

WHAT WORKS?

Helen Roberts¹
Professor of Child Health
City University, London

Abstract

“What works?” is a fundamental question for policy makers, practitioners and service users. As well as asking what works, and how to implement it, we also need to think about what does not work, and how to stop it. If, as many of us believe (and as there is good evidence to demonstrate), the most effective time to make a difference to outcomes is in childhood, then it is likely that this is also a time when we need to be careful of doing damage. Intervening in children’s lives is not just a research, policy and practice issue for those of us at the supply end. It is also a rights issue for children and young people. Those of us in the evidence-based arena who have been pushing policy makers and practitioners to adopt evidence-based practice need to be careful that we do not sell the “What works?” agenda as a simple way to solve problems. Social and educational interventions are complex and are capable of doing as much or even more harm than medical ones. Drawing on examples that relate to children and young people, this paper will suggest that the public and NGO sectors need to invest more heavily in the “D” part of R&D, and that we need to stimulate demand for research-based interventions. In the world of childbearing and HIV/AIDS, services in the United Kingdom have been transformed by powerful alliances between the evidence and the “user” lobbies. What would the world look like for young people if those having difficulties with their education, in trouble with the law, or in the care of the state, were to ask what the evidence is behind the services they are offered?

1 Acknowledgements

I am grateful to my colleagues in the ESRC-funded What Works for Children initiative (www.whatworksforchildren.org.uk), whose work has informed mine, and to other colleagues in the Child Policy and Research Unit of City University. My observations on Charles Booth and the Fabians were drawn largely from the Charles Booth Archive at the London School of Economics (LSE), and conversations with Rodney Barker, Professor of Government at LSE. My colleague Mark Petticrew allowed me to draw on our forthcoming book on systematic reviewing (Petticrew and Roberts 2005) and to reproduce a diagram from our paper “Horses for courses” (Petticrew and Roberts 2003). I am particularly grateful to the Ministry of Education and to Martin Connelly, Senior Manager of Education Management Policy in the Ministry, for inviting me to the conference.

INTRODUCTION

It is important not only to know what we know, but that we know what we do not know. (Lao-Tze, Chinese philosopher)

As we know, there are known knowns. There are things we know we know. We also know there are known unknowns. That is to say we know there are some things we do not know. But there are also unknown unknowns, the ones we don't know we don't know. (Donald Rumsfeld, United States Defence Secretary²)

As this conference, the one that preceded it and the considerable work done by public and social policy, practice and academic colleagues in New Zealand make clear, "What works?" is a fundamental question for policy makers, practitioners and service users. Indeed, it is a fundamental question for all of us. "What works?" and "Does this work better than that?" or "Is it better to do something, or do nothing at all?" are questions we can relate to in making small decisions as well as big ones, both in our everyday lives and in our professional activities.

New Zealand has been a leader in this field, and has addressed some important questions. The best evidence syntheses on the Ministry of Education website³ provide good pragmatic examples of evidence use, matching research design to the particular questions asked. SPEaR itself produces helpful practice guidelines on the policy cycle.⁴ In 2002, the New Zealand Treasury asked what we know about the effectiveness of early years services (Annesley et al. 2002). New Zealand academics have asked about the effect of incomes on outcomes (Connor et al. 1999). Your country questioned the evidence for going to war in Iraq, the evidence for which has been the subject of so much discussion in the United Kingdom since (Butler 2004).

At present, practitioners, parents and children and young people looking for good research evidence on common problems will often find the evidence cupboard disappointingly bare. Where it is not bare, the outcomes measured will not always be those that seem most important to these players. It is refreshing that in New Zealand, the learning outcomes for schooling derived from best evidence syntheses include not just achievement and skills, but wellbeing and whānau spirit, as well as respect for others (rangimārie), tolerance, non-racist behaviour, caring or compassion (aroha), diligence and hospitality or generosity (manaakitanga) (Ministry of Education 1993:17).

2 Feb 12 2002, Department of Defence News Briefing http://www.defenselink.mil/transcripts/2002/t02122002_t212sdv2.html (Petticrew and Roberts 2005).

3 www.minedu.govt.nz

4 www.spear.govt.nz/SPEAR/documents/best-practice/background-paper-series-1.doc

The role of formal research evidence in the formation of social and public policy, at least in the United Kingdom, has a relatively short history. There has probably been a role for “intelligence” for as long as countries have had foreign policies, and this is likely to have been a mixture of sound research, know how, and something a bloke said in a bar (or possibly at a cocktail party if we are talking about the diplomatic corps). Much of this will have involved matters such as trade, military competition and conflict. The relative distance of the majority of the expatriate British from local populations, so well described by E.M. Forster, George Orwell and Doris Lessing suggests that the learning that could have been had from different countries’ or indigenous peoples’ ways of doing things was not a core part of the task of those who spent time overseas.

In the 19th century there were certainly a number of “investigations” which bear the hallmarks of good qualitative research. One example of “listening to children” can be seen in the report of Andrew Doyle to the local government board at Whitehall on the emigration of pauper children to Canada. He spoke to many children, some of whom had had happy experiences, some less so. Of course if the United Kingdom’s policy of sending orphan children to New Zealand, Canada, Australia and Rhodesia (as it then was) had been subjected to a traditional evaluation, the report may well have said that the children looked well cared for, the food, transport and child-to-adult ratios were adequate, and the youngsters, asked about their experiences might well have said: “It’s OK here” (or whatever the 19th century equivalent was). Doyle managed to collect the other side of the story, often untold even in current policy evaluations. Of the sea voyage, he was told: “we sicked all over each other,” and a child patiently explained the meaning of adoption: “Doption sir, is when folks get a girl to work without wages” (Doyle 1875).

