From 'nothing works' to 'post-truth': The rise and fall of evidence in British probation

Peter Raynor [ID]
Swansea University, UK

Abstract

Since its origins over half a century ago, evaluative research on probation services has swung between optimism and pessimism. This article, based largely on England and Wales, describes and reviews the long journey from over-optimism, via 'nothing works' in the 1970s, to programmes based on Risk-Need-Responsivity principles, introduced on a large scale from the late 1990s but limited in their impact owing largely to problems in implementation. After this, evaluation researchers developed greater interest in implementation, in organisational culture and, in particular, in practitioners' skills. In the process, researchers have developed a better understanding of the necessary social science methods for evaluation and have begun to learn from new sources such as desisting former offenders. In the meantime, in spite of encouraging research, the political context in some countries has become hostile, and research has to survive in a new context of 'post-truth' and politically motivated denigration of expertise.

Keywords

Effectiveness of probation, 'nothing works', evaluation research methodology, core correctional practices, post-truth.

Introduction: Why social science has three legs

Evaluating the effectiveness of probation has always been a knotty problem for social science. Like most (perhaps all) social sciences, it depends on achieving the right combination of different sources and kinds of knowledge. This article reviews key stages in the development of evaluation research on probation services, with a major but not

Corresponding author:
Peter Raynor, Swansea University, College of Law, Vivian Tower, Swansea, SA2 8PP, UK.
Email: p.raynor@swansea.ac.uk

exclusive focus on England and Wales, and ends with some suggestions about the future. Research on the skills used by probation staff marks a particular and important step in this development.

Social science, as used in evaluative research, is a three-legged creature supported by three sources of knowledge or forms of investigation: understanding, measurement and comparison. (Three is usually the minimum number of legs required to support a stable structure.) We need understanding, usually acquired by qualitative research methods, to bring into focus the aims of social actors, their beliefs about the processes they are involved in, and the meanings they attach to what they do and to what happens to them. The criminologist David Matza (1969) called this 'appreciation', meaning the attempt to understand social situations from the point of view of those involved. Human action is socially constructed and our social environments are structured by our actions and by those of others (particularly those more powerful than ourselves). Understanding also requires awareness of our own assumptions and our ways of interpreting and shaping experience, because what we learn will be the product of an interaction between our own perceptions and those of our research subjects. This is why qualitative researchers have to try hard to be guided by what they actually find rather than what they expect or hope or prefer to find. However, it is not clear how social science can be *social* without an attempt to understand the meanings of social experience for the people involved. Some notable recent examples of probation research have relied on qualitative methods: for example, research on the occupational culture of probation staff (Mawby and Worrall, 2013) and on their beliefs about the quality of probation work and the nature of good practice (Robinson et al., 2014). However, evaluation research in probation needs to go beyond practitioners' beliefs to develop more independent and objective ways to measure the impact of probation practice: what does it change? What difference does it make?

Here, we need to depend more on the quantitative procedures of measurement and comparison. Can we actually identify and measure a difference in outcome? Can we reliably estimate whether it is likely to recur? Can we show that the difference is likely to be due to some probation practice or process, not something that would have happened anyway or which occurs simply by chance? These are the scientific procedures that allow us to claim social investigation as a *science*, capable of generating reliable knowledge and building a cumulative knowledge base. Evaluation research, being centrally concerned with whether professional intervention makes a difference, depends on getting these procedures right so that we can learn what works, how it works and in what circumstances, and how we might make it work better. Without understanding, we cannot get far, but without measurement and comparison it is difficult to turn understanding into evidence-based statements about the effectiveness of probation practice. It is this effectiveness that has historically, and repeatedly, been called into question. It is of course important to make the right comparison: that is, the comparison that is relevant to testing your hypothesis or answering your research question. Much methodological writing has tried to establish a hierarchy of research strategies ranked according to the degree of certainty they offer about the validity of findings. Many argue that controlled trials based on random allocation and a strict experimental model are required (e.g. Harper and Chitty, 2004), while others point out that well-designed quasi-experimental studies offer an almost comparable degree of certainty and are much more feasible in the criminal

justice environment (e.g. Hollin, 2008). This article cites examples of both, and both have made contributions to probation research.

Early days: A 'landmark in penal history', optimistic practice and sceptical research

Official aspirations to develop an evidence-based or evidence-informed Probation Service in Britain date back at least to the period of ground-breaking and comprehensive social policy reform that followed the end of the Second World War. This delivered a National Health Service and most of the structure of our current Welfare State, still recognisable in spite of the damage done by more recent hostile governments. There had been earlier attempts to articulate a scientific basis for probation (described in detail by Vanstone, 2004) but official awareness of probation's potential contribution as an integral part of the State's criminal justice and welfare policies emerged strongly during the late 1940s and 1950s. The White Paper 'Penal Practice in a Changing Society' (Home Office, 1959), which set out new aspirations for the prison system, gives a clear insight into official thinking: custody, particularly for the growing number of young offenders, was to lose its punitive emphasis and concentrate on finding the appropriate 'treatment' for offenders who were to be allocated to suitable regimes through a process of assessment and classification, led by psychologists: 'methods of training have been progressively extended and improved, notably in the application of psychiatry and psychology' (Home Office, 1959: 46). Their effectiveness was to be evaluated by the recently established Home Office Research Unit ('The Research Unit is at present studying the effectiveness of different forms of treatment when applied to different types of offender' (p. 5)) and the methodology for the evaluations was to be pragmatically eclectic, but quantitative where possible ('The Research Unit will apply the basic principles of scientific method and attempt to produce its results in quantitative terms. It will not cling to the methods of any particular discipline or school of thought, but will seek to provide answers to specific questions by whatever means appear most appropriate' (p. 6)). The overall ambition was to achieve a step change towards an evidence-based and rehabilitative system:

> A fundamental re-examination of penal methods, based on studies of the causes of crime, or rather of the factors which foster or inhibit crime, and supported by a reliable assessment of the results achieved by existing methods, could be a landmark in penal history and illumine the course ahead for a generation. (p. 7)

The whole document, from nearly six decades ago, breathes modernity and optimism, faith in new human sciences, and an enlightened rejection of purely punitive approaches. As often happened then and more recently, new thinking about prisons had consequences for probation: for example, new post-custodial after-care responsibilities were acquired, but more fundamentally the new approach set out in the 1959 White Paper was clearly meant to encompass probation. A significant part of the work of the Home Office Research Unit and of the Cambridge Institute of Criminology (supported largely by Home Office funds) was to focus on probation over the next 17 years.

