studies:

- *Was there an observable impact on performance?* Yes, this was identified from correlations between pupil outcome data and survey data, but great caution is needed in the interpretation of this, as the effect size is small.
- *Only one person in each school was asked to respond to the questionnaire; therefore is the survey data essentially based on only one person's opinion?* It was agreed that this is true and the question was debated at some length; however, the design was constrained because of the resources available.
- *What are the implications of the low response rate to the survey?* The research team outlined their considerable efforts to improve the response rate, and their post hoc comparison of respondent schools with the national population, which showed that their sample was broadly representative. It is also the case that low response rates are increasingly a problem in all social research, and the DfES is investigating the problem.
- *What are the issues with schools' self-designation of stage of development?* The research team agreed that this would indeed raise questions of validity if there was no other source of data; but the in-depth case study findings tended to confirm the survey data.

It was noted in discussion that a research project on 'learning to learn' being funded by the ESRC as part of the Teaching and Learning Research Programme is finding similar factors in relation to building learning cultures and communities in schools, and thus there is likely to be some strength in the accumulation of findings.

The Transfer of Good Practice (TGP) project attracted some discussion about current policy and practice – for example, a questioning of the emphasis on 'practice transfer' as a mode of dissemination in several national initiatives. The discussion queried the evidence base for such decisions and made reference to Robert Slavin's point about 'policy fads' [see his keynote address: *Translating research into practice on a large scale: lessons from Success for All (US)*]. The TGP research team pointed out that such policy initiatives may not be based on research at all, but rather on an assumption that successful outcomes have been brought about by strong, explicit and transferable practice in the originating schools that can be readily grasped by others. The TGP research team make a useful distinction between the transfer of knowledge and the transfer of practice, and have produced a guide for practitioners (RR615) that sets out the kinds of issues to address when planning to share practice.

The similarities with some of the findings and implications of the recently-published systematic review of the impact of continuing professional development (http://eppi.ioe.ac.uk/EPPIWeb/home.aspx?page=/reel/review_groups/CPD/review_one.htm) were also highlighted, for example, in terms of the importance of peer observation of practice and subsequent feedback/dialogue.

It was also noted that a focus on teacher-to-teacher and department-to-department transfer is probably going to be more salient than school-to-school transfer – especially since head teachers often delegate initiatives and processes for 'transfer'. Advanced Skills Teachers, for example, have professional autonomy, explicit pedagogical and subject expertise and pre-existing relationships, on all of which they can build.

The audience was very appreciative of both presentations, and the interesting methodological and substantive issues they raised.

Workshop: influences on the decision to stay in learning post-16

Chair: Trevor Fellowes, DfES

The influence of the school in the decision to participate in learning post-16

Nick Foskett, School of Education, University of Southampton

Research in 24 schools across nine LEAs involved qualitative interviews with young people in years 10, 11 and 12, and with head teachers, heads of year and heads of careers, as well as a postal survey of parents. The intent was to understand the key influences, to identify implications for schools, and to model patterns.

Four school-based factors were found to have a very strong influence on young people's decisions about post-16 education, training and career pathways:

- Whether or not the school had a sixth form;
- The characteristics of school leadership, ethos and values;
- The socio-economic status of the school's catchment area;
- The organisation and delivery of careers education guidance at the school.

In general, schools in higher socio-economic areas see themselves as developing pupils for university; while those in areas whose population is largely from lower economic backgrounds have a stronger commitment to vocational pathways. Although careers information offered at schools with sixth forms was perceived as narrower in scope, it was rated more highly by pupils and parents.

Factors within the school identified in the survey as key influences on the decision whether to stay on were:

- Presentations by Connexions and other advisers;
- Inspirational teachers;
- Good-quality careers advice/lessons;
- Enjoyable teaching;
- Work experience.

The results were analysed according to a classification of schools into different types. Student-centred schools produced a much broader range of outcomes, in part because the school linked with more enabling agencies. By contrast, school-centred schools channelled young people to the school's own preferred outcomes. The message from young people, themselves, was that they wanted the experience of a broader curriculum, and more opportunity to taste post-16 opportunities.

Transitions from compulsory education and young people not in education, employment or training: evidence from the evaluation of the EMA pilots

Sue Middleton, Centre for Research in Social Policy, Loughborough University.

The description of the experiences of young people not in education, employment or training (NEET) immediately after the end of compulsory education focused on aspects of their school lives during years 10 and 11, and assessed the sources available to them for information, advice and guidance. There have been important strides in tackling low achievement, addressing disaffection by changing the National Curriculum, reshaping formal information, advice and guidance through Connexions, and the introduction of Education Maintenance Allowances.