There were other investigations which we would certainly view as research evidence now. Charles Booth’s work in the United Kingdom in the late 19th century charted inequalities in health and welfare. Between 1886 and 1903 Charles Booth (and Mary Booth, about whom we hear rather less) worked on the life and labour of the people of London – 17 volumes in all – working with, among others, Beatrice Webb. He lodged with working-class families, and observed both the inequalities and the resilience and coping strategies of families living with disadvantage that we continue to learn from today. Booth wrote:

The children in class E [“... above the line of poverty”], and still more in class D [“... the poor”], have when young less chance of surviving than those of the rich, but I certainly think their lives are happier, ...always provided they have decent parents. It is the constant occupation, which makes the children’s lives so happy. They have their regular school hours. They have for playground the back yard, [or] the even greater delights of the street. Let it not be supposed,

however, that on this I propose to base any argument against the desire of this class to better its position. Very far from it. [...] the uncertainty of their lot, whether or not felt as an anxiety, is ever present as a danger. (Booth 1902: 159-60)

Booth was not only a keen social observer and essayist. He produced maps descriptive of London poverty which would be the pride of a 21st century social geographer, with investigators collecting data while accompanying policemen on their rounds. The Fabian Society, of which Booth was a member, was founded in 1884, and committed to gradual social reform. Also active in the Fabians were Beatrice and Sidney Webb, who saw research very specifically as informing social policy in a linear way which we might now consider unrealistic. The London School of Economics, now part of the University of London was founded by the Webbs in 1895 to inform social and public policy. Its motto, “*rerum cognoscere causas*” – to seek the cause of things, remains one key question at the heart of evidence-informed policy today. Another is what to do when we do know the cause of things, and a further one is to learn the consequences of things.

While there were huge scientific advances in all fields in the 20th century, some benign or better, some less so, the project of uniting research policy and practice became less central as a key intellectual interest. Indeed, even today there are those who recoil at the implied positivism of research-informed policy and practice. Although the social sciences and agriculture had been in the forefront of the development and use of research evidence, and in particular experimental methods, it was the establishment of the worldwide Cochrane Collaboration in health care in 1993 that gave fresh impetus to the use of sound research evidence to inform practice. In the late 1990s, and to some extent prompted by both social scientists and clinicians involved in Cochrane, the Economic and Social Research Council (ESRC) in the United Kingdom made available funding for a centre, and a number of “nodes” to work on research evidence. One of these ESRC-funded “nodes” is the Research Unit in Research Utilisation, whose co-director, Sandra Nutley, spoke at the 2003 Social Research and Evaluation Conference (Nutley 2003). Another is *What Works for Children?* which is a collaborative project between City University, the University of York and Barnardo’s.

In this paper, I am drawing on work we have done as part of the *What Works for Children* enterprise and on joint work with Mark Petticrew from the University of Glasgow on ways of answering different kinds of research question, and of carrying out systematic reviews (Petticrew and Roberts 2003, Petticrew and Roberts 2005). In order to illustrate some of the problems (and steps towards solutions) in the world of evidence-based policy and practice, I am going to:

- briefly describe some of the work on the What Works for Children project

- describe some of the ways in which systematic reviewing can help move us towards making sense of large volumes of research
- discuss the gap between the “R” and the “D” of R&D in much academic work
- speculate on the emancipatory potential of the fuller involvement in the evidence agenda of end-point users of services or policies
- sound a note of caution on promising too much from the evidence agenda without sufficient investment, and the difficulties that can arise when the research evidence suggests that something does not work

WHAT WORKS FOR CHILDREN?

Academic life tends to be about the generation of new knowledge. Scholarship and primary studies are at the core of university activity, and rightly so. But for those of us working on the boundaries between research, policy and practice there is another agenda, also with a scholarly component, which involves dissemination and implementation. *What Works for Children* focuses on sharing and implementing research evidence rather than generating “new” evidence. With a population group, children, rather than a discipline or a single profession as the core unifying factor, our work covers a range of disciplinary and professional interests, emphasising interventions to address children’s wellbeing in its broadest sense.

Successful research impact has been identified as depending on more than dissemination (Barnardo’s R&D 2000, Nutley et al. 2002), and facilitation may play a key role in the implementation of evidence in practice. This is by no means always the case, however, and a single primary study or review can have massive impact if the findings say the right things to the right people at the right time. In Morton Hunt’s influential work on the way in which science “takes stock”, for instance, he describes how Professor Eric Hanushek wrote an article at the start of the Reagan administration in 1981 maintaining that empirical studies showed that increased spending on education did not increase student achievement (Hanushek 1981). Hunt writes:

This startlingly counter-intuitive finding was catnip to conservatives [and] Hanushek soon became an expert witness for the defence at hearings ... against school boards ... accused of miserly budgeting. (Hunt 1997:54)

Since we did not have a dramatic piece of work likely to attract this kind of attention, we proceeded with the evidence-based notion that facilitation makes a difference, and on the basis of this, our *What Works for Children* project incorporates a post of development or implementation officer, putting the “D” into R&D. She supports practitioners and service planners, and both she and the university researchers are exploring the levers and barriers to the use of research in a real world setting. Alongside this, the research team at City University has produced a range of tools for

practitioners; an *Evidence Guide*, *Evidence Nuggets* (which are summaries of evidence, including costings and examples), and research overviews. These and other resources are on the *What Works for Children* website (www.whatworksforchildren.org.uk).