In spite of the confidence about scientific and progressive criminal justice expressed in the 1959 White Paper, doubts about the evidence base and effectiveness of probation were already being expressed in parts of the social science and social policy community. The authors of the White Paper had few doubts: they saw the Probation Service as a 'nation-wide network of qualified social case-workers' (Home Office, 1959: 20) fit to assume demanding new tasks in the after-care of prisoners. Others, however, were sceptical of the claims of the young social work profession, with which probation was largely identified at that time (and still should be, though that is another story: see, among others, Raynor and Vanstone, 2015). As early as 1943, the sociologist C. Wright Mills argued that attributing social problems to individual malfunctions distracted attention from the need for wider policy reforms (Mills, 1943). Even earlier, in 1931, Dr Richard Cabot's presidential address to the American Association of Social Workers called for more evaluation research, resulting eventually in the Cambridge Somerville Youth Study (Powers and Witmer, 1951). This substantial and methodologically sophisticated experiment was based on a sample of adolescent boys of whom half were randomly allocated to supervision by social workers, and their subsequent level of offending was compared with that of the control group of boys who were not allocated to social workers. This design was strong on measurement and comparison, but involved little understanding or control of what the social workers were actually doing. As was widely reported at the time (though not much discussed within British social work), the experimental group did no better than the controls; in fact, they were reported to have offended slightly but not significantly more. Thirty years later, in a remarkable follow-up study (McCord, 1978), it was found that the experimental group had continued to do worse, this time significantly and on a range of indicators including crime, unemployment, alcohol abuse, mental illness, stress related illness and earlier death.

In addition to the White Paper, 1959 also saw the publication of Barbara Wootton's *Social Science and Social Pathology* (Wootton, 1959) that contained a pointed critique of social casework theory. Other American studies such as the Vocational High experiment (Meyer et al., 1965) and the Chemung County study of family services (Wallace, 1967) showed no clear benefits from social casework. By the 1970s, there were enough of these studies to lead to significant anxiety in American social work (Fischer, 1973, 1976; Grey and Dermody, 1972). Joel Fischer, who undertook a comprehensive review of social work evaluations up to the 1970s, famously summed up his findings like this: 'The bulk of practitioners in an entire profession appear to be practicing in ways which are not helpful or even detrimental to their clients, and, at best, operating without a shred of empirical evidence validating their efforts' (Fischer, 1976: 140).

What stands out about the American social work evaluations of that era, when tested against the three-legged model of social science, is that they tended to be relatively well executed with regard to measurement and comparison but they did not really examine, unpack and understand what social workers were actually doing. 'Casework' was evaluated as if it was one uniform activity, stable and consistent like a standardised 'treatment'. This was a weakness in the area of understanding, which left open the possibility (recognised by Fischer) that outcomes that showed no overall benefit might be concealing the fact that some practitioners were doing beneficial work, but that their effect was being cancelled out by the poorer work of others so that aggregated effects showed no

significant differences from control groups that received no service. (An interesting study in Britain around the same time showed almost the opposite design: clients of a family service agency were interviewed to find out what they thought of the service, using a careful qualitative approach but with little attempt at measurement or comparison (Mayer and Timms, 1970). The clients, who tended to see their problems in a practical way, often did not understand what the caseworkers were trying to do: why so many questions about early childhood?)

Some similar problems were emerging in research on probation services. For example, an early study of the results of probation by Radzinowicz (1958), who was a strong advocate of probation, found that reconviction rates looked promising but included no comparison with similar offenders receiving different sentences. When Wilkins (1958) published a similar study but included relevant comparisons, people sentenced to probation had outcomes no better than those receiving other sentences. There are clear parallels in this respect with American social work research. The Home Office Research Unit also carried out, throughout the 1960s and early 1970s, a series of carefully designed and methodologically resourceful descriptive studies of probation aimed largely at developing empirical classifications of probationers and their problems. A very detailed study by Martin Davies (1974) attempted to describe the impact of probation on the social environment of probationers, and argued that only a properly controlled comparative study could show conclusively whether probation was having a positive impact. Such a study was in fact under way, known as IMPACT (Intensive Matched Probation and After-Care Treatment: Folkard et al., 1976): probationers were randomly assigned to normal or 'intensive' caseloads, and subsequent reconviction rates were compared.

The result was a slight but non-significant difference in favour of the control cases: more probation input did not seem to lead to better results. The only group that appeared to benefit from smaller caseloads was a fairly small number of offenders with high self-reported problems and low 'criminal tendencies', who were not very representative of offenders in general. The greater responsiveness of this group might tell us something about the methods officers were using in the extra time made available by lower caseloads: perhaps they were offering a form of counselling, which might help this group more than it helped others. Unfortunately, we know little about what the officers were actually doing: understanding of the process is missing, so that measurement and comparison can only give us the results of an input about which all we know is that some people received more of it than others. We cannot tell what methods were in use or how well they were being implemented. This also represented a missed opportunity to start a fuller investigation of the impact of caseload size on effectiveness: if lower caseloads did not necessarily improve results, there might still be threshold effects leading to poorer results when caseloads are simply too high to allow adequate individual attention, but this common-sense expectation, which most practitioners would support, has never been fully tested.