However, the diversity of the NEET group needs to be recognised so that policies can be tailored to particular needs.

Only 10% of the NEET group had had no aspirations or thoughts about what they would do post -16, which posed the question of why they became NEET. Key factors were:

- disagreements with their parents as to the best choice;
- poor experience of school;
- poor exam results.

Young people in NEET had a more positive view on the benefits of staying on than did those in work or training. They reported getting less advice than their counterparts who had chosen to stay on, and often their parents said that they had been unable to offer any advice, because they did not feel equipped to do so. A smaller proportion of the young people in the NEET group had had one-to-one interviews with Connexions personal advisers, compared with their equivalents who chose to stay on in education.

Workshop: synthesising research

Chair: Judy Sebba, University of Sussex

Issues in synthesising research

Wynne Harlen, University of Bristol and
Richard Andrews, University of York

Wynne Harlen outlined issues in and criticisms of educational research. For example, in 2003, the Commission on the Social Sciences suggested that social science research addresses issues other than those that are central or directly relevant to the political and policy debate. It commented that research was of questionable quality, because it employed limited procedures for gathering evidence and used opportunistic samples with low return rates, which add little to the understanding of particular events or wider issues of methodology. It was noted that communication about research findings and accessibility to it can be as problematic as the research itself, limiting the chances of the key messages being acted upon.

Wynne Harlen suggested that, in theory, these issues can be addressed through systematic reviewing.

She went on to describe briefly the stages in an EPPI (Evidence for Policy and Practice Information and Co-ordinating Centre) review and some key findings from reviews. These reviews address some of the methodological concerns by explicitly including judgements of the methodological soundness and weight of the evidence that a particular study provides, for answering the questions posed in that review. There is an important role for reviews and syntheses of research in bringing together the best evidence relating to a particular policy issue.

Some of the difficulties encountered in the systematic reviewing process were then described including identifying the research questions, synthesising results from studies with different foci and communicating findings to different user audiences. The tension was identified between reporting findings in which individual studies still show through and adopting a level of generalisation that enables real synthesis, but may be seen as going beyond the data.

Richard Andrews distinguished between meta-analysis and synthesis. Meta-analysis is a specific tool involving the statistical combination of data from different studies, leading to a quantitative summary of the findings. Synthesis involves quality-assuring and combining conclusions from different studies in order to provide a view of the best evidence to date on a particular question. In educational research, there is insufficient homogeneity between the studies to enable meta-analyses to be conducted. Much of education research is bottom-up and fails to accrue into generalisable evidence that, in turn, can provide the foundation for theory-building. Studies are likely, therefore, to be divergent in design, making it difficult to synthesise across them. He suggested that education is a field of inquiry drawing on many disciplines, rather than a discipline in itself.

Narrative synthesis is a synthesis in words rather than in numbers. It is an advanced form of summary or review that it is bound by the research question and aims for a relatively dispassionate answer or answers to that question. It is directed not so much at the audience, as at achieving the best possible account of the research being examined. The process of synthesis is very much an attempt to minimise bias.

Richard Andrews provided an example from a systematic review of the evidence on the impact of teaching grammar on quality of writing. The review had been unable to demonstrate support for teaching grammar on the basis of impact alone. However, the commentary made clear that this does not mean that the teaching of such aspects of grammar is not interesting or useful in its own right, but rather that there may be better ways of teaching writing. He reported that the experience of researchers engaging in the reviewing process is that it has an impact on the quality of their subsequent research design, methodology and reporting of research.

Discussion

Concerns were expressed that, because the model of reviewing is imported from healthcare, it attributes greater value to randomly controlled trials than to other research methodologies. Richard reiterated that the systematic reviewing he had been describing includes the full range of methodologies. It is important to distinguish between the need for more experimental work in education, and the status given to randomly controlled trials in reviews. It was noted that there is a need to present reviews in accessible forms and to extend practitioner involvement in research.

Realistic synthesis of evidence on mentoring

Ray Pawson, University of Leeds

The review starts with a hunt for theories from the literature and, having detected and developed a theory, it then looks for studies that match the theory. The review tests out and refines the theory. Case studies of individuals can be tested out against the theories. This leads to better formulation of the theories. This is the basis for 'realistic synthesis'. The review on mentoring is used here as an example, demonstrating this approach.