Rather than describing in detail a project that is fully documented on our website, and well described by my colleague Kristin Liabo in this journal (2005), and elsewhere in published papers (Roberts et al. 2004, Liabo et al. 2003, Lucas et al. 2003, Stevens et al. 2005), I want to draw on our project for examples of the complexity of finding, appraising and where appropriate using research evidence – or, put more simply, the evidence of what works – which relate to children and young people. I want to argue in this paper that despite some excellent examples of cross-professional and cross-disciplinary working, such as the work of Davies et al. (2002) and Wood and Kunze (2004) in New Zealand, for instance, the public sector needs to invest more heavily in the “D” part of R&D. Moreover we need to stimulate greater demand for, as well as a steady supply of research-based interventions. In the world of childbearing and HIV/AIDS, services in the United Kingdom have been transformed by powerful alliances between the evidence lobby and the “user” lobbies. What would the world look like for children and young people if those in trouble with the law, or in the care of the state, were to ask what the evidence is behind the services they are offered?

FINDING OUT WHAT WORKS

- Is water fluoridation effective in reducing dental caries in children?
- Do children learn better in small classes?
- Can young offenders be “scared straight” through tough penal measures?
- Can the steep social class gradient in fire-related child deaths be reduced by installing smoke alarms?

As anyone attending this conference will be well aware, finding out “what works” is no mean feat. Many evaluations do not have the methods or the tools to demonstrate whether something has had the intended outcomes, though they can often give us really good information on the processes that underpin a successful service. And if we cannot get people in through the door to use services (and keep them there), then no service is going to have a chance of proving its success.

To give a practical example of problems that may arise with evaluation: suppose someone planning services reads a high-quality study showing that for children with a poor start in life, having parents who take an interest in their children’s education seems to lead to better outcomes in school and later on. In the light of this, a school-based discussion group to encourage parental support for children’s education is set up. Suppose that the programme runs for eight weekly sessions, with before and after questionnaires. It is clear from the “after” questionnaires that it is considered by both

workers and parents to have been a success (though there are no data on those who dropped out, or who did not attend in the first place). Their judgement is based on the self-reports of the parents, the workers' observations of parents' increasing confidence in the school setting, and improvements in the apparent attention span and behaviour of the children (who have been cared for in an after-school group during the parents' discussions). The work is carefully carried out and gets good local publicity. The researchers present the work well, and it is picked up for a national early morning news programme. Before asserting that "discussion groups work" we need to be as certain as possible that in this particular case:

- improvements *have* taken place
- they have been brought about as a result of the discussion group.

It is difficult to do this if we do not have mechanisms for ruling out competing explanations, such as the following.

- The parents might have become more involved in the school in any case with the passage of time. There is evidence from other fields that many problems improve spontaneously over time in two-thirds of cases (Rachman and Wilson 1980). This provides a reason for having a "no treatment control", so that we can compare the effects of our intervention with the passage of time.
- The children might have become more settled through spending time with skilled after-school club workers.
- Other external factors might be responsible for changes, such as improved income support or additional help from social services.
- The perceived improvement in the parents might be due to their having learned the "right" things to say in the course of the intervention, having been asked the same kind of questions at the beginning and end of the programme, and having become familiar with the expectations of the workers.
- The parents who stayed the course might have been highly motivated and would have improved anyway. Alternatively, parents who dropped out might have done just as well as those in the programme. We simply don't know (adapted from Macdonald and Roberts 1995).

What is more, we also need to know what other studies have shown, and how good those studies are. Maybe this study showed it "worked" because of the play of chance, and not because the intervention really was generally effective. Maybe other studies show that it does not work, or that it only works in some types of setting, or for certain types of parent or child. After all, we are well used to seeing research studies reported in the media and elsewhere that show that something works one year, only to be contradicted by a different study (or a different researcher) the next. What happens if we take the results of all these studies together? Will we still conclude that discussion groups "work"? Or, having seen all the relevant evidence, will we conclude the

opposite? Or remain uncertain? These problems demonstrate the difficulties that arise whenever we plan to do something that we claim produces clear outcomes, whether good or bad.

This example also illustrates the not particularly earth-shattering observation that some research methods are better than others for answering different kinds of question. In effect, we need the right horse for each particular course. A particular cause of upset, adversarial sabre rattling and misrepresentation of the views of those one disagrees with has been the hierarchy of evidence, initially developed by the Canadian Task Force on the Periodic Health Examination, and subsequently adopted by the United States Preventive Services Task Force to help decide on priorities when searching for studies to answer clinical questions (see Table 1).