As in broader social work research, findings of 'no difference' were the norm at this time. There were occasional exceptions, such as Margaret Shaw's (1974) study of pre-release help to prisoners, in which those who were randomly allocated to receive more attention from prison welfare officers were reconvicted less, but there was little practical

follow-up of this finding until the development of 'resettlement' services some decades later. The general picture, in the USA as well as Britain, seemed to be one of failure, and this was summed up by Robert Martinson (1974) in his unauthorised, over-simplified but highly influential summary of the large research review of the effectiveness of rehabilitation carried out for the New York State government (Lipton et al., 1975). Martinson's (1974: 25) conclusion that 'the rehabilitative efforts that have been reported so far have had no appreciable effect on recidivism', although not a fully accurate summary of the review, was widely reported as meaning 'nothing works'. Although not all criminologists accepted this, it had a political impact, particularly in the Anglophone world where many politicians were looking for justifications for reductions in public spending. In Britain, the director of research in the Home Office summed up as follows:

> Penal 'treatments', as we significantly describe them, do not have any reformative effect . . . The dilemma is that a considerable investment has been made in various measures and services . . . Are these services simply to be abandoned on the basis of the accumulated research evidence? Will this challenge evoke a response . . . by the invention of new approaches and new methods? (Croft, 1978: 4)

What happened to 'new approaches and new methods'

After IMPACT official research on the effectiveness of probation virtually ceased in England and Wales for about 20 years. The 'new approaches and new methods' did gradually appear (Vanstone, 2004) but were mostly not systematically evaluated until much later. In the meantime, rehabilitative criminal justice was overshadowed for a while by the 'justice model' of desert-based proportional sentencing (Hood, 1974; Von Hirsch, 1976) and the work of probation and youth justice was increasingly concentrated on creating opportunities for diversion, or 'alternatives to custody'. If their methods could not reliably change people's behaviour, at least they could use their role in Court to influence decisions and to encourage the use of sentences which were seen as less harmful (and usually cheaper). Such approaches attracted support from criminologists (e.g. Bottoms and McWilliams, 1979) and eventually from Government, which articulated a role for the Probation Service in encouraging the use of non-custodial penalties instead of prison (Home Office, 1984). This strategy worked particularly well in juvenile justice (Rutherford, 1986). However, too much weight was given during this period to the 'nothing works' research, which was actually quite limited in one main respect. Although often strong on measurement and comparison, it tended to be weak in its understanding of inputs: 'probation' or 'prison' were seen as treatments in their own right, rather than needing to be unpacked to see what different inputs were actually being offered by a range of practitioners, and how well they were being delivered.

Although some research on probation's effectiveness continued outside Government in England and Wales (for example Raynor, 1988; Roberts, 1989), the next major steps in research on rehabilitation did not come until research began to benefit from better understanding and control of inputs. Correctional researchers who had never accepted Martinson's verdict began to carry out meta-analyses looking at the

characteristics of different programmes and regimes for offenders to see which were more often associated with positive outcomes. The most influential of these (Andrews et al., 1990) combined understanding of inputs with a strong focus on measurement and comparison, and proposed what became the most influential recent approach to rehabilitation, the Risk-Need-Responsivity or RNR model (Bonta and Andrews, 2017). Other meta-analyses reached broadly similar conclusions (e.g. Lipsey, 1992; Lösel, 1995; McGuire, 2002; Redondo et al. 2002) and a Scottish research review (McIvor, 1990) helped to raise awareness of this kind of work in Britain, as did a series of 'What Works' conferences (McGuire, 1995). The new focus on understanding and describing the service that was actually provided, and the explicit aim of distinguishing between effective and ineffective practice in order to encourage the former, led to a number of innovations and in particular to the development of structured group programmes using cognitive-behavioural methods. These aimed to ensure the right inputs from staff by providing detailed manuals and training, and they emphasised programme integrity, that is, delivery as designed.

One of the first programmes of this kind to be tried in probation in Britain, and the first to be thoroughly evaluated, was the Reasoning and Rehabilitation programme developed in Canada (Ross et al., 1988) and introduced in a Welsh probation area (Raynor and Vanstone, 1996, 1997; Raynor 1998). Unlike some later, larger evaluations (e.g. Hollin et al., 2004), the research on this programme used not only measurement and comparison, with modestly positive findings from a two-year reconviction follow-up, but also used qualitative approaches, interviewing all programme graduates and a number of staff, and documenting the implementation process through participant observation. Some of the lessons learned from this study about, for example, listening to probationers, taking time to do implementation properly, and involving staff through thorough consultation, seemed later to be forgotten: there was a centrally-driven rush to roll out programmes on a massive scale to take advantage of the short-term funding available in the Government's Crime Reduction Programme from 1999 to 2002 (Maguire, 2004; Raynor, 2004).

The early results of this huge effort were not as good as had been hoped or expected, with many problems of implementation including poor selection and poor retention of programme participants (Hollin et al., 2004). Thus, the overall message so far seems to be that group programmes, if properly designed, targeted and delivered, and supported where necessary by appropriate individual supervision, can make a useful contribution to the effectiveness of probation services. It is, however, very unfortunate that rushed implementation during the Crime Reduction Programme caused some front-line staff to be suspicious and resentful of 'what works' (Raynor, 2004). The associated research, which produced some positive results and led to many ideas about possible improvements, shows the benefits of a three-legged approach combining measurement and comparison of outcomes with a degree of clarity about inputs, which were defined by the programme designs and manuals at least to the extent of knowing what staff were *meant* to be doing. Many programmes also benefited from analysis of video recording of programme sessions to check integrity of delivery. For the first time this gave researchers a clearer grasp of what inputs were likely to be producing the measured outputs.