The realistic synthesis undertaken was of mentoring relationships. It looked at what makes for a good partnership in mentoring rather than whether mentoring works. A theory was built from the literature relating the type of mentoring relationship (for example, coaching or advocacy) to the type of task (for example, direction-setting or befriending). Ray concluded that there are pathways of youth mentoring that involve different agencies such as family and peers, youth and community services, education, training and work services and career services, which all engage in fire-fighting in the mentoring process. He also identified the importance of confidence building, resilience building and trust building in the mentoring processes.

Ray presented some methodological conclusions about realistic synthesis, which identified what might be expected from it. He suggested that it did not provide verdicts, procedural uniformity or novice input. It did, however, provide theory development and refinement, advice on implementation and targeting, realistic expectations on programme goals and assessments of research quality – but only in relation to the theories under test. Finally, findings from realistic synthesis are open to challenge.

Discussion

The discussion covered the key features of realist synthesis that are similar and those that differ from systematic reviews. There was also debate about how practitioners engage in realist synthesis.

Further information on Ray Pawson's approach is available at:
www.ccsr.ac.uk/methods/publications/RMPmethods2.pdf

Workshop: research designs for complex social interventions

Chair: Dr Barbara Maughan, Institute of Psychiatry

Workshop led by Dr Stephen Scott, Institute of Psychiatry, King's College London

There have been relatively few rigorous evaluations of educational interventions; this is in contrast to the health field, where in the last 25–30 years randomised controlled trials (RCTs) have become far more widespread and have on occasion turned up surprising results. Thus, for example, 'common sense' has not always found to be sustained once evaluated rigorously by an RCT.

For example, the Cambridge–Somerville project was thought in its time to be a state-of-the-art intervention for antisocial teenage boys who were failing at school. A male mentor social worker visited the home once or twice a week; the boys were given remedial teaching where necessary, and took part in constructive after-school activities such as sports. They were also sent on summer camps where they could take part in sporting activities and so forth. A control group were given no particular extra care. Evaluation two years later showed that the intervention group were smoking more cannabis, committing more crime on self-report, and had had more police contact and arrests. Subsequent qualitative research has suggested that it was the summer camps that did the damage: the young people learned new tricks from each other, and came to see involvement in antisocial behaviour as a way of gaining prestige in the eyes of their peers.

In a similar vein, an RCT at the Oregon Social Learning Centre compared parent training alone, parent training and a peer group, and a peer group alone as interventions for antisocial teenagers. Again, those in the peer group did worse, for similar reasons.

In the preventative arena of early enrichment programmes, one or two headline projects (such as the Perry/Highscope project) have had good effects, even when participants were followed up in adulthood. However, several others have had little effect, or have failed to be replicated by practitioners other than the originators of the programme. Findings from the current SureStart evaluation have, perhaps, been less hopeful than was originally expected: 24 out of 25 variables did not show any change; the only variable that improved was 'maternal acceptance of the child' (NESS, 2004).

The reasons why some replications in real life have failed to show positive effect are complex, but two major factors stand out. The first is that evidence-based interventions are not used, and are not delivered with skill. There is an increasing literature on the importance of treatment fidelity, meaning that if people stick to a particular programme with skill, outcomes are much better. Thus, multi-systemic therapy for antisocial behaviour has failed to show good outcomes when it is not done with great skill, but is very effective when done skilfully. Likewise, the Institute of Education's Effective Provision of Pre-School Education (EPPE) study (http://www.ioe.ac.uk/cdl/eppe/pdfs/eppe_brief2503.pdf) has shown that quality of nursery

provision for young children in the UK has a big effect on whether their attainment is enhanced before they arrive at primary school; where the quality is low, however, early childcare can be positively deleterious.

The second factor is failure to engage hard-to-reach families, and to put in place special measures for those who fail to turn up.

In general, RCTs provide much stronger tests of the effectiveness of interventions than assumptions drawn from longitudinal studies, which may face undetected confounding factors. For example, parental reading with children has been associated with better outcomes in longitudinal studies. In part, however, this finding might reflect parental or child characteristics: more able parents may be more likely to read with their children, or more able children may enjoy reading with parents. Reading with parents may improve outcomes, but other effects may be operating as well.

This question was taken up in Stephen Scott and Kathy Sylva's recent study on Supporting Parents on Kids Education in Schools (SPOKES), an RCT of parenting groups and reading support held in schools for reception and year one children (*see the separate report elsewhere in this publication*). After a year, the children were less antisocial; their reading age had improved by six months; their level of hyperactivity was less; and parenting – assessed independently on videotape – had improved. The project is an example of a tightly controlled intervention that is likely to lead to better long-term outcomes; however, this needs testing through further follow-up studies.

