# 6 WHAT EVIDENCE FOR YOUTH JUSTICE?

## DAVID SMITH

This chapter argues that the claims of successive governments since 1997 that their policies on youth justice are 'evidence-based' are, at best, only partly justified. It contends that, apart from the inherent difficulty of implementing genuinely evidence-based policies in a field as contested and politicised as youth justice, recurrent problems have arisen from misconceptions on the part of policy-makers – and sometimes of practitioners – about what evidence actually exists, and what should be expected of research intended to produce such evidence. These misconceptions lead, it is proposed, to a tendency for the government's position to swing between the poles of excessive optimism and excessive pessimism. The confusions that have resulted are illustrated by an examination of relevant material recently produced by the Youth Justice Board for England and Wales (YJB). The chapter concludes with some reflections on how far youth justice policy can be evidence-based, and what it might look like if it were informed by a realistic understanding and interpretation of the evidence.

## DIVERSION AND THE LABOUR GOVERNMENTS

When the first 'New' Labour government was elected in 1997, it was publicly committed to the 'modernisation' of the processes of government (Cabinet Office, 1999a, 1999b). Modernisation was taken to mean greater policy coherence ('joined-up' government) and the development of policies based on evidence of 'what works'. As applied to youth justice, which it enthusiastically was, 'modernisation' turned out to mean both more and less than the implementation of coherent, empirically grounded policies. As Jones (2002) demonstrated, the incoming government already knew what it wanted to do

about youth justice, which was why it was able to act as quickly as it did, with a White Paper (*No More Excuses*) in 1997 and a very substantial piece of new legislation, the Crime and Disorder Act, in 1998. The Act, among much else, established the YJB to take charge of what the government apparently considered 'a fractured and immature youth justice system' (House of Commons Justice Committee, 2011: 6). Among the Board's responsibilities were 'identifying and promoting effective practice' and 'commissioning research and publishing information' (Ministry of Justice, 2011). So far, so reasonable-sounding, but these duties were to be undertaken within a culture of responsibility that was to replace the 'culture of excuses' from which youth justice had supposedly been suffering. Jones wrote of the Board's 'apparent expurgation of all youth justice knowledge and practice prior to 1998' (2002: 15); while what was to count as evidence was not specified, it was clear that the skills and experience of practitioners would not form part of it.

In particular, the YJB reversed the assumption that it was better to divert young people from the formal system if at all possible, which had been a key principle of youth justice practice from the early 1980s (Fergusson, 2007; Smith, 2010). In fact, the rate of diversion had been in decline since the early 1990s, reflecting the 'repoliticisation' of youth crime by the Conservative government (in which Labour followed), in the wake of youthful disturbances in the early 1990s (Campbell, 1993) and in particular the killing of James Bulger in 1993 (Bateman, 2012; Goldson, 1997; Hay, 1995). The principle of diversion where possible was thus already under pressure, for ideological reasons rather than considerations of effective practice; but the formal rejection of it as a principle was still a striking example of the priority of political over evidential considerations.

The subsequent history of diversion as an element of youth justice policy is interesting, however, in that it shows at least a partial rejection of, or at least a drift away from, the pro-intervention, 'no more excuses' position. Whether this further change was any more evidence-based than the original turn away from diversion is doubtful, but – without actually using the term – the YJB in 2006 introduced a new performance indicator that implied that diversion was back in favour: one of the ways youth justice was to be judged was by its success in reducing the number of 'first time entrants' to the system. This was made government policy in 2008 (Bateman, 2009, 2012) and became one of the Ministry of Justice's three key indicators for youth justice, the others being a fall in 'proven reoffending' and a decrease in the use of custody. So something like the principles that were ruthlessly disavowed by both Conservative and Labour governments in the 1990s were reinstated – and the YJB's apparent success in relation to these indicators seems to have been among the reasons why, in late 2011, the Coalition government reversed its decision to abolish it as part of its 'cull' of 'quangos' (House of Commons Justice Committee, 2011; Ministry of Justice, 2012).

This begins to look like an example of evidence-based policy-making, or at least of an evidence-based change of mind: it could have been difficult to go

ahead with the abolition of the Board in the face of such apparent success (though the part played by the Board itself in contributing to these positive outcomes is not easy to identify), and the Justice Committee was told that the support of the Board was valued by members of local Youth Offending Teams (YOTs). In any case, the reinstatement of diversion and the apparent determination to drive down custody was, for many commentators and practitioners, an encouraging reversal of the previous policy, though, as Bateman (2012; this volume) argues, it would be unwise to assume that the downward trends in system involvement and in the use of custody will prove stable. They might have a better chance of being so if they were clearly and publicly defended on crime-reduction grounds, but no one involved in high-level discussions about the future of the YJB seems to have mentioned that there is strong research evidence in favour of diversion both from the system and from custody. Impressive evidence that targeted early intervention in the lives of young people 'at risk' may increase rather than reduce the prospect of serious offending, and that the best policy in some cases (if the aim is to reduce the likelihood of serious offending) is to do nothing, has come from the quasi-experimental Edinburgh Study of Youth Transitions and Crime (McAra and McVie, 2007, 2010; see also Goldson, 2008; Soothill et al., 2008). This evidence is far stronger than that available in the 1980s: then, practitioners certainly used social science – in the form of labelling theory – to guide their practice (Thorpe et al., 1980), but neither they nor the government (which supported an increase in cautioning instead of prosecution) could have claimed that there was much empirical evidence to support their preference for minimum intervention. Whether the much more persuasive evidence that emerged from the Edinburgh study and elsewhere will get the attention it deserves remains to be seen, though it would be unwise to hold one's breath; McAra and McVie draw the broad conclusion that the policy direction implied by their findings is 'the provision of services proportionate to need and offering maximum diversion' (2010: 201), implying a commitment to a welfare-based system that is in direct contradiction to dominant trends in politics and practice in Scotland as well as in England and Wales.