Shifting the focus from programmes to skills

In reality, probation has always depended more on individual supervision than on group programmes, and it continues to do so. The next step in understanding the inputs from practice dates mainly from the early years of the current century. Important precursors were the work of Chris Trotter in Australia in the 1990s (1993, 1996) and the recognition of 'Core Correctional Practices' in Canada (Andrews and Kiessling, 1980; Dowden and Andrews, 2004). Researchers interested in skills carried out a number of studies that, in spite of some differences of method and focus, were all concerned to study the impact of better practice skills. Among the best known of these have been Bonta's STICS study in Canada, which used a random allocation design to compare reconviction rates after supervision by officers who had received additional training in evidence-based practice skills with reconvictions after supervision by officers who had not received the training (Bonta et al., 2011); Trotter's continuing series of studies of the impact of particular practice skills (e.g. Trotter 2013, 2017); Taxman's (2008) study of the effects of training in 'proactive community supervision', and the STARR study (Robinson et al., 2012) that looked at the impact of skills on re-arrest rates. The quasi-experimental Jersey Supervision Skills Study (JS3) identified a range of skills used by probation staff in videotaped interviews and found significantly lower reconviction rates among people supervised by more skilled staff (Raynor et al., 2014). The differences in reconviction rates found in these studies are substantial, comparable with or greater than those typically reported in programme evaluations: for example, 21 percentage points in STICS, 14 in STARR, 32 in JS3.

A recent meta-analysis of skills-based research (Chadwick et al., 2015) reports on a number of studies comparing different levels of skill in supervision, which consistently show that more skilled supervision is more effective. In addition, other recent work suggests that successful implementation of initiatives for improvement in skills may depend on good management and appropriate agency culture (Bonta et al., 2013). In England and Wales the SEEDS programme (Skills for Effective Engagement, Development and Supervision: Rex and Hosking, 2014) aimed at a similar effect through staff training, and although the eventual outcomes are unclear, supervision by SEEDS-trained officers has been shown to result in slightly higher levels of compliance (Sorsby et al., 2017). The initiative seems to have been welcomed by staff and managers and SEEDS-based practices are still continuing in some places. Similarly, in Jersey the probation staff have themselves devised a process of staff development using the research instruments from the JS3 study to assess and discuss each other's interviews, which they record on video. It appears that this kind of research, combining measurement and comparison with an informed understanding of what practitioners are actually doing, has begun to illuminate a significant part of the probation service's input into supervision and is readily translatable into training initiatives that have the potential to improve practice. There is much more work to be done in this area, and it appears likely that the most productive approaches to future probation research will combine measurement and comparison with a detailed understanding of what practitioners are doing, and why. In the meantime, we have come a very long way from 'nothing works', and the three-legged approach to research methods holds plenty of promise for the future.

The next steps?

Although in principle practitioners have much to gain from applying the rationale and methods of evaluative research, these are not always welcomed. Ideally, practice is embedded in a 'culture of curiosity' (Raynor and Vanstone, 2001) in which practitioners want to know if what they are doing is getting results and how they might improve them. Evaluation is part of evidence-based practice, and evidence-based practice works best when practitioners understand it as something that can help them to achieve the outcomes that they look for in their work. When evaluation is seen as a management tool to increase control, or as something done by researchers for their own purposes or careers, or simply as lacking relevance to day-to-day work, it tends to be resisted. In England and Wales, there is currently a high level of insecurity, anxiety and unpredictable organisational change in probation services, which sometimes leads to anxiety about how research might be used. However, it should be clear from the examples outlined earlier in this article that in favourable circumstances such obstacles can be overcome if evaluation is rooted in an understanding of practice and a shared goal of effective service. As we have seen, successful evaluations tend to combine qualitative and quantitative approaches: qualitative, to understand the processes and perceptions and goals that point the way to what is worth measuring, and quantitative because without measurement there is no basis for comparison, and comparison is fundamental to the question of what works better or worse than something else.

Finally, it is clear that we are not yet looking productively at everything that matters. We have learned to measure some aspects of service quality, but others remain to be developed. Until recently, it was unusual to include the quality and behaviour of organisations in evaluations of practice, and the approaches needed to begin to do this depend heavily, so far, on evaluation instruments developed in Canada (such as the Correctional Programs Assessment Inventory) that may not be a perfect fit in other organisational cultures (though experience so far suggests that the fit is mostly quite good). Experience also suggests that the attempt to identify good services could make more use of two fields of enquiry which are not yet integrated into evaluation research. First, and perhaps more immediately accessible, is the understanding gained from service users themselves and articulated within studies of desistance, which draw on people's own accounts and narratives of how they came to desist from offending.

Studies of desistance in Britain have tended to rely heavily on qualitative methods, mostly interviews, unlike some American studies (e.g. Laub and Sampson, 2003) that also use statistical analysis. (An exception is the Sheffield Desistance Study (Bottoms and Shapland, 2011), which produced a large amount of quantitative data, but this has been unusual in British writing about desistance.) Qualitative studies contribute to understanding the desistance process and restoring service users to their rightful place at the centre of that process. This has undoubtedly helped some practitioners to think about the nature and aims of their work. The strengths of British desistance research have been in the area of understanding rather than measurement or comparison, and this has so far tended to limit its contribution to the evaluation of probation work (which, to be fair, was not its main aim). In the decade since McNeill suggested a 'desistance paradigm' for probation practice (McNeill, 2006), desistance scholars have produced little in the way

of specific guidance for probation practice; this has not been their role, and it is one they have explicitly resisted (Weaver and McNeill, 2010). Sometimes their work has been presented as an alternative to 'what works' (Farrall et al., 2014; see also McNeill et al., 2015) rather than as a complementary perspective (Ward and Maruna, 2007).