Table 1 An Example of the “Hierarchy of Evidence”

Type of Evidence
<ul style="list-style-type: none"> • Systematic reviews and meta-analyses • Randomised controlled trials with definitive results • Randomised controlled trials with non-definitive results • Cohort studies • Case-control studies • Cross-sectional surveys • Case reports

With growing interest in the effectiveness of social interventions, a single hierarchy of methods has become unhelpful as well as divisive, and at present certainly misrepresents the interplay between the question being asked and the type of research most suited to answering it. For this reason, a matrix, or a typology, may be a useful construct. As Table 2 illustrates, different research methods are, after all, more or less good at answering different kinds of research question. A randomised controlled trial, well conducted, can tell us which kind of smoke alarm is most likely to be functioning 18 months after installation, but it cannot tell us the best way to work effectively with housing managers to make sure smoke alarms are installed effectively and cost effectively, and that the households of the most vulnerable tenants are included. The obstacles and levers for the uptake of research findings are likely to be understood through methods different from those usually found at the top of the evidence hierarchy. It may therefore be more useful to think of how one can best use the wide range of evidence available – and particularly to consider what types of study are most suitable for answering particular types of question.

Table 2 An Example of a Typology of Evidence (for Social Interventions in Children)

Research question	Qualitative Research	Survey	Case Control Studies	Cohort Studies	RCTs	Quasi-experimental Studies	Non-experimental Evaluations	Systematic Reviews
Effectiveness								
Does this work? Does doing this work better than doing that?				+	++	+		+++
Effectiveness of service delivery								
How does it work?	++	+					+	+++
Sallience								
Does it matter?	++	++						+++
Safety								
Will it do more good than harm?	+		+	+	++	+	+	+++
Acceptability								
Will children/parents be willing to or want to take up the service offered?	++	+			+	+	+	+++
Cost effectiveness								
Is it worth buying this service?					++			+++
Appropriateness								
Is this the right service for these children?	++	++						++
Quality								
How good is the service?	++	++	+	+				+

Source: Petticrew and Roberts 2002 (adapted from Muir Gray 1997)

A related problem lies in the stark use of the term “evidence”. It is not uncommon for discussion papers to use the terms “evidence”, “evidence based” and “hierarchies of evidence,” while avoiding any discussion of what sort of evidence they are advocating (or rejecting). For epidemiological questions relating to “real world” risk factors which are not amenable to randomisation (e.g., does passive smoking in the home cause cancer in later life to children exposed to cigarette smoke?) a particular sort of data are required, with prospective cohort studies at the top of the hierarchy. Qualitative studies, expert opinion and surveys, on the other hand, are likely to have crucial lessons for those wanting to understand the process of implementing an intervention, what can go wrong, and what the unexpected adverse effects might be when an implementation is rolled out to a larger population. A different sort of hierarchy is again implied.

Overall, information on both outcomes and processes is of value. Knowing that an intervention works is no guarantee that it will be used, no matter how obvious or simple it is to implement. For example, it is around 150 years since Semmelweis showed that handwashing reduces infection, yet healthcare workers’ compliance with handwashing remains poor. Even the most simple, cost-effective and logical intervention fails if people will not carry it out. The British government is currently exercised by the problems of MRSA (Methicillin-resistant *Staphylococcus Aureus*), and a number of solutions are being explored. While it may have strong common sense value for a patient to say to a doctor or nurse, “Have you washed your hands?” it is potentially deeply offensive to cast doubt on the personal cleanliness habits of others. To do so from the vulnerable position of a hospital bed may be particularly difficult, and possibly unwise.

An example of an area where policy and research have been evolving in tandem is in the growing use of child public health interventions, which can be effective in both the immediate and the longer term in improving outcomes for children (Roberts 1997, Glass 1999, Glass 2001). Highscope, Headstart, parenting education, home visiting and mentoring provide examples of well-designed programmes that have been the subject of robust evaluations, some of them using complex randomised controlled trials (Schweinhart and Weikart 1993, Barlow 1999, Webster Stratton et al. 1989, Olds et al. 2002, Tierney et al. 2000, Grossman and Tierney 1998, Dubois et al. 2002). But “parent education”, “home visiting” and mentoring, as many of their proponents and evaluators would agree, largely remain black boxes with a great many unanswered questions about what exactly the intervention involves, and *how* it works.

Meanwhile, a climate has been created in a number of countries, including my own, where it is widely held that these interventions “work” and national programmes have been established. The questions of who delivers the service, the kind of young people who might benefit, and the content of services likely to be effective can be lost in the

drive to get the show on the road. These programmes can gain momentum because they have strong face validity. They look like the sort of things that should work, our “gut” feelings tell us that they will work, and we want them to work. Not only may this result in premature roll-out on the basis on insufficient evidence or a simplistic interpretation of the evidence, but it may then be difficult to stop or change direction after programmes have been launched.

Take an example of a big social problem – anti-social behaviour in young people (Scott 1998) – and an intervention – mentoring – where a sound meta-analysis (Dubois et al. 2002) has demonstrated benefit. Anti-social behaviour in young people is a problem for families, for the young people themselves, for the police, for communities, and for politicians. This makes finding a solution a political as well as a therapeutic imperative – a potent driver to “do something”. Mentoring is non-invasive and medication-free. It is easy to see why it might work, and why it is attractive to politicians and policy makers. In February 2003, Lord Filkin, then a minister in the Home Office, announced £850,000 of funding for mentoring schemes in England:

Mentors can make a real difference to ... some of the most vulnerable people ... and help to make our society more inclusive. There are ... excellent examples of schemes which really work.⁵

There was equal enthusiasm on the other side of the pond. In his state of the union address in January 2003, President Bush announced plans for a \$450 million initiative to expand the availability of mentoring programmes for young people. This included \$300 million for mentoring at-risk pupils and \$150 million to provide mentors to children of prisoners:

I ask Congress and the American people to focus the spirit of service and the resources of government on the needs of some of our most vulnerable citizens – boys and girls trying to grow up without guidance and attention and children who must walk through a prison gate to be hugged by their mom or dad.⁶

WHAT DOES THE RESEARCH SHOW?