It was strongly proposed that funders should invest in more research into intervention studies. There are many questions that need answering, such as: whether the literacy and numeracy hours are effective (they should be tested against alternative teaching methods); the degree of training teachers need to improve their performance; and which classroom behaviour management techniques are best in helping teachers gain control (the commonest reason given by teachers for leaving teaching is not pay, but children's antisocial behaviour).

Dr Stephen Scott next took workshop delegates through the design and ethical issues that would arise when devising a multi-centre RCT to evaluate whether special schools, separate from the mainstream, were better than inclusion in mainstream schools for children with antisocial behaviour and dyslexia.

References

The National Evlauation of Sure Start (NESS) (2004) *The impact of Sure Start local programmes on child development and family functioning: a report on preliminary findings.* London: Birkbeck University of London.

EPPE study from the Institute of Education.
http://www.ioe.ac.uk/cdl/eppe/pdfs/eppe_brief2503.pdf

Workshop: lifelong learning

Chair: Hugh Tollyfield, Deputy Director Skills for Success, DfES

Evaluation of the Employer Training Pilots

Jim Hillage, Institute for Employment Studies

The full conference paper – is included elsewhere in this publication

Raising employers' demand for skills

Geoff Mason, National Institute of Economic and Social Research

In the 2003 National Employer Skills Survey (ESS), only 4% of employers reported external recruitment difficulties in relation to skills shortage vacancies, and only 22% reported internal skills gaps. Cross-country comparisons suggested that a relatively large proportion of British output was concentrated towards the more standardised, less complicated, lower end of the quality spectrum, leading to a relatively low demand for skills.

Such a simple categorisation of product strategies was problematic, but empirical evidence showed:

- Strong product strategy–skills association at establishment level;
- Both positively related to foreign competition;
- Marked variations within and between industry sectors;
- Sector upgrading of product strategies, leading to polarisation of skill levels;
- Low-end product strategies significantly associated with decline in sales, and not sustainable in the medium term; but many companies in low value-added production could survive comfortably in existing markets.

However, ESS 1999 had indicated that 40% of private sector companies were looking to take at least some steps up the value chain, most of which would lead to additional skills requirements. A problem highlighted by the Skills Task Force was that human resources practice often lagged behind company product strategies, disguising 'latent' skills gaps. The conclusion was that companies making profits from low value-added production would be unlikely to respond to government exhortation or encouragement to move up-market, and that any rate of change would be gradual. Business support policy should, therefore, be based on segmentation and should focus on:

- Companies seeking to move up-market, but constrained from doing so;

- Companies needing to move up-market, but too immersed in day-to-day problems to do so effectively.

Identification of such companies would be the challenge.

Discussion

The discussion that followed tested out the main points through reference to particular sectors (such as hospitality and construction), and sought to distinguish between generic employability skills, such as problem-solving and communications, and more specific vocational or technical skills.

There was a recognition that training tended to provide workforce productivity benefits beyond those associated simply with skills acquisition, and that the capacity of employers to deploy workforce skills effectively was an issue associated as much with low-value product strategies as with the acquisition of the skills by the workforce.

Overall, it was felt that the short-term nature of business planning in many companies led to a failure to identify skills gaps.

Workshop: research communication and utilisation in policy and practice

Chair: Audrey Brown, DfES

Time to be bold: mainstreaming models of educational research that respond to user needs

Marilyn Leask, Teacher Training Agency

The full conference paper is included elsewhere in this publication.

Using research in practice: what's known and what's needed

Philippa Cordingley, Centre for the Use of Research and Evidence in Education

This presentation focused on practitioner-based research, in particular the work of the National Teacher Research Panel (NTRP). The issue of the accessibility of findings was identified, as well as the need to produce convincing findings that teachers care about or have to do something about. The discussion of experiences in undertaking systematic reviews for the Evidence for Policy and Practice Information and Co-ordinating Centre included the following points:

- It is important to have systematic, user-orientated syntheses;
- 'Stories' of the findings would be helpful in making reports more accessible;
- There is a need for practitioner-researchers to enhance practice;
- There is also a need for practitioner input when setting priorities.

Closing the gap between practitioners and research

David Mallows, National Research and Development Centre for Adult Literacy and Numeracy

This presentation on the work of the National Research and Development Centre for Adult Literacy and Numeracy looked at why teachers should be helped in gaining access to, understanding and using research. It considered how teachers consume research, for example:

- As questioners (for instance, when reading articles, attending conferences etc.);

- As adopters (for example, through involvement in practitioner forums to consider the implications of research for practice);
- As proactive consumers (involvement in research as part of continuing professional development);
- As producers (sustained involvement in their own or external research).