One presumably unintended consequence has been that some practitioners have seen the desistance perspective as an ally against managerialist attempts to impose 'what works' models, and therefore as an endorsement of existing practice. This limits its potential influence on practice: for example, consistent and supportive relationships between service users and supervisors are seen as a useful aid to desistance, and suggest an obvious overlap with thinking about 'core correctional practices' which aims to address the same issues, but little work has yet been done to link these two bodies of work because they belong to different research traditions. (Some recent work by Kirkwood (2016) on an interactional approach to desistance is an exception to this rule, and it is hoped that further analysis of recorded interviews from Jersey and Australia will help to show more detail of how interaction and communication with probation staff can support beliefs and behaviour consistent with desistance.) Overall, however, most scholars of desistance have preferred not to engage in quantitative procedures such as measurement and comparison, and this has limited the capacity of this body of work to accumulate reliable knowledge of the kind that is most useful in evaluation research. This is not because evaluation researchers are not interested in qualitative methods: early 'what works' studies made extensive use of qualitative interviewing to elicit service users' perspectives (Raynor and Vanstone, 1997). However, such attempts have not yet been much influenced by desistance research, and there is an agenda for future work here.

The second set of issues that arguably should be taken more into account in evaluation research concerns the influence of the policy context on practice. Briefly, the dramatic policy shifts of recent years have included the nationalisation of formerly local services in 2001, the establishment of NOMS (the National Offender Management Service) in 2004 and the privatisation of most of the Probation Service's work in 2015, which is described and discussed further below. NOMS itself has been replaced by a new Prison and Probation Service, HMPPS; is this because NOMS was a failure, or a success, or neither? These changes cannot have failed to affect the quality of practice in a variety of ways. However, they also bring into question the role and consequences of politics.

Evaluating criminal justice in post-truth Britain

As suggested in the preceding section, evaluation researchers who study probation may need to pay more attention to the wider social policy contexts in which their work is located. This can touch on politically sensitive issues: for example, desistance studies point clearly to difficulty in finding employment as an obstacle to desistance (Bottoms and Shapland, 2011), and levels of unemployment reflect wider social structures and, in part, political decisions. More broadly, comparative research on penal systems shows that some societies are consistently more punitive than others (Cavadino and Dignan, 2006) and that variations in the use of imprisonment can be linked to social inequality (Wilkinson and Pickett, 2009): societies with greater

inequality of income tend to make proportionately more use of imprisonment than more equal societies. Britain has been becoming more unequal during recent decades, when the prison population has also been rising. This context has an impact on what penal policy can achieve and on the opportunities for service development, and needs to be taken into account in thinking about the actual and potential effectiveness of services. Social and political context also affect the value attached to evaluative research and the extent of its influence on policy.

Finding evidence is not in itself the whole answer: persuading people that it is in their interests to pay attention to it is another challenge, and the nature of this challenge, and the uses and meanings of evidence, change over time. Before the 1990s, the limited research available had little impact on policy and practice in England and Wales (though an early attempt by McGuire and Priestley (1985) to promote evidence-based practice was widely read). The Home Office had largely given up on the search for effectiveness and showed limited interest in research from overseas, or even from Scotland. This changed when probation's search for 'what works' coincided with the dissemination (often by psychologists) of new research on effective practice. The New Labour government elected in 1997 promised evidence-based policy-making and was prepared to invest in the 'Pathfinder' experiments although, as described above, the time-scale was too short and the implementation too uneven to deliver the kind of results that probation's leaders hoped for. The peak years for evidence-based development in probation lasted from 1997 to about 2003; by then, politicians who had learned to see evidence as a useful resource began to look for evidence to support preferred policies rather than choosing policies to fit the evidence.

Within criminal justice evidence has been used to support practice developments such as the accreditation of programmes, but major policy changes are based on little or no evidence, or on reports specially commissioned to support them. An early example was the abolition of consent to probation in 1997 (see Raynor, 2014); more recent examples are the abolition of probation orders in 2000; the creation of the national Probation Service and marginalization of the judiciary in 2001; the creation of the National Offender Management Service in 2004 following a report by a businessman (Carter, 2003); the inclusion of a 'punitive requirement' in every community sentence in 2013, in line with an earlier report by a civil servant (Casey, 2008), and finally, in the clearest example of evidence-free policy-making, the decision to break up and sell off the majority of the Probation Service to private providers in 2014–15, the creation of Community Rehabilitation Companies and the attempt to incentivise better performance through 'Payment By Results' (PBR) (Ministry of Justice, 2013).

Originally a thorough programme of piloting was planned to precede the implementation of the new policies, in line with the original plans set out in 'Transforming Rehabilitation' (Ministry of Justice, 2013). However, a new Justice Secretary, Christopher Grayling, decided to accelerate the process by abandoning the pilots which were intended to provide guidance on feasibility and implementation. (This was probably in order to ensure that the changes could be rushed through before the General Election of 2015, to avoid the possibility that an incoming Government might not proceed with the changes – just as the break-up and privatisation of Britain's railways was rushed into effect before the General Election of 1997.) Instead, two small pilots of after-care for short-term

prisoners, which actually measured the impact of providing an after-care service rather than the impact of PBR (Disley et al. 2015; Pearce et al. 2015), were wrongly claimed by politicians as evidence of the success of PBR. As Grayling (2014) explained to a Policy Exchange meeting in 2014, 'I don't believe you need to pilot professional and operational freedom'. Evidence was unnecessary to support a policy guided simply by ideology and political conviction. Since the privatisation, almost all the evidence collected by auditors and independent inspectors (e.g. National Audit Office, 2016; HM Inspectorate of Probation, 2016a, 2016b, 2016c) shows that it has made community sentences less reliable and less safe, and has done little to create the new resettlement services for short-term prisoners that were part of the rationale for the policy. In addition, at the time of writing only two of the 21 CRCs were on track to meet the performance threshold for PBR (Ministry of Justice, 2018).