A problem with interventions that become politically attractive, and to which large amounts of cash are attached, is that research may be used for support rather than illumination. There is indeed robust research that indicates benefits from mentoring for some young people, for some programmes, in some circumstances, in relation to some outcomes. There are also good descriptive evaluations that suggest that those young

5 www.homeoffice.gov.uk/docs/capital_mentoring_grants.html [Accessed November 12th 2004]

6 www.cnn.com/2003/ALLPOLITICS/01/28/sotu.transcript/ [Accessed 12-11-2004]

people who stay on in programmes are inclined to report favourably on the experience (St James-Roberts and Singh 2001, Tarling et al. 2001).

As part of our *What Works for Children* project, we reviewed the evidence on whether one-to-one, non-directive mentoring programmes targeted towards young people at risk of, or already involved in, offending can improve behaviour. We then contacted David Dubois in Chicago who, with colleagues, has authored the most complete meta-analysis on youth mentoring. He reviewed our work, pointed us to further evidence and with us co-authored a publication on mentoring in the *British Medical Journal* suggesting that it is less than 100% effective for all young people at all times. In effect, on the basis of existing reviews we concluded that research on mentoring programmes does not provide evidence of measurable gains in outcomes for mentees who enter programmes as a result of, or to solve problems of offending, truanting or involvement in other anti-social behaviours.

In fact it looks from the evidence as if mentoring programmes for vulnerable young people may have a negative impact, and adverse effects associated with mentor-mentee relationship breakdowns have been reported (Grossman and Rhodes 2002). Worryingly, a 10-year follow-up study of one well-designed scheme found that a sub-group of mentored young people, some of whom had previously been arrested for minor offences, were unexpectedly found to be more likely to be arrested after the project than those not mentored (O'Donnell et al. 1979) On the basis of findings such as these, we concluded rather cautiously that non-directive mentoring programmes delivered by volunteers cannot be recommended as an effective intervention for young people at risk for, or already involved in, anti-social behaviour or criminal activities.

We were not suggesting that mentoring cannot work. There are many different kinds of mentoring and some show better evidence of effect than others. Our current state of knowledge on the effectiveness of mentoring is similar to that of a new drug that shows promise, but remains in need of further research and development. There is no equivalent of the National Institute of Clinical Excellence (NICE) or the Food and Drugs Administration (FDA) for social interventions. If there were, no more than a handful of programmes might have realistic hopes of qualifying. And even then, it would have to be acknowledged that a full understanding of the safeguards needed to ensure that young people are not harmed by participation is lacking. This observation was picked up by Mark Fulop, the Director of the National Mentoring Center in the United States, as part of a mentoring exchange digest (www.nwrel.org/mentoring). Fulop has a rather different perspective:

Mentoring is not a “drug” or a “treatment” that needs FDA approval. Mentoring is a process of community-building and community organization. If we have to measure with a microscope to see if youth mentoring is making a difference then we are not making enough difference in the first place. Every mentoring program is a living lab that is telling a story each and every day but the story is not whether Johnny is skipping few classes or that Suzie is now doing her homework. The outcomes of mentoring need to be measured at the community level. Collectively, is our community holding our children closer to our hearts? (Mentor exchange digest, 5.3.2004).

Fulop makes the important point that social value judgements are involved in the outcomes we choose to measure, and clearly, the amount of investment a community makes in its children could be a useful and important outcome. While holding children close to our hearts might be harder to measure, the point is a sound one. We need to make sure that we are thinking about children as whole people, in the context of their families and communities, rather than as trainee adults. Meanwhile, for some of the most vulnerable young people, mentoring programmes as currently implemented may become one more intervention that fails to deliver on its promises.

Showing that something works (or not) is one thing, and difficult enough. But what happens next in real-life settings?

In our case, we summarised the research findings and presented them to practitioners and planners tasked with implementing mentoring and allocated the funds to do so. Unsurprisingly, their response was not, “Let’s send back the funds and go home”. What they did do was to roll up their sleeves and ask “How can we make it work?” They asked for evidence on what seemed to have a more positive effect for the group of children with whom they were working, and decided to include a directive element in their approach, drawing on the evidence of the apparent effectiveness of cognitive behavioural therapeutic approaches in attaining some of the outcomes sought, including delivery via a mentoring-like component (Davidson et al. 1987, Cavell and Hughes 2000). They also drew on practices that seem to be correlated with stronger benefits for young people, such as ongoing training for volunteer mentors and involvement of parents (DuBois et al. 2002). Of course, without implementing these innovations in a trial setting, we will never know whether these approaches are better, worse or much the same as doing nothing, or implementing the standard local means of delivering mentoring.