## USING SOCIAL SCIENCE THEORIES

Practitioners' awareness of labelling theory in the 1980s, and their success in finding space within the system to develop forms of practice that were in line with its predictions, is a (rare) example of the large-scale practical application of a social science theory, or in this case a group of theories. Their use of it might prompt reflection on the nature of knowledge in the social sciences, what counts as evidence, and how it might be used. When they look at social science, research policy-makers, understandably enough, want it to yield clear, unambiguous guidance on how to solve a problem; they tend to look for 'one best way'

answers (Smith, 2004). But this expectation is based on a misunderstanding of the kind of knowledge the social sciences can be expected to produce. The misunderstanding arises from a crude version of positivism, meaning in this context 'the assumption that knowledge in the social sciences is essentially similar in kind to knowledge in the natural sciences, and that if social science is properly conducted, it can produce universal truths that are as stable and reliable as those of, for example, chemistry' (Smith, 2006: 83). There are good reasons why this is not the case, though sometimes social scientists themselves have succumbed to the temptation to give managers and bureaucrats what they demand, and claimed that their research can produce results that can be treated as certain and universal.

The essential reason why this is impossible is that the social world can never be made entirely predictable (MacIntyre, 1985). The logic of theory in the social sciences is, therefore, necessarily different from that of natural science theories. MacIntyre writes that 'What managerial expertise requires for its vindication is a justified conception of social science as providing a stock of law-like generalisations with strong predictive power' (1985: 90). But this requirement, according to MacIntyre, is unachievable. One of the examples he gives to show the difference between generalisations in the social and the natural sciences is Oscar Newman's (1973) well known and enduringly influential theory of defensible space. This predicted, among other things, that crime rates would rise with the height of residential buildings but level off when they reached 13 storeys. This was subjected to close research by positivist criminologists, who found disconfirming as well as supportive cases. But it does not follow that because a prediction of the theory is not right all the time, in every context and in every place, we should abandon the theory or at least radically modify it. It is doubtful if anyone really expects social science theories to deliver universal truths, but managers and policy-makers, including – as will be seen below – the YJB, often behave as if they do. Sensible practitioners and policy-makers meanwhile continue to draw on the social sciences as and when they find them helpful, without expecting them to be right all the time. They are able to do this because they know that human life is not and cannot be entirely predictable and controllable.

Within criminology, a helpful formulation of these ideas is found in John Braithwaite's (1993) paper on how criminologists who want their work to be useful to practitioners should move 'beyond positivism' and think in terms of 'contextual integrated strategies' as they try to deal with problems of crime. Braithwaite argues that criminologists should 'develop a range of theories that are sometimes useful' and that for purposes of practical problem-solving 'it is contextualized usefulness that counts, not decontextualized statistical power' (1993: 386–8). While recognising the importance of good positivist research, not least in exposing some theories as consistently wrong (the evaluation of 'Scared straight' programmes provides an example (Petrosino et al., 2013), though this has not led to their demise), Braithwaite contends that criminologists should 'reject the view that the ultimate value in science is discovering that

single set of law-like statements that offers the best explanation of the phenomenon' (1993: 394). The same ought to apply to criminal justice managers and practitioners.

The status of knowledge in the social sciences might seem an abstract and esoteric issue, but misunderstanding of it has real practical effects. First, the inevitable failure of research on 'what works' to produce universal law-like generalisations can contribute to a rejection by policy-makers and practitioners of the very idea that research can ever be useful. At the level of policy-making, this leaves the field open for evidence-free policy motivated by ideological preferences – as in the case of 'Scared straight' programmes – and/or initiatives that appear to have an obvious populist appeal (e.g. the proposed privatisation of most probation work in England and Wales; Ministry of Justice, 2013). At the level of practice, a belief that research has (and can have) nothing to offer would leave practitioners with nothing to guide them but intuition and habit, and with no basis for reasoned resistance to ideologically driven managerial demands.

## MAKING SENSE OF WHAT WORKS

Second, and more pertinently to the discussion that follows, misunderstanding what social science can deliver leads to an erratic and irrational oscillation on the part of policy-makers (who may also be commissioners of research) between excessive optimism and excessive pessimism (the latter probably being the default position of recent years in England and Wales). Thus, following a period when the orthodox position was apparently characterised by 'nothing works' despondency, well-conducted empirical research began to appear that suggested that something might work after all (Raynor and Vanstone, 1996, 1997). The researchers themselves expressed their conclusions modestly and tentatively. But, according to Mair (2004), there was nothing tentative about the Home Office's reaction to the research: probation managers were encouraged to see cognitive-behavioural offending-focused groupwork programmes as the single answer to the question 'What works?' and to ensure that they became the core of probation practice. A sensible response to the evidence would have been cautious optimism and curiosity about what more could be learned: for example, about the contextual features which enabled such programmes to be effective (Pawson and Tilley, 1997) or the importance of individual staff characteristics (Bonta et al., 2011; Burnett and McNeill, 2005; Raynor et al., 2013). Instead, cognitive-behavioural programmes were 'rolled out' with very little attention to contexts and processes of implementation and with, eventually, a disregard for some basic messages from the 'what works' research. For example, the 'risk principle' that intensity of intervention should be proportional to the assessed risk of reoffending was apparently forgotten: since cognitive-behavioural programmes were taken to comprise *the* answer, everyone should

get them – which led the chief inspector of probation to complain of 'programme fetishism' (HM Inspectorate of Probation, 2002, 2004). The research was interpreted in this way, according to Mair (2004), because of New Labour's commitment to 'modernisation' and therefore to evidence-based policy, and because people in key positions in the Home Office and the probation inspectorate were enthusiastic about the research and believed it could be straightforwardly used to make practitioners work more effectively (by following the prescribed groupwork curriculum).