Grayling's evidence-free privatisation was an early example of what, during 2016, would become known in Britain and the USA as 'post-truth' politics. A tendency to shape the evidence to support a pre-existing policy line was not new: for example, it was a feature of Tony Blair's foreign policy leading up to the invasion of Iraq in 2003, and the dangers of this approach were thoroughly exposed by the Chilcot enquiry into the invasion and its aftermath (Chilcot, 2016). However, recent developments have gone beyond simply shaping the evidence to managing without it altogether, or to a complete disregard for facts. In 2016, a plot by a group of 'Eurosceptic' Conservative politicians in England to unseat their party leader led to a referendum on continued membership of the European Union, in which the successful campaign for a 'leave' vote was based on deliberately misleading propaganda, taking full advantage of modern means of mass communication. Inconvenient facts were ignored, and one senior 'leave' campaigner, Michael Gove (for a short while Christopher Grayling's successor as Justice Secretary) claimed that 'The British people have had enough of experts' (reported in the *Financial Times*, 3 June 2016). In the USA, a similar populist rhetoric was a feature of Donald Trump's successful election campaign, and commentators drew parallels with the disastrous populist movements of the 1930s that used similar appeals to popular prejudice based on widely disseminated and emotive falsehoods. The modern equivalents include 'fake news' and so-called 'alternative facts' (as proposed by Kellyanne Conway, a senior aide to President Trump, in a hilarious press briefing reported in the *Guardian*, 22 January 2017, and all other reputable media).

In short, the 21st century so far shows a progression from being *guided* by evidence, to using evidence as a *resource to support* policy decisions already made, to *creating* evidence to support policies, and eventually to dispensing with evidence altogether. The erosion of the traditional basis of probation has been accelerated by a series of Home Secretaries and Justice Secretaries who seemed over-concerned with maintaining a reputation for being 'tough on crime', from Michael Howard through to Jack Straw, David Blunkett, John Reid and finally Grayling. Some of these gentlemen may have felt that too much support for probation would compromise their tough image. As a result, the majority of probation staff trying to contribute to rehabilitation find themselves in a legislative and organisational environment that is not helpful, fails to support their work properly and in many cases exposes them to the risk of redundancy as CRCs try to protect their profits by reducing their wage bill.

The emerging style of politics (perhaps not so much post-truth as post-Enlightenment) does not provide a promising environment for evaluation research or evidence-based policy in those countries where it is prevalent. However, it is not prevalent everywhere: probation research is flourishing in Europe, and in some of the devolved jurisdictions within the British Isles such as Scotland, Northern Ireland and the Channel Islands. Even within England and Wales there are examples of local research that engages directly with service providers. When post-truth policies fail, factual research on how to make probation more effective will be needed to support the necessary evidence-based reform. In the meantime, in Britain we face continuing uncertainty about how best to finance our research, where to place it and, in particular, how to communicate it effectively to people who may put it to some use.

Funding

Note

This article was originally presented as a paper at the meeting of the European Society of Criminology Working Group on Community Sanctions, Barcelona, 18–19 May 2017. It is largely based, with the permission of the editors, on a book chapter: 'The search for impact in British probation: From programmes to skills to implementation', which appears in Ugwudike P, Raynor P and Annison J (eds) *Evidence-Based Skills in Criminal Justice: International Research on Supporting Rehabilitation and Desistance*, Bristol: Policy Press, published in December 2017. It also includes some material that first appeared in another book chapter: Raynor P (2016) Effective probation in England and Wales: The rise and fall of evidence. In: Vanstone M and Priestley P (eds) *Probation and Politics*. Basingstoke: Palgrave. Parts of this paper also formed the basis of a keynote address to the World Congress of Probation in Tokyo, September 2017.

ORCID iD

Peter Raynor [iD] https://orcid.org/0000-0002-2628-3489

References

Andrews DA and Kiessling J (1980) Program structure and effective correctional practices: A summary of the CaVIC research. In: Ross R and Gendreau P (eds) *Effective Correctional Treatment*. Toronto: Butterworth, pp. 441–463.

Andrews DA, Zinger I, Hoge RD et al. (1990) Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology* 28(3): 369–404.

Bonta J and Andrews DA (2017) *The Psychology of Criminal Conduct*, 6th edn. Abingdon: Routledge.

Bonta J, Bourgon G, Rugge T et al. (2011) An experimental demonstration of training probation officers in evidence-based community supervision. *Criminal Justice and Behavior* 38(11): 1127–1148.

Bonta J, Bourgon G, Rugge T et al. (2013) Taking the leap: From pilot project to wide-scale implementation of the Strategic Training Initiative in Community Supervision (STICS). *Justice Research and Policy* 15(1): 17–35.

Bottoms AE and McWilliams W (1979) A non-treatment paradigm for probation practice. *British Journal of Social Work* 9(2): 159–202.

Bottoms AE and Shapland J (2011) Steps towards desistance among male young adult recidivists. In: Farrall S, Hough M, Maruna S and Sparks R (eds) *Escape Routes: Contemporary Perspectives on Life after Punishment*. Abingdon: Routledge, pp. 43–80.

Carter P (2003) *Managing Offenders, Reducing Crime: A New Approach*. (Correctional Services Review.) London: Home Office.

Casey L (2008) *Engaging Communities in Fighting Crime: A Review*. London: Cabinet Office.

Cavadino M and Dignan J (2006) *Penal Systems: A Comparative Approach*. London: SAGE.