Box 1 – Key Messages

- Mentoring children and young people at risk for, or already involved in, anti-social behaviours has become popular, but research evidence to support the most commonly used programmes is lacking.
- There is evidence that failed mentoring relationships may have a detrimental effect on a sub-group of children and young people.
- A commitment to research-based practice needs to focus on what works in implementation as well as evidence of effect. Mentoring practices vary widely.
- In order to know more, we need further trials, with end-point users and practitioners involved from the outset in study design

Source: adapted from Roberts et al. 2004

MOVING FORWARD

“Under no child left behind, schools are being set up to fail”
 “Good intentions, bad results”
 “No child left behind produces unintended negative results”
 “Youth programs in Queens are educational, fun and often free”
 “Why mentoring programmes and relationships fail”
 (headlines from newspapers in the United States)

The research (if any) behind the kinds of headlines above is generally given much greater credence than it merits. With a clear message and a good communicator, they can take off in much the same way as Hanushek’s work on spending on schools. But there are few studies that are so methodologically sound, whose results are so generalisable and that leave us so certain that the results represent a good approximation of the “truth” that we should accept their findings outright.

A problem with interpreting and using research is that it is often so far removed from real-life settings that it may be difficult for policy makers or the public to know whether the results are to be taken seriously, or whether they represent no more than the latest unreliable dispatch from the world of science. The more sceptical research-informed policy maker may simply wait patiently, on the grounds that another researcher will soon publish a paper saying the opposite. If one study appears claiming that what delinquents need is a short sharp shock, another is sure to follow suggesting that what they actually need is a teambuilding adventure holiday.

But if one problem is faced by those sceptical of single studies, quite another is faced by the researcher, policy maker or practitioner who tries to range more broadly in his or her reading and thinking. With new journals launched yearly, and thousands of research papers published, it is impossible for even the most energetic policy maker or researcher to keep up-to-date with the most recent research evidence, unless they are interested in a very narrow field indeed. The increasing amount of research information, which varies in quality and relevance, can make it difficult to respond to these pressures, and can make the integration of evidence into practice difficult. An example of information overload is provided in Box 2.

Box 2 – Stopping bullying: information overload (adapted from Petticrew and Roberts 2005)

Teachers, parents and pupils interested in preventing bullying, or stopping it when it does happen, will have no shortage of information. There are over a quarter of a million sites which refer to school bullying on the web. Among the approaches to this problem described by one government organisation, the Department for Education and Skills in the United Kingdom (<http://www.parentcentre.gov.uk>) are:

- co-operative group work
- Circle Time
- Circle of Friends
- befriending
- Schoolwatch
- the support group approach
- mediation by adults
- mediation by peers
- active listening/counselling based approaches
- quality circles
- assertiveness training groups.

How can those using the web work out which sites to trust, and which interventions might actually work? Some sites suggest that certain interventions such as using sanctions against bullies can be ineffective, or even harmful – that is, actually increase bullying (e.g., <http://www.education.unisa.edu.au/bullying>). Other sites suggest that such approaches may work (http://www.education-world.com/a_issues/issues103.shtml). The same intervention may appear to work for some children, but not for others – younger children, for example – and some types of bullying, such as physical bullying, may be more readily reduced than others, such as verbal bullying.

For bullying, as for other types of social problem, one can quickly become swamped with well-meaning advice. Navigating one's way through the swamp is tricky, but systematic reviews provide stepping stones – differentiating between the boggy areas (the morass of irrelevant information) and the higher ground (the pockets of reliable research information on what works and for whom, and where and when).

Systematic reviews can provide a means of synthesising information on bullying, or aspects of bullying, and give a reliable overview of what the research literature can tell us about what works. For example, a systematic review of school-based violence prevention programmes identified 44 trials in all, and concluded that while more high-quality trials are needed, three kinds of programmes may reduce aggressive and violent behaviours in children who already exhibit such behaviour (Mytton et al. 2002).

Systematic literature reviews are a method of making sense of large bodies of information, and a means of contributing to the answers to questions about what works and what does not. They are a method of mapping out areas of uncertainty, and identifying where little or no relevant research has been done but where new studies are needed.

Systematic reviews are a method of critically appraising, summarising and attempting to reconcile the evidence in order to inform policy and practice, and they provide a synthesis of robust studies in a particular field of work which no policymaker or practitioner, however diligent, could possibly hope to read themselves. Systematic reviews are thus unlike “reviews of the studies I could find”, “reviews of the authors I admire,” “reviews which leave out inconveniently inconclusive findings or findings I don't like,” and “reviews which support the policy or intervention I intend to introduce”. Not only do they tell us about the current state of knowledge in an area, and any inconsistencies within it, but they also clarify what we still need to know.

The systematic review adopts a particular methodology in an endeavour to limit bias, with the overall aim of producing a scientific summary of the evidence in any area. In this respect, systematic reviews are simply another research method, and in many respects they are very similar to a survey – though in this case they involve a survey of the literature, not of people. It is less of a discussion of the literature, and more of a scientific tool; but it can also do more than this, and can be used to summarise, appraise and communicate the results and implications of otherwise unmanageable quantities of research. It is widely agreed, however, that at least one of these elements – communication – needs to be greatly improved if systematic reviews are to be really useful.

WHEN TO DO A SYSTEMATIC REVIEW

It can help to do a systematic review:

- when there is uncertainty (for example, about the effectiveness of a policy or a service, and where there has been research on the issue)
- in the early stages of development of a policy, when evidence of the likely effects of an intervention is required
- when it is known that there is a wide range of research on a subject but where key questions remain unanswered – such as questions about treatment, prevention, diagnosis, or causation, or questions about people’s experiences of being on the receiving end of an intervention
- when a general overall picture of the evidence in a topic area is needed to direct future research efforts.