This over-optimism – not so much about what the research said but about how its implications could be implemented in a changed approach to practice – was, perhaps inevitably, succeeded by over-pessimism. The Home Office's Crime Reduction Programme (CRP), launched in 1998, was the largest and best-funded crime reduction initiative ever attempted in Britain – and a product of the modernising optimism identified by Mair. It was meant to run for 10 years, but was closed down in 2002 (Maguire, 2004). Much of what was envisaged was never implemented, or not implemented as had been hoped. Maguire summarises the problems it encountered:

> Ultimately, few projects were implemented as planned, with the knock-on effect of a dearth of conclusive research findings … it was undermined significantly by inherent risks and tensions that became increasingly prominent as circumstances (and the political climate) changed. While initially conceived as research-driven, it was 'sold' to politicians as contributing to the government's challenging crime reduction targets, an aim which progressively took priority over research … It was over-ambitious in scale and raised unrealistic expectations of its outcomes. It suffered from major practical problems … Low commitment to project integrity, cultural resistance among practitioners, and insufficient attention to differences between academics' and policy makers' understandings of research, also contributed to its problems. (2004: 213–4)

He concludes that 'the ideal of "evidence based policy" may be more effectively pursued as a quiet iterative process over the longer term, rather than through a risky investment in one high profile and rapidly implemented "programme" which promises more than it can guarantee to deliver' (Maguire, 2004: 213–4).


# INTERPRETING RESEARCH

The response of civil servants responsible for research was to criticise the academics involved in evaluating the programme, who were told that 'we did the wrong kind of research' (Raynor, 2004: 319). This entailed a tacit self-criticism, since the advice evaluators had been given at the start of the CRP was very different from the prescriptions for future evaluations that were given after it ended. The initial advice had mentioned randomised controlled trials (RCTs) as a method of evaluation but rejected them as 'usually … not possible for practical reasons' (Colledge et al., 1999: 16). After the end of the CRP, the position of the Home Office was that randomised controlled trials were indisputably

superior to any other method of evaluation. The Home Office review of 'what works' published in 2005 was much more sceptical than an earlier review (Goldblatt and Lewis, 1998) about the state of knowledge on the effectiveness of interventions designed to reduce reoffending. The introduction (Friendship et al., 2005) noted the difficulty of transferring findings from one setting to another (i.e. the Canadian research which had informed much of the development of cognitive-behavioural programmes in Britain might not be relevant after all). It went on to complain about the poor quality of most British evaluative research on the topic, and in particular about the lack of RCTs. The conclusion of the review reiterated this claim: Chitty complained that 'many of the results in this volume say a great deal about implementation, its problems and its effects on outcomes rather than the true effects of interventions' (though what this distinction means is obscure) and called for the use of RCTs in the 'correctional services, so that our knowledge of what works is truly improved and the existing equivocal evidence is replaced with greater certainty' (2005: 79–80). The same line was taken in the 'minimum standards for reconviction studies' produced by the Home Office in 2004 and incorporated into the YJB's Research Strategy for 2008–11 (YJB Research Team, 2008). As will be shown below, this enthusiasm for RCTs remains a feature of the latest thinking about evaluation in the YJB.

Can RCTs really produce definitive, unambiguous results, as Harper and Chitty (2005) and the YJB Research Team appear to believe they can? The experience of the last time the Home Office funded RCT research, in the late 1960s and early 1970s, is not encouraging in this respect. In that period there were four projects concerned with the effect of 'treatment' on reconvictions that used an RCT design (Nuttall, 2003). These were a study of Kingswood Approved School by Clarke and Cornish (1972; 1975); a study of social work in prison by Margaret Shaw (1974); the IMPACT study of the effects of intensive probation (Folkard et al., 1976); and what was presented as a replication of Shaw's study by Fowles (1978). None produced results that could reasonably be described as clear-cut, though taken together they were believed by policy-makers to give broad support to the thesis that nothing worked (Nuttall, 2003). This was despite the fact that the most apparently definite results, from Shaw's prison study, were also the most positive, in terms of suggesting that something had worked. But this message was apparently unwelcome in official circles: the then head of the Home Office Research Unit, I.J. Croft, wrote in his Foreword that 'until they have been repeated, the experiments in social casework described in this report should not necessarily be regarded as the answer to the penologists' prayer' (Shaw, 1974: iii). And, reassuringly enough for penological pessimists, they were not repeated in the study by Fowles (1978), supposedly a replication (though in fact a very approximate one) of Shaw's research – though neither were its findings (like those of the IMPACT study) as negative as they were often presented as being (Nuttall, 2003: 276).