Chadwick N, Dewolf A and Serin R (2015) Effectively training community supervision officers: A meta-analytic review of the impact on offender outcome. *Criminal Justice and Behavior* 42(10): 977–989.

Chilcot J (Chair) (2016) *The Report of the Iraq Inquiry*. London: HMSO.

Croft J (1978) *Research in Criminal Justice*. Research Study 44. London: HMSO.

Davies M (1974) *Social Work in the Environment*. Research Study 21. London: HMSO.

Disley E, Giacomantonio C, Kruithof K and Sim M (2015) *The Payment by Results Social Impact Bond pilot at HMP Peterborough: Final process evaluation report*. London: Ministry of Justice.

Dowden C and Andrews D (2004) The importance of staff practice in delivering effective correctional treatment: A meta-analysis. *International Journal of Offender Therapy and Comparative Criminology* 48(2): 203–214.

Farrall S, Hunter B, Sharpe G and Calverley A (2014) *Criminal Careers in Transition*. Oxford: Oxford University Press.

Fischer J (1973) Is casework effective? A review. *Social Work* 18(1): 5–20.

Fischer J (1976) *The Effectiveness of Social Casework*. Springfield: C. C. Thomas.

Folkard MS, Smith DE and Smith DD (1976) *IMPACT Volume II: The results of the experiment*. Research Study 36. London: HMSO.

Grayling C (2014) Speech to Policy Exchange, March.

Grey A and Dermody H (1972) Reports of casework failure. *Social Casework*, November: 534–543.

Harper G and Chitty C (2004) *The impact of corrections on re-offending: A review of 'what works*. Research Study 291. London: Home Office.

HM Inspectorate of Probation (2016a) *Transforming Rehabilitation: Early Implementation 5*. Manchester: HMIP.

HM Inspectorate of Probation (2016b) *An Inspection of Through-the-Gate Resettlement Services for Short-Term Prisoners*. Manchester: HMIP.

HM Inspectorate of Probation (2016c) *The Effectiveness of Probation Work in the North of London*. Manchester: HMIP.

Hollin C (2008) Evaluating offending behaviour programmes: Does only randomization glister? *Criminology and Criminal Justice* 8(1): 89–106.

Hollin C, Palmer E, McGuire J et al. (2004) *Pathfinder Programmes in the Probation Service: A Retrospective Analysis*. Home Office Online Report 66/04, London: Home Office.

Home Office (1959) *Penal Practice in a Changing Society*, Cmnd. 645, London: HMSO.

Home Office (1984) *Probation Service in England and Wales: Statement of National Objectives and Priorities*. London: Home Office.

Hood R (1974) *Tolerance and the Tariff*. London: NACRO.

Kirkwood S (2016) Desistance in action: An interactional approach to criminal justice practice and desistance from offending. *Theoretical Criminology* 20(2): 222–237.

Laub J and Sampson R (2003) *Shared Beginnings, Divergent Lives: Delinquent boys to age 70*. Cambridge, MA: Harvard University Press.

Lipsey M (1992) Juvenile delinquency treatment: A meta-analytic enquiry into the variability of effects. In: Cook T, Cooper H, Cordray DS et al. (eds) *Meta-Analysis for Explanation: A Case-book*. New York: Russell Sage, pp. 83–127.

Lipton D, Martinson R and Wilks J (1975) *The Effectiveness of Correctional Treatment*. New York: Praeger.

Lösel F (1995) The efficacy of correctional treatment: A review and synthesis of meta-evaluations. In: McGuire J (ed.) *What Works: Reducing Reoffending*. Chichester: Wiley, pp 79–111.

McCord J (1978) A thirty-year follow-up of treatment effects. *American Psychologist* 33(3): 284–289.

McGuire J (2002) Integrating findings from research reviews. In: McGuire J (ed.) *Offender Rehabilitation and Treatment*. Chichester: Wiley, pp. 3–38.

McGuire J (ed.) (1995) *What Works: Reducing Reoffending*. Chichester: Wiley.

McGuire J and Priestley P (1985) *Offending Behaviour: Skills and Stratagems for Going Straight*. London: Batsford.

McIvor G (1990) *Sanctions for Serious or Persistent Offenders*. Stirling: Social Work Research Centre.

McNeill F (2006) A desistance paradigm for offender management. *Criminology and Criminal Justice* 6(1): 39–62.

McNeill F, Raynor P, Herzog-Evans M and Durnescu I (2015) Rethinking what helps? Beyond probation and desistance. *Howard Journal* 54(3): 311–317.

Maguire M (2004) The Crime Reduction Programme in England and Wales: Reflections on the vision and the reality. *Criminal Justice* 4(3): 213–37.

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

Matza D (1969) *Becoming Deviant*. Englewood Cliffs: Prentice-Hall.

Mawby R and Worrall A (2013) *Doing Probation Work*. Abingdon: Routledge.

Mayer J and Timms N (1970) *The Client Speaks*. London: Routledge.

Meyer H, Borgatta E and Jones W (1965) *Girls at Vocational High*. New York: Russell Sage Foundation.

Mills CW (1943) The professional ideology of social pathologists. *American Journal of Sociology* 49(2): 165–180.

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

Ministry of Justice (2018) *Final and Interim Proven Reoffending Statistics for the Community Rehabilitation Companies and the National Probation Service*. London: Ministry of Justice, 25 January.

National Audit Office (2016) *Transforming Rehabilitation*. London: National Audit Office.

Pearce S, Murray D and Lane M (2015) *HMP Doncaster Payment By Results Pilot: Final process evaluation report*. London: Ministry of Justice.

Powers E and Witmer H (1951) *An Experiment in the Prevention of Delinquency*. New York: Columbia University Press.

Radzinowicz L (ed) (1958) *The Results of Probation*. A report of the Cambridge Department of Criminal Science. London: Macmillan.