PUTTING THE “D” INTO R&D

If, as many of us believe, and as there is good evidence to demonstrate, the most effective time to make a difference to outcomes is in childhood, and the earlier the better, then it is likely that this is a time when we need to be particularly careful of doing damage. In other words, as well as asking what works and how to implement it, we also need to think about what does not work and how to stop it.

It is not part of the initial training of most academics to work on policy or practice development, and it tends not to be part of research budgets to provide cash for the “D” component, where “D” means development, although increasingly, funders provide time and funds for dissemination, which can be a first step to development.

There are clearly important training needs here, with an eye to those who really do know about the “D” of R&D, such as pharmaceutical companies, and exchanges and secondments between academic life, policy and practice. But none of these will work well if the interventions being proposed are not fit for purpose; are not meaningful to those who are intended to receive them, or are culturally inappropriate.

For this reason, the inclusion of end-point users at every point in the R&D process is not just good democratic practice – it is likely to result in better work and more effective interventions, to say nothing of more fun in the working day.

CONCLUSION

Social interventions are complex and are capable of doing as much or even more harm than medical ones. They need to be subject to as much if not more evaluation before and after implementation.

There are no simple solutions to complex problems, and therein lies a problem for policy makers and politicians. To find a quick-acting solution to an important social problem is a great prize, particularly if the solution is one that will have a result before the next general election. When that solution seems to have common sense behind it, and is not too expensive, it becomes even more compelling.

Those of us who have been pushing policy makers and practitioners to adopt evidence-based policy need to be careful that we do not sell it as a simple way to solve problems. We need a lot more work on how to collaborate effectively with policy makers dealing with complex interventions and evidence. It is probably even clearer to practitioners, policy makers and front-line users of services than it is to researchers that there are massive evidence gaps, sometimes because the right questions are not being asked. For many social interventions, there will be little evidence to review – few primary studies, even fewer that are sufficiently robust to affect policy. But we must be careful not to confuse absence of evidence with evidence of absence.

The R&D agenda in health and social care needs huge investment if we are to develop adequate social interventions for big problems. At present, practitioners, parents and children, and young people themselves looking for good research evidence on common problems will find the evidence cupboard disappointingly bare.

Intervening in children's lives is not just a research policy and practice issue for those of us at the supply end. It is also a rights issue for children and young people. Young people have the right to evidence-based interventions. We know from the past that many well-meaning attempts to do good resulted in harm, but we now have the means through systematic review, trials, sound evaluations and good qualitative work, to do better.

REFERENCES

- Annesley, B., P. Christoffel, R. Crawford, V. Jacobsen, G. Johnston and N. Mays (2002) *Investing in Children's Well-Being from a Life Course Perspective: A Preliminary Analytical Framework and Overview of the Literature*, New Zealand Treasury, Wellington.
- Barlow, J. (1999) *Systematic Review of the Effectiveness of Parent-Training Programmes in Improving Behaviour Problems in Children Aged 3–10 Years* (second edition), Health Services Research Unit, University of Oxford.
- Barnardo's R&D (2002) *What Works? Making Connections Linking Research and Practice*, Barnardo's R&D team, Barnardo's, Barkingside.
- Booth Charles (1902) *Life and Labour of the People in London*, vol. 1, Macmillan, London.
- Butler (2004) *Review of Intelligence on Weapons of Mass Destruction*, Return to an address of the Honourable the House of Commons, dated July 14th, 2004, Report of a Committee of Privy Counsellors, Chairman The Rt Hon the Lord Butler of Brockwell, KG, GCB, CVO, <http://www.butlerreview.org.uk/report/index.asp>
- Cavell, T.A. and J.N. Hughes (2000) "Secondary prevention as context for assessing change processes in aggressive children" *Journal of School Psychology*, 38:199-235.
- Connor J., A. Rodgers and P. Priest (1999) "Randomised studies of income supplementation: A lost opportunity to assess health outcomes" *Journal of Epidemiology and Community Health*, 53:725-730.
- Davidson, W.S., R. Redner, C.H. Blakely, C.M. Mitchell and J.G. Emshoff. (1987) "Diversion of juvenile offenders: An experimental comparison" *Journal of Consulting and Clinical Psychology*, 55(1):68-75.
- Davies, E., B. Wood and R. Stephens (2002) "From rhetoric to action: A case for a comprehensive community-based initiative to improve developmental outcomes for disadvantaged children" *Social Policy Journal of New Zealand*, 19:28-47.
- Doyle, A. (1875) *Pauper Children (Canada)*, return to an order of the Honourable the House of Commons, dated 8 February 1875.
- DuBois, D.L., B.E. Holloway, J.C. Valentine and H. Cooper (2002) "Effectiveness of mentoring programs for youth: a meta-analytic review" *American Journal of Community Psychology*, 30:157-197.
- Glass, N. (1999) "Sure Start: the development of an early intervention programme for young children in the UK" *Children and Society*, 13(4):257-264
- Glass, N. (2001) "What works for children: The political issues" *Children and Society*, 15(1):14-20
- Grossman, J.B. and J.E. Rhodes (2002) "The test of time: Predictors and effects of duration in youth mentoring programs" *American Journal of Community Psychology*, 30:199-206.
- Grossman, J.B. and J.P. Tierney (1998) "Does mentoring work? An impact study of the Big Brothers Big Sisters program" *Evaluation Review*, 22:403-426.