As a result of his experience with the Kingswood research, Ron Clarke became permanently disillusioned with 'treatment' approaches and with RCTs

as an evaluation method, and turned to other forms of research on other approaches to crime reduction, notably situational preventive measures, which evolved into what is claimed as the new discipline of 'crime science' (Clarke, 2004). Nuttall (2003) gives Clarke's views as one local and specific reason for the negative interpretation of the findings of the four RCTs. It may also be that it was difficult for Home Office researchers to accept an interpretation of the findings that was contrary to the emerging orthodoxy that nothing worked (for which Martinson (1974) is usually held responsible). On a larger scale, the response to this research could be interpreted as a product of the emerging disillusionment with treatment approaches which Garland (2001) and others have identified as one element of a much broader movement away from penal modernism and faith in rehabilitative expertise towards a more populist and punitive set of policy assumptions (Pratt et al., 2005). At an intermediate level, the acceptance of the negative view by practitioners – and by most academic commentators – could be explained by their awareness that while the results were not entirely negative, they were not as positive as enthusiasts for offender 'treatment' had hoped.

Whatever level of explanation is adopted – and all are likely to be relevant – the reception of these research reports is a clear example of how research is always delivered into a particular political and ideological context, and how it is that context, and not the findings of the research alone, determines how it will be used. In other contexts the response to the findings might have been 'Interesting – let's see what more we can find out', instead of a near-total loss of interest in evaluative research on interventions with offenders on the part of the government, which lasted for over 20 years. As discussed above, its revival, when it came, was fragile, as an ostensible commitment to evidence-based policy was overtaken by what were seen as political imperatives. Still, we are not quite back to 'nothing works' even though – as the following discussion of the uses of research in youth justice shows – there is not much sign of confidence that we know what does 'work'.

## WHAT WE KNOW AND DON'T KNOW IN YOUTH JUSTICE: THE OFFICIAL STORY

It seems fair to say that at the time of writing (late 2013) the status of evidence on youth justice in official circles is not entirely clear. In late 2011 the YJB seemed to accept that it had failed to make practitioners sufficiently aware of what counts as effective practice. Giving evidence to the House of Commons Justice Committee, the Board's chief executive, John Drew, said that 'Effective practice is probably the area of the YJB where we have met our mandate least satisfactorily' (House of Commons Justice Committee, 2011: Ev. 20). He was responding to a question about the National Audit Office's (NAO) finding in the previous year that 'seventy-six per cent of YOT managers agreed with the statement, "It is

difficult to find evidence on 'what works' for certain areas of our work'", from which the NAO report's authors seemed to draw the grand but doubtfully justified conclusion that 'Practitioners in the youth justice system do not know which interventions have the most impact on reducing offending' (National Audit Office, 2010: 8). It would be understandable if John Drew was on the defensive in giving his evidence, since the Justice Committee was specifically examining the proposed abolition of the YJB, but it is not clear how defensive he needed to be, since, as he told the Committee, the YJB had published 73 research studies, 31 of which were concerned with the outcomes of interventions and 'about another dozen' were 'in the pipeline'. He was surely justified in saying that the YJB's contribution to research was not 'negligible', and that 'we know quite a lot about what works'. All the same, he said that recent scrutiny of the YJB's work on effective practice (by the NAO and in the internal review chaired by Dame Sue Street; Department for Education, 2010) had been 'a real wake-up call' and that the YJB was 'in the process of reformulating our entire offer in relation to effective practice' and as a result would be 'much more focused'.

YOT staff who look at the YJB's website, as they are presumably expected to do from time to time, are unlikely to be struck by a lack of material on 'effective practice'. They might, however, be baffled by the variety of the material that they find. As well as a range of research reports on specific topics, some, as John Drew said, evaluating the results of interventions, the enquiring practitioner will find (among other things) 10 statements of the Key Elements of Effective Practice (KEEPs), each with an accompanying 'source document' that gives the basis in evaluative research findings for the brief guidance offered in the KEEP document. These KEEP statements date from 2008, when the original versions were revised (Prior and Mason, 2010). With a bit more effort, the practitioner will also find some more recent documents, which taken together make up the 'effective practice framework'. These include an account of the YJB 'Practice Classification System' (Archer, 2013a) and a statement on 'Effective Practice Identification and Dissemination' (Archer, 2013b), as well as reports from the previous two years on the 'Effective practice prioritisation exercise'. There is also an 'Effective Practice Library', which is meant to 'provide practitioners and commissioners in youth justice with easy access to examples of effective practice' and to 'allow those developing and using innovative practice to share what they have found to be effective' (Youth Justice Board, 2012). (How far it succeeds in these aims is briefly discussed below.) On the face of it, youth justice practitioners do not lack advice from the YJB on 'what works'; presumably the post-2008 documents represent part of the YJB's response to criticism by the NAO and others, but even before then there had hardly been a shortage of material. There are, however, interesting differences between the earlier and later sets of documents in the view they take of what counts as evidence of effectiveness. I look first at the earlier set, the KEEPs, and in particular their source documents.

The 10 KEEPs follow a standard format and, after suggestions on how they should be used, give general guidance to practitioners and managers on 'delivery', 'operational management' and 'strategic management and partnership

working'. They reiterate the importance of monitoring and evaluation at each level and, by means of cross-referencing, they encourage their readers to refer to the relevant source documents. There is little in the KEEPs that anyone could sensibly object to, but much of the guidance seems so general as to be unlikely to be very helpful: many readers might well feel that they know all this already, and the NAO (2010: 36) found that most YOT respondents would have liked 'more practical guidance'. The source documents are a different matter, being generally authoritative statements, based on systematic reviews of the literature, on what evidence there is that could helpfully inform practice. Only 50 per cent of respondents to the NAO (2010) thought that the information they got from the YJB had a convincing evidence-base, but given the effort put into the source documents it seems unreasonable to blame the Board (or only the Board) for this.