Raynor P (1988) *Probation as an Alternative to Custody*. Aldershot: Avebury.

Raynor P (1998) Attitudes, social problems and reconvictions in the STOP probation experiment. *Howard Journal* 37(1): 1–15.

Raynor P (2004) The probation service 'pathfinders': Finding the path and losing the way? *Criminal Justice* 4(3): 309–325.

Raynor P (2014) Consent to probation in England and Wales: How it was abolished, and why it matters. *European Journal of Probation* 6(3): 296–307.

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(4): 272–284.

Raynor P and Vanstone M (1997) *Straight Thinking On Probation (STOP): The Mid Glamorgan Experiment*. Probation Studies Unit Report No. 4. Oxford: University of Oxford Centre for Criminological Research.

Raynor P and Vanstone M (2001) Straight Thinking On Probation: Evidence-based practice and the culture of curiosity. In: Bernfeld G, Farrington D and Leschied A (eds) *Offender Rehabilitation in Practice*. Chichester: Wiley, pp. 189–203.

Raynor P and Vanstone M (2015) Moving away from social work and half way back again: New research on skills in probation. *British Journal of Social Work* 46(4): 1131–1147.

Raynor P, Ugwudike P and Vanstone M (2014) The impact of skills in probation work: A reconviction study. *Criminology and Criminal Justice* 14(2): 235–249.

Redondo S, Sanchez-Meca J and Garrido V (2002) Crime treatment in Europe: A review of outcome studies. In: McGuire J (ed.) *Offender Rehabilitation and Treatment*. Chichester: Wiley, pp. 113–141.

Rex S and Hosking N (2014) Supporting practitioners to engage offenders. In: Durnescu I and McNeill F (eds) *Understanding Penal Practice*. Abingdon: Routledge, pp. 271–280.

Roberts C (1989) *Hereford and Worcester Probation Service Young Offender Project: First evaluation report*. Oxford: Department of Social and Administrative Studies.

Robinson C R, Lowenkamp CT, Holsinger AM et al. (2012) A random study of Staff Training Aimed at Reducing Re-arrest (STARR): Using core correctional practices in probation interactions. *Journal of Crime and Justice* 35(2): 167–188.

Robinson G, Priede C, Farrall S et al. (2014) Understanding quality in probation practice: Frontline perspectives in England and Wales. *Criminology and Criminal Justice* 14(2): 123–142.

Ross RR, Fabiano EA and Ewles CD (1988) Reasoning and rehabilitation. *International Journal of Offender Therapy and Comparative Criminology* 32(1): 29–35.

Rutherford A (1986) *Growing Out of Crime*. Harmondsworth: Penguin.

Shaw M (1974) *Social Work in Prisons*. Research Study 22. London: HMSO.

Sorsby A, Shapland J and Robinson G (2017) Using compliance with probation supervision as an interim outcome measure in evaluating a probation initiative. *Criminology and Criminal Justice* 17(1): 40–61.

Taxman FS (2008) No illusions: Offender and organizational change in Maryland's proactive community supervision efforts. *Criminology and Public Policy* 7(2): 275–302.

Trotter C (1993) *The Supervision of Offenders – What Works? A Study Undertaken in Community Based Corrections, Victoria*. Melbourne: Social Work Department, Monash University and Victoria Department of Justice.

Trotter C (1996) The impact of different supervision practices in community corrections. *Australian and New Zealand Journal of Criminology* 28(2): 29–46.

Trotter C (2013) Effective supervision of young offenders. In: Ugwudike P and Raynor P (eds) *What Works in Offender Compliance: International Perspectives and Evidence-Based Practice.* Basingstoke: Palgrave Macmillan, pp. 227–241.

Trotter C, Evans P and Baidawi S (2017) The effectiveness of challenging skills in work with young offenders. *International Journal of Offender Therapy and Comparative Criminology* 61(4): 397–412.

Vanstone M (2004) *Supervising Offenders in the Community: A History of Probation Theory and Practice*. Aldershot: Ashgate.

von Hirsch A (1976) *Doing Justice: The Choice of Punishments*. Report of the Committee for the Study of Incarceration. New York: Hill and Wang.

Wallace D (1967) The Chemung County evaluation of casework service to dependent multiproblem families: Another problem outcome. *Social Service Review* 41(4): 379–389.

Ward T and Maruna S (2007) *Rehabilitation*. Abingdon: Routledge.

Weaver B and McNeill F (2010) Travelling hopefully: Desistance theory and probation practice. In: Brayford J, Cowe F and Deering J (eds) *What Else Works? Creative Work with Offenders*. Cullompton: Willan, pp. 36–60.

Wilkins LT (1958) A small comparative study of the results of probation. *British Journal of Delinquency* 8(3): 201–209.

Wilkinson R and Pickett K (2009) *The Spirit Level*. London: Allen Lane.

Wootton B (1959) *Social Science and Social Pathology*. London: Allen and Unwin.

Author biography

Peter Raynor is emeritus research professor in the Department of Criminology, Swansea University, in Wales, UK. A former probation officer, he has been carrying out and publishing research on probation and related topics since the 1970s. Recent books include *Rehabilitation, Crime and Justice* (with Gwen Robinson, Macmillan, 2009); *Offender Supervision* (with Fergus McNeill and Chris Trotter, Willan, 2010); *What Works in Offender Compliance* (with Pamela Ugwudike, Macmillan, 2013) and *Reintegrative Justice in Practice* (with Helen Miles, Ashgate, 2014). He is a member of the Correctional Services Advisory and Accreditation Panel for England and Wales and of the National Offender Management Service's Evidence Review Panel, and a Fellow of the Academy of Social Sciences. [Email: p.raynor@swansea.ac.uk]