- Hanusheck, Eric A. (1981) "Throwing money at schools" *Journal of Policy Analysis and Management*, 1:19-41.
- Hunt, M. (1997) *How Science Takes Stock: The Story of Meta Analysis*, Russell Sage Foundation, New York.
- Liabo, K. (2002) "What works for children? An evidence-based information source for children's social care" *Learning and Skills Research* 5(3):50-51.
- Liabo, K. (2005) "What works for children and what works in research implementation? Experiences from a research and development project in the united kingdom" *Social Policy Journal of New Zealand*, 24:.
- Liabo, K., P. Lucas and H. Roberts (2003) "Can traffic calming measures achieve the Children's Fund objective of reducing inequalities in child health?" *Archives of Disease in Childhood*, 88(3):235-36.
- Lucas, P., K. Liabo, H. Roberts (2003) "Do behavioural treatments for sleep disorders in children with Down's syndrome work?" *Archives of Disease in Childhood*, 87(5):413-414.
- Macdonald, G. and H. Roberts (1995) *What Works in the Early Years?* Barnardo's, Barkingside.
- Ministry of Education (1993) *The New Zealand Curriculum Framework*, Ministry of Education, Wellington.
- Muir Gray JA (1997) *Evidence Based Healthcare*, Churchill Livingstone, Edinburgh.
- Mytton, J., C. DiGiuseppi, D. Gough, R. Taylor and S. Logan (2002) "School-based violence prevention programming: Systematic review of secondary prevention trials" *Archives of Pediatrics and Adolescent Medicine*, 156(8):752-762.
- Nutley, S., H. Davies and I. Walter (2003) "Evidence-based policy and practice: Cross-sector lessons from the United Kingdom" *Social Policy Journal of New Zealand*, 20:29-48.
- Nutley, S., I. Walter and H. Davies (2002) *From Knowing to Doing: A Framework for Understanding the Evidence into Practice Agenda*, Research Unit for Research Utilisation, Department of Management, University of St Andrews, www.st-andrews.ac.uk/~ruru/RURU/%20publications%20list.htm.
- O'Donnell, C.R., T. Lydgate and W.S.O. Fo (1979) "The Buddy System: Review and follow-up" *Child Behavior Therapy*, 1:161-169.
- Olds D. L., J. Robinson, R. O'Brien, D.W. Luckey, L.M. Pettitt, C.R. Henderson Jr., R.K. Ng, K.L. Sheff, J. Korfmacher, S. Haitt and A. Talmi (2002) "Home visiting by paraprofessionals and by nurses: A randomized, controlled trial" *Pediatrics*, 110(3):486-496.
- Petticrew, M. and H. Roberts (2003) "Evidence, hierarchies and typologies: Horses for courses" *Journal of Epidemiology and Community Health*, 57:527-529.
- Petticrew, M. and H. Roberts (2005) *Systematic Reviews in the Social Sciences: A Practical Guide*, Blackwells, Oxford.
- Rachman, S. and G.T Wilson (1980) *The Effects of Psychological Therapy*, Pergamon, London.

- Roberts, H. (1997) "Socio-economic determinants of health: Children, inequalities and health" *British Medical Journal*, 314(7087):1122-1125.
- Roberts, H., K. Liabo, P. Lucas, D. DuBois and T.A. Sheldon (2004) "Mentoring to reduce antisocial behaviour in childhood" *British Medical Journal*, 328(7438):512-514.
- Schweinhart, L. and D. Weikart (1993) *A Summary of Significant Benefits: The High-Scope Perry Pre-school Study Through Age 27*, High Scope, Ypsilanti, Michigan, and the United Kingdom.
- Scott, S. (1998) "Fortnightly review: Aggressive behaviour in childhood" *British Medical Journal*, 316:202-206.
- St James-Roberts, I. and C. Singh (2001) *Can Mentors Help Primary School Children with Behaviour Problems?* Final report of the Thomas Coram Research Unit between March 1997 and 2000, 233 Home Office Research, Development and Statistics Directorate, Home Office Research Study, London.
- Stevens, M., K. Liabo, S. Frost and H. Roberts (2005 in press) "Using research in practice: A research information service for social care practitioners" *Child and Family Social Work*, 10(1):67-75.
- Tarling, R., J. Burrows and A. Clarke (2001) *Dalston Youth Project Part II (11-14): An Evaluation*, 232 Home Office Research, Development and Statistics Directorate, Home Office Research Study, London.
- Tierney, J.P., J.B. Grossman and N.L. Resch (2000) *Making a Difference: An Impact Study of Big Brothers Big Sisters*, Public/Private Ventures, Philadelphia.
- Webster-Stratton, C., T. Hollinsworth and M. Kolpacoff (1989) "The long-term effectiveness and clinical significance of three cost-effective training programs for families with conduct-problem children" *Journal of Consulting and Clinical Psychiatry*, 57(4):550-553.
- Wood, E. and K. Kunze (2004) *Making New Zealand Fit For Children: Promoting a National Plan of Action for New Zealand Children* (Violence, Exploitation and Abuse Section), UNICEF New Zealand, Wellington.