## WHAT KINDS OF EVIDENCE COUNT?

It may be, however, that the Board's expectations of the authors of the source documents led to an unnecessary limitation on what kinds of evidence were to count as valid. As Prior and Mason (2010) explain, the brief for the source documents emphasised that authors should conduct a systematic review of research in the field following the guidelines of the Campbell Collaboration (www.campbell-collaboration.org/). In practice, Campbell Collaboration reviews of research usually include only the findings of experiments or quasi-experiments, with a strong preference for RCTs (see e.g. Strang et al. (2013) on restorative justice, and the authors' explanation in Appendix A of why even some RCTs were excluded as insufficiently rigorous). Full adherence to the Campbell position that only RCTs count as valid would have meant that the KEEP source documents – with the possible exception of the one on offending behaviour programmes (Wikström and Treiber, 2008) – would have been a good deal thinner than they are. This is not just because many topics that could in principle have been researched using RCTs or something approaching them have not been, but, as Prior and Mason say, because 'certain research questions are, in effect, rendered "unaskable" because they cannot be addressed using experimental methods' (2010: 219). One such question is their own, on how best to go about *Engaging Young People Who Offend* (Mason and Prior, 2008). They found no studies of 'engagement' that met the systematic review criteria, and, rather than producing a 'very slim' source document that would be of little help to anyone, they agreed after discussions with YJB staff that they should move beyond the Campbell principles and conduct a broader review that would include what 'robust' evidence they could find specifically on techniques of engagement, findings from other YJB reviews that were relevant to these techniques, and 'key messages' on engagement from the practice literature. As they say, the last category is of special interest because it 'opens up the possibility of a quite distinctive conception of what might count as "evidence"

in attempts to establish "what works"' (Prior and Mason, 2010: 213–4). As has long been accepted in thinking about evidence-based medicine – but, oddly, less so in social work and youth justice – the skills and experience of practitioners ought to count as a valid source of evidence (Sackett et al., 1997; Smith, 2006).

It is to the credit of the YJB staff involved that they were open to the possibility that helpful evidence might be found in places other than the results of RCTs. Most of the other KEEP source documents share this view – necessarily, since if they were to confine themselves to RCTs they would have been very thin indeed. It is, therefore, slightly odd to find that the most recent (post-NAO report) material on effective practice on the YJB website looks like a strong restatement of the view that RCTs are superior to any other kinds of evaluative research. Thus Archer (2013a), in explaining how the YJB's Effective Practice Classification Panel will work, reproduces from Friendship et al. (2005: 7) the 'Scientific Methods Scale adapted for reconviction studies'. The scale ranks research from 1 to 5, with RCTs at Level 5, followed by studies with a well-matched comparison group, with an unmatched comparison group and with no comparison group (e.g. using a risk predictor to compare actual and expected outcomes), down to Level 1, which are before and after studies with no point of comparison. The scale encourages the view that anything below Level 4 is barely worthy of attention – and is certainly unlikely to be funded. What Archer does not reproduce is Friendship et al.'s (2005: 8) suggestion that for an expected reduction of 2.5 percentage points in the rate of reconviction a sample size of 5,024 is needed for both intervention and control groups – and that even on the optimistic assumption of a 10 percentage point reduction the minimum number required for each group is 325. It is not surprising that Friendship et al. remarked that few UK studies had met these requirements and concluded that Level 5 had rarely been achieved in reconviction studies – but Archer (2013a) pays no attention to this surely critical question of whether RCTs, or even Level 4 studies on a large enough scale, are practical propositions.

To be fair, the YJB does not go so far as to dismiss all other kinds of evaluation research entirely. Archer (2013a: 7–9) gives considerably more space to qualitative than to quantitative methods, and provides a sensible list of factors to be considered in appraising qualitative evaluations (a shorter and perhaps more user-friendly list than the original version in Spencer et al., 2003). He also acknowledges the limitations of quantitative methods, including their typical lack of attention to processes and the problem of how far the results of a controlled study can be generalised to different contexts. Correspondingly, he notes the virtues of qualitative methods in illuminating 'why' and 'how' questions about interventions, and he is clear about the importance of the context in which any intervention is undertaken and of practitioner variables. But he also notes that qualitative methods 'do not offer the same scientifically rigorous certainty [as RCTs, presumably] that it was the practice or programme being evaluated that produced the results seen, or to what extent those results were achieved' (Archer, 2013a: 12). Overall, the message is that quantitative methods are superior, and the more rigorously they are applied the better.