It was concluded that practitioners should be engaged throughout the research process, not just at the end.

Discussion

The following issues were raised in the discussion:

Will summarising research for teachers result in researchers failing to consider this, as someone else will do it? It was agreed that:

- Dissemination should be part of the research process;
- Reports need to be organised more systematically, with different levels of reporting;
- It is important to consider the audience: for policy, a final project report or summary is appropriate; for practice, a vision of how it works in the classroom is useful;
- This question was not a criticism of the research community, but of funders, as dissemination is often not factored into projects.

How can the policy-makers who are responsible for research projects ensure that findings are useful for practitioners? It was agreed that:

- It is essential to get practitioners involved early in the research;
- Reports need to give more weight to discussion of the implications of the findings. It might be worth drawing in fresh people to consider the implications of findings;
- However, DfES research is often commissioned to inform policy; therefore, impact on practice often lags behind.

Many researchers would not agree that the role of research is to improve practice; instead this 'muddies the purity of the enterprise'.

- The response emphasised that funding is crucial, and shapes the focus and design of research.

Personnel changes on externally funded projects, particularly long-term projects, often lead to changed requirements, or to reports presented to people who weren't originally involved in the project. It was agreed that:

- It might be helpful for policy-makers to get a better understanding of the research process, and equally that researchers should try to understand how policy-makers work;
- It is not possible to influence the way the Department is organised;

- It would be good to get research upstream of policy – we need to think about what the big issues will be three years hence.

The Teaching and Learning Research Programme (TLRP) raised similar issues several years ago.

- There was some agreement that the TLRP programme should follow up the funding in schools;
- However, there was some criticism that the model is not sufficiently coherent and can lead to a wide range of projects;
- It was also claimed that there is little research on teacher education (as the bids were not very strong).

Workshop: Cross-cutting

Chair: David Hopkins, DfES

Success for All in England: Implementation and outomes of a comprehensive literacy reform for primary schools

Robert E Slavin, Johns Hopkins University, and Judith Wordsworth, Sucess for All Foundation, UK

The full conference paper is included elsewhere in this publication

Simple principles for analysing complex problems

Ray Pawson, University of Leeds

One of the most dramatic changes in public policy in recent years has been the rise of complex, multi-objective, multi-site, multi-agency, multi-subject programmes. The reason for all this multiplication is clear enough. The roots of social problems intertwine. A health deficit may have origins in educational disadvantage, labour market inequality, environmental disparities, housing exclusion, differential patterns of crime victimisation, and so on. By twisting the kaleidoscope and beginning this little list at any other point, one can make more or less the same argument about any social problem. All have multiple sources and a location within cycles of deprivation. Decision makers have, accordingly, begun to ponder whether single-measure, single-issue interventions might be treating just the symptoms, at best inducing short-term gains before losing their grasp on subjects who then sink back into deeper-seated disadvantage.

Such thinking has inspired the upsurge of a new breed of 'super interventions', some examples of which follow. The hot-spots of social deprivation are often highly concentrated with the result that these interventions characteristically take the form of area-based initiatives (ABIs). Programmes like *Health Action Zones and Education Action Zones,* and are thus located, across the country, in 'sink estates', 'inner-city neighbourhoods', 'social flight zones' and so forth. Another key feature, exemplified perhaps by Sure Start and the Connexions service, is the idea of joining up existing welfare services in order to begin to match provision to the complexity of the underlying problems. The Connexions Service takes this idea to the limit, involving as it does a massive assimilation of Careers Advice and Youth Work professionals. The third and final feature highlighted here is the long chain of command. This applies to all the aforementioned examples but I might single out the *New Deal for Communities* as a champion specimen, with its particularly dispersed implementation chain stretching back and forth from Whitehall to regional government to local government to local agencies and to community leaders.

I make no further comment on the extent, or indeed the wisdom, of the shift to such

agglomerated interventions. My purpose is view them from the perspective of the researcher and ask ‘how?’ or, perhaps better, ‘how on earth?’ can they be evaluated.

Copies of this publication can be obtained from:

DfES Publications
P.O. Box 5050
Sherwood Park
Annesley
Nottingham
NG15 0DJ

Tel: 0845 60 222 60
Fax: 0845 60 333 60
Minicom: 0845 6055560

Published by the Department for Education and Skills

ISBN: 1 84478 463 0
Ref No: CR2004

www.dfes.gov.uk/research