A What Works Centre for Probation: Challenges and possibilities

Michael Sanders
What Works Centre for Children’s Social Care, UK; King’s College London, UK

Louise Jones and Eleanor Briggs
What Works Centre for Children’s Social Care, UK

Abstract
The What Works Movement in the UK Government has seen the establishment of 12 centres to focus on evidence-based policy in different domains. In this paper, we present the challenges and opportunities posed by a What Works Centre (WWC) for Probation, based on our prior experience of establishing WWCs in other areas. Although there are legitimate and substantial challenges to some of the methodological approaches of ‘What Works’, we conclude that Probation is in an unusually strong starting position for such a centre to thrive.

Keywords
What Works, research methods, organisational change, organisational culture, practitioners

Introduction
Is there a space for a What Works Centre in the field of Probation in the UK? This was the broad topic of a roundtable hosted on the 26th November 2020 by Kent, Sussex and Surrey Community Rehabilitation Company (KSSCRC). The roundtable,
which attracted an expert panel as well as knowledgeable and curious participants, considered a wide-ranging discussion of the topic, and we were pleased to offer the perspective of a friendly outsider at What Works for Children’s Social Care.

There are currently 12 ‘What Works Centres’ (WWCs) in the UK, 13 of which have been established in the last decade starting with the establishment of the Education Endowment Foundation in 2012. Although these centres have very different focuses – ranging from early years to ageing, via crime reduction and financial capability – they share three core purposes:

1. **Collation** of the existing evidence – drawing together what is known already from the academic and ‘grey’ literature;
2. **Creation** – filling in the gaps of evidence that are found through collation, by either conducting or commissioning original research; and
3. **Translation** – making sure that all of this evidence is presented in an accessible and easy to access format and trying to make sure that both policy and practice come to reflect this progress.

The centres are also ultimately interested in the answer to a particular kind of research question; ‘What Works?’ – or, more scientifically, questions of causal inference. These research questions are typically quite instrumental or practical in their nature and can often be boiled down to ‘If I do X, what will happen to some outcome Y that I care about?’

In answering these questions, the What Works movement most commonly (although far from exclusively) turns to Randomised Controlled Trials (RCTs). The methodology, which is borrowed from medicine where it is used routinely to test drugs – and new vaccines – is straightforward. You start with a large group of people in your sample who are eligible for your new intervention, or approach – let’s say all men under 40 who have been released from prison following a custodial sentence for domestic abuse. This large group is divided in two at random – Heads in one group, Tails in another. One group gets the new intervention you want to test – let’s say a programme of counselling support designed to reduce recidivism – and the other half gets business as usual. After an appropriate period of time, you return and see how many from each has reoffended. Because of the random assignment, we’d expect the two groups to be the same except for the new intervention, and so any differences we find between the two groups can be said to be caused by that new intervention. RCTs aren’t the only way of establishing ‘what works’, but they are the most straightforward, and the most reliable – although even among advocates of a ‘What Works’ approach, it must be acknowledged that opinions differ as to the extent to which RCTs can and should be used both between and within different fields of inquiry. A straightforward guide to conducting RCTs can be found in Haynes et al. (2012).

As we say, these research questions are practical, but they are not uncontroversial – they are often seen as overly prescriptive, overly managerial in their nature. RCTs which place a great emphasis on the ability of a researcher to control
what’s going on, need to clear a high ethical bar to take place, and risk imposing,
or assuming, an artificial order on the complexity of the real world. The ethical
considerations associated with withholding an intervention that is believed to be
beneficial, and which has strong theoretical underpinnings – from half of a group
that could benefit, cannot be overlooked; nor can concerns about participants’
ability to meaningfully consent when trials involve professionals in positions of
authority.

What Works in probation?
As was reflected by the panel and in group discussion, ‘What Works’ research
is too rarely integrated into the practice of probation professionals – either the
production of evidence, or even its use. Although this to some extent reflects the
extent of the RCT literature in the field, the picture is rosier in this regard than
many might think. Randomised trials have been conducted in probation contexts
of alcohol use screening in England (Newbury-Birch et al., 2009), Cognitive
Behavioural interventions in the UK (Pearson et al., 2016), Restorative Justice in
the UK (Shapland et al., 2008) and Motivational interviewing in Sweden
(Forsberg et al., 2011), among others. Where RCTs have been possible,
interventions have often been discrete, manualisable, and in several cases,
imported from other, more clinical fields. Quasi-experimental approaches,
which seek to statistically emulate the analysis of an RCT without randomisation
having taken place, have also been used, for example to evaluate the effective-
ness of the series of programmes that made up the Home Offices’ Pathfinder
Projects (Hollin et al., 2004). Efforts have already been taken, for example in
the ‘Reducing Reoffending’ project hosted by Manchester Metropolitan Uni-
versity, to bring together evidence in this area. This position is in many ways
stronger than that in children’s social care when we began our journey to form
our What Works Centre.