This is also the view of the YJB's partner in matters of research, the Social Research Unit at Dartington, which in turn draws inspiration from the 'Blueprints' project at the University of Colorado.[1] The Blueprints criteria are exacting: before a 'program' can be classified as 'promising' it must have been tested by one high-quality RCT or two high-quality quasi-experiment evaluations; and the requirement for a 'model program' is two high-quality RCTs or one such RCT plus a high-quality quasi-experiment. Few programmes qualify as models: eight out of over 1,100 'youth promotion programs' were reviewed. The list includes some that are so general that their implementation must vary widely across contexts and by the skills and commitment of practitioners, even if the content could somehow be controlled: Functional Family Therapy, LifeSkills Training and Multisystemic Therapy, for example. It is hard to see how this categorisation can help practitioners (or 'commissioners', in YJB-speak). It is equally unclear what they are meant to do when faced with the YJB's 'Effective Practice Library' (Youth Justice Board, 2012), which takes a similarly parsimonious position in bestowing its approval. Functional Family Therapy and Multisystemic Therapy are among eight (or so) out of 168 types of intervention (including tools or aids to intervention as well as 'programs') that are awarded 'research-proven' status; a few others are 'promising' (fortunately, the Blueprints criteria have not been applied in their full rigour to all of these); and the great majority are 'emerging', which is a polite way of saying that they have not been evaluated. Nobody could reasonably say that the YJB has not made an effort to respond to the criticisms of the NAO and others, but it is reasonable to wonder how productive this effort has been.

In particular, the cause of evidence-based practice would be seriously harmed if the idea that only RCTs can produce valid results were taken seriously. As we have seen, RCTs on interventions with offenders are in short supply everywhere, and almost non-existent in the UK. There is every reason to think that this will continue to be the case. The demands of RCTs articulated by Friendship et al. (2005) are such as to make them practically impossible to carry out – especially in a time when public sector spending is being deliberately cut. RCTs have their enthusiastic advocates, notably Farrington (1997) and Sherman (2009). The main example Sherman gives to support his case that RCTs can be executed effectively is the work of Shapland et al. (2008) on restorative justice schemes and their effects on reconviction. This was funded from 2001 under the CRP discussed above. Over four reports Shapland et al. provided rich material on both processes and outcomes and went some way towards answering the question that RCTs are often criticised for not answering (or asking): what was it about the interventions that made the difference? In many respects this research is exemplary, as good as research on criminal justice programmes is likely to get. But Sherman's (2009) claim that it shows the feasibility and value of the RCT approach is doubtful: Shapland et al. (2008) themselves say that

---

[1] See http://dartington.org.uk/projects/blueprints-for-success/.

only one of the three schemes they evaluated (though the largest of them) used random allocation to restorative justice conferencing, so for the other two schemes they had to identify a comparison group, matched as far as possible. They therefore rate their work on the largest project at Level 5 on the Scientific Methods Scale, and their work on the other two projects at Level 4. To achieve sufficient numbers for statistically significant results, cases from the different schemes had generally to be aggregated, but since 'restorative justice' meant different things in different places, the aggregated outcomes were not all produced by the same intervention. None of this is to disparage the work of Shapland et el., but it is to suggest that its results were not as clear-cut as Sherman (2009) suggests. While persuasive, and encouraging for advocates of restorative justice, they inevitably leave questions unanswered (Smith, 2012).

## CONCLUSIONS

Maguire (2004) was quoted above as arguing that evidence-based policy should be seen as a 'quiet iterative process' that will take time to achieve. Also reflecting on the CRP, Raynor concluded that 'The business of using research evidence to improve services is more incremental, provisional, iterative and gradual than big gestures would like it to be' (2004: 322). I would apply this to 'big gestures' in research as well as in offender-related programmes: to conduct an RCT in the UK with the numbers specified by Friendship et al. (2005) – even if we think of the lower figure (325 in each group) and not the higher one (5,024 in each group) – would require an almost unimaginable input of resources. It seems reasonable to conclude that it is not going to happen any time soon. So the claim that such an RCT – or several of them – would finally provide *the* answers that have previously eluded us, is to mislead policy-makers about what kind of research is feasible. It also reflects a mistaken view of the nature of evaluation, the kind of evidence it can produce, and the nature of the social sciences as a whole.

It is unrealistic to expect that youth justice policy will be shaped by research (Fergusson, 2007), but it ought to be realistic to expect that it will be informed and influenced by it (Goldson, 2010). The argument of this chapter suggests that this is most likely to be achieved if policy-makers are helped to be realistic about what they should and should not expect of research. They should not look for MacIntyre's (1985) 'law-like generalisations', but for empirically informed ideas on what approaches look promising, and what conditions are required for them to be implemented properly. Researchers, therefore, have a responsibility not to pander to politicians' demands for certainty, for single true answers. The claims they make for evaluative research should be more modest in one sense (they are not claiming that they can reveal universal truths) but in another sense they can be more confident, since they will not be trapped by the belief that only RCTs, or something very like them, can tell us anything useful.

# REFERENCES

Archer, B. (2013a) *YJB Practice Classification System*. London: Youth Justice Board.

Archer, B. (2013b) *Effective Practice Identification and Dissemination*. London: Youth Justice Board.

Bateman, T. (2009) 'Figures show a fall in the number of "first time entrants" to the Youth Justice System in England and Wales', *Youth Justice*, 9 (1): 87–8.

Bateman, T. (2012) 'Who pulled the plug? Towards an explanation of the fall in child imprisonment in England and Wales', *Youth Justice*, 12 (1): 36–52.

Bonta, J., Bourgon, G., Rugge, T., Scott, T., Yessine, A.K., Gutierrez, L. and Li, J. (2011) 'An experimental demonstration of training probation officers in evidence-based community supervision', *Criminal Justice and Behavior*, 38 (11): 1127–48.

Braithwaite, J. (1993) 'Beyond positivism: learning from contextual integrated strategies', *Journal of Research in Crime and Delinquency*, 30 (4): 383–99.

Burnett, R. and McNeill, F. (2005) 'The place of the officer-offender relationship in assisting offenders to desist from crime', *Probation Journal,* 52 (3): 221–42.

Cabinet Office (1999a) *Modernising Government*. London: The Stationery Office.

Cabinet Office (1999b) *Professional Policy Making in the Twenty-First Century*. London: The Stationery Office.

Campbell, B. (1993) *Goliath: Britain's Dangerous Places*. London: Methuen.

Chitty, C. (2005) 'The impact of corrections on re-offending: conclusions and the way forward', in Harper, G. and Chitty, C. (eds), *The Impact of Corrections on Re-offending: A Review of 'What Works'*, 2nd edn, Home Office Research Study 291. London: Home Office.

Clarke, R.V. (2004) 'Technology, criminology and crime science', *European Journal on Criminal Policy and Research*, 10 (1): 5–63.

Clarke, R. V. G. and Cornish, D.B. (1972) *The Controlled Trial in Institutional Settings: Paradigm or Pitfall for Penal Evaluators?* Home Office Research Study 15. London: HMSO.

Clarke, R. V. G. and Cornish, D.B. (1975) *Residential Treatment and its Effects on Delinquency.* Home Office Research Study 32. London: HMSO.

Colledge, M., Collier, P. and Brand, S. (1999) *Programmes for Offenders: Guidance for Evaluators*, CRP Guidance Note 2. London: Home Office.

Department for Education (2010) *Safeguarding the Future: A Review of the Youth Justice Board's Governance and Operating Arrangements*. London: Department for Education.

Farrington, D.P. (1997) 'Evaluating a community crime prevention program', *Evaluation*, 3 (2): 157–73.

Fergusson, R. (2007) 'Making sense of the melting pot: multiple discourses in youth justice policy', *Youth Justice*, 7 (3): 179–94.

Folkard, M.S., Smith, D.E. and Smith, D.D. (1976) *IMPACT Volume II: The Results of the Experiment*, Home Office Research Study 24. London: HMSO.

Fowles, A.J. (1978) *Prison Welfare: An Account of an Experiment at Liverpool*, Home Office Research Study 45. London: HMSO.

Friendship, C., Street, R., Cann, J. and Harper, G. (2005) 'Introduction: assessing the evidence', in Harper, G. and Chitty, C. (eds), *The Impact of Corrections on Re-offending: A Review of 'What Works'*, 2nd edn, Home Office Research Study 291. London: Home Office.

Garland, D. (2001) *The Culture of Control: Crime and Social Order in Late Modernity*. Oxford: Clarendon.

Goldblatt, P. and Lewis, C. (1998) *Reducing Offending: An Assessment of Research Evidence on Ways of Dealing with Offending Behaviour*, Home Office Research Study 187. London: Home Office.

Goldson, B. (1997) 'Children in trouble: state response to juvenile crime', in Scraton, P. (ed.), *'Childhood' in 'Crisis'?* London: UCL Press.

Goldson, B. (2008) 'Early intervention in the youth justice sphere: a knowledge-based critique', in Blyth, M. and Solomon, E. (eds), *Prevention and Youth Crime*, Researching Criminal Justice Series. Bristol: Policy Press.

Goldson, B. (2010) 'The sleep of (criminological) reason: knowledge-policy rupture and New Labour's youth justice legacy', *Criminology and Criminal Justice*, 10 (2): 155–78.

Harper, G. and Chitty, C. (eds) (2005) *The Impact of Corrections on Re-offending: A Review of 'What Works'*, 2nd edn, Home Office Research Study 291. London: Home Office.

Hay, C. (1995) 'Mobilization through interpellation: James Bulger, juvenile crime and the construction of a moral panic', *Social and Legal Studies*, 4 (2): 197–223.

HM Inspectorate of Probation (2002) *Annual Report 2001–2002*. London: Home Office.

HM Inspectorate of Probation (2004) *Annual Report 2003–2004*. London: Home Office.

House of Commons Justice Committee (2011) *The Proposed Abolition of the Youth Justice Board. Tenth Report of Session 2010–2012* (HC 1547). London: The Stationery Office.

Jones, D.W. (2002) 'Questioning New Labour's youth justice strategy: a review article', *Youth Justice*, 1 (3): 14–26.

MacIntyre, A. (1985) *After Virtue: A Study in Moral Theory*, 2nd edn. London: Duckworth.

Maguire, M. (2004) 'The crime reduction programme in England and Wales: reflections on the vision and the reality', *Criminology and Criminal Justice*, 4 (3): 213–37.

Mair, G. (2004) 'The origins of what works in England and Wales: a house built on sand?', in Mair, G. (ed.), *What Matters in Probation*. Cullompton: Willan.

Martinson, R. (1974) 'What works? Questions and answers about prison reform', *The Public Interest*, 35: 22–54.

Mason, P. and Prior, D. (2008) *Engaging Young People Who Offend: Source Document*. London: Youth Justice Board.

McAra, L. and McVie, S. (2007) 'Youth justice? The impact of agency contact on patterns of desistance from offending', *European Journal of Criminology*, 4 (3): 315–45.

McAra, L. and McVie, S. (2010) 'Youth crime and justice: key messages from the Edinburgh study of youth transitions and crime', *Criminology and Criminal Justice*, 10 (2): 179–209.

Ministry of Justice (2011) *Consultation on Reforms Proposed in the Public Bodies Bill* (Consultation Paper 10/2011). London: Ministry of Justice.

Ministry of Justice (2012) *Government's Response to the Justice Committee's Report: The Proposed Abolition of the Youth Justice Board*. London: Ministry of Justice.

Ministry of Justice (2013) *Transforming Rehabilitation: A Strategy for Reform* (Cm 8619). London: The Stationery Office.

National Audit Office (2010) *The Youth Justice System in England and Wales: Reducing Offending by Young People*. London: National Audit Office.

Newman, O. (1973) *Defensible Space: Crime Prevention through Urban Design*. London: Architectural Press.

Nuttall, C. (2003) 'The Home Office and random allocation experiments', *Evaluation Review*, 27 (3): 267–89.

Pawson, R. and Tilley, N. (1997) *Realistic Evaluation*. London: Sage.

Petrosino, A., Turpin-Petrosino, C., Hollis-Peel, M.E. and Lavenberg, J.G. (2013) *Scared Straight and Other Juvenile Awareness Programs for Preventing Juvenile Delinquency: A Systematic Review*. Available at: www.campbellcollaboration.org/lib/project/3/ (accessed 9.19.14). Oslo: The Campbell Collaboration.

Pratt, J., Brown, D., Brown, M., Hallsworth, S. and Morrison, W. (eds) (2005) *The New Punitiveness: Trends, Theories, Perspectives*. Cullompton: Willan.

Prior, D. and Mason, P. (2010) 'A different kind of evidence? Looking for "what works" in engaging young offenders', *Youth Justice*, 10 (3): 2121–22.

Raynor, P. (2004) 'The probation service "pathfinders": finding the path and losing the way?', *Criminal Justice*, 4 (3): 309–25.

Raynor, P. and Vanstone, M. (1996) 'Reasoning and rehabilitation in Britain: the results of the Straight Thinking on Probation (STOP) programme', *International Journal of Offender Therapy and Comparative Criminology*, 40: 279–91.

Raynor, P. and Vanstone, M. (1997) *Straight Thinking on Probation (STOP): The Mid-Glamorgan Experiment*. Oxford: Centre for Criminological Research.

Raynor, P., Ugwudike, P. and Vanstone, M. (2013) 'The impact of skills in probation work: a reconviction study', *Criminology and Criminal Justice*, July. Available online at: http://crj.sagepub.com/content/early/2013/07/14/1748895813494869.full.pdf+html (accessed 16.12.13).

Sackett, D.L., Richardson, W.S., Rosenberg, W. and Haynes, R.B. (1997) *Evidence-based Medicine: How to Practise and Teach EBM*. Edinburgh: Churchill Livingstone.

Shapland, J., Atkinson, A., Atkinson, H., Dignan, J., Edwards, L., Hibbert, J., Howes, M., Johnstone, J., Robinson, G. and Sorsby, A. (2008) *Does Restorative Justice Affect Reconviction? The Fourth Report from the Evaluation of Three Schemes*, Ministry of Justice Research Series 10/08. London: Ministry of Justice.

Shaw, M. (1974) *Social Work in Prison*, Home Office Research Study 22. London: HMSO.

Sherman, L.W. (2009) 'Evidence and liberty: the promise of experimental criminology', *Criminology and Criminal Justice*, 9 (1): 5–28.

Smith, D. (2004) 'The uses and abuses of positivism', in Mair, G. (ed.), *What Matters in Probation*. Cullompton: Willan.

Smith, D. (2006) 'Youth crime and justice: research, evaluation and "evidence"', in Goldson, B. and Muncie, J. (eds), *Youth Crime and Justice*. London: Sage.

Smith, D. (2010) 'Out of care 30 years on', *Criminology and Criminal Justice*, 10 (2): 119–35.

Smith, D. (2012) 'Arguments about methods in criminal justice evaluation', in Bowen, E. and Brown, S. (eds), *Perspectives on Evaluating Criminal Justice and Corrections*. Bingley: Emerald.

Soothill, K., Ackerley, E. and Francis, B. (2008) 'Criminal convictions among children and young adults: changes over time', *Criminology and Criminal Justice*, 8 (3): 297–315.

Spencer, L., Ritchie, J., Lewis, J. and Dillon, L. (2003) *Quality in Qualitative Evaluation: A Framework for Assessing Research Evidence*. London: Cabinet Office.

Strang, H., Sherman, L.W., Mayo-Wilson, E., Woods, D. and Ariel, B. (2013) *Restorative Justice Conferencing (RJC) Using Face-to-Face Meetings of Offenders and Victims: Effects on Offender Recidivism and Victim Satisfaction. A Systematic Review*. Oslo: The Campbell Collaboration. Available at: www.campbellcollaboration.org/lib/project/63/ (accessed 16.12.13).

Thorpe, D.H., Smith, D., Green, C.J. and Paley, J.H. (1980) *Out of Care: The Community Support of Juvenile Offenders*. London: Allen and Unwin.

Wikström, P.-O. and Treiber, K. (2008) *Offending Behaviour Programmes. Source Document*. London: Youth Justice Board.

YJB Research Team (2008) *Research Strategy 2008–11*. London: Youth Justice Board. Available at: www.justice.gov.uk/downloads/publications/yjb/2008/ResearchStrategy200811.pdf, (accessed 16.12.13).

Youth Justice Board (2012) *Purpose of the Effective Practice Library*. London: Justice. Available at: www.justice.gov.uk/youth-justice/effective-practice-library (accessed 13.10.14).