However, their use, and their deployment into practice, is clearly much more
limited. And here would be a clear benefit of a What Works Centre. The role of such
a centre, as well as working out what is known and pushing out the frontiers of
knowledge, would be to communicate – to practitioners, managers, policymakers,
and politicians – what is already known, and to help ensure that this changes the
reality on the ground.

Challenges
There are legitimate challenges to the legitimacy of a What Works Centre in pro-
bation – as with any other policy domain. However, the intelligent design of a What
Works Centre can help address many of these. Others, which are principally phi-
losophical or political, must be reconciled. Many of the most valid forms of these
objections relate to RCTs per se, and can be addressed best in that context, as we
seek to do so here.
RCTs are inherently top down, and discourage participatory research

Because an RCT requires a large sample size, and quite a bit of researcher control, it can certainly favour centralisation of control – pushing power upwards and towards the middle. This, ironically, runs the risk of stifling innovation – which will often arise not as the brainchild of a minister or an official, but because of individual practitioners, experts in doing their job, thinking of and creating a new, potentially better way, of doing some aspect of their work. However, as an individual there is a limit to the sample size you can reasonably put together. You can influence your own practice – and perhaps your team’s – but that’s not enough for an RCT. If we are to truly value relationship based, dynamic practice, then producing the standard of evidence that the What Works movement asks for through this approach is clearly not practical.

There is an easy way to avoid this, however. We must recognise that an RCT is the end of an evidence journey and not the beginning. A new idea at practice level can and should be tested first qualitatively, to understand how people relate to it, developing an understanding of the idea’s ‘theory of change’ – how it proposes to make the world different (Drucker, 1995). After this, testing whether it can be delivered at a reasonable scale can be done through a feasibility study, alongside information about how both practitioners and clients respond to it. Only when there is indicative evidence at all of the prior stages, need an RCT be done. By this stage, the idea’s originator should have been empowered or positioned to run a larger scale project themselves.

This idea, that we can begin evidence generation around even quite early-stage ideas, is at the core of our Practice in Need of Evidence (PINE) programme at What Works for Children’s Social Care. PINE combines support from our researchers with an online platform that gives practitioners free access to research tools that enable them to begin their evidence journey. Since we began PINE 12 months ago, we’ve worked with more than 20 partners in this way, on projects including early parental assessments, and a fathers’ parenting group. Our hope is that this, as well as other efforts to work with and engage the grass roots, flips the narrative described above. By democratising evidence production, we’re hoping to push the power balance away from large central agencies and towards individual practitioners with bright ideas.

RCTs cannot answer questions beyond ‘if I do X, what happens to Y?’

This is another criticism with which we are very familiar. An RCT can only answer one, quite basic question; that of causal impacts. This is indeed the crucial purpose of an RCT – they are very good for answering questions of this type, but not very much use for much else – or at least, not by themselves. But the belief that an RCT only answers this kind of question seems to be rooted in a medical model of randomised trial, which is principally concerned with the quantitative outcomes, and relies more on basic scientific research to understand mechanisms. RCTs in social policy, however, almost always include qualitative components – which aims to
understand how the intervention works (or doesn’t), and why. It can also help to give a sense of for whom it might work – which can be tested quantitatively as well.

**It is unethical to withhold the intervention from some people**

A common ethical challenge to RCTs is that it is unethical to withhold interventions from people who might benefit from them – especially when there are real world consequences of doing so, and/or the participants in the study are particularly vulnerable. Arguments of consent must be considered in line with this – and the extent to which we can truly gain consent in a situation where the state has substantial coercive power over participants. But we must, as John List (2011) argues, consider not just the cost of conducting the trial, but also the benefits. If a trial is not conducted, we lose out on the learning we might gain from it. We cannot so easily stop doing that which does not work or continue doing that which does. The trial, and the evidence it produces, can be used to powerfully make the case for an intervention to policymakers, and ensure its greater roll out – as we have seen, for example, in the recent launch of the National Tutoring Programme by the Education Endowment Foundation (EEF). This programme is a large investment (£350 million) by the Department for Education in delivering tutoring to hundreds of thousands or millions of young people that have experienced educational disadvantage as a result of the pandemic, and which follows meta-analytic research on the impacts of tutoring (D’Agostino and Harmey, 2016). The risks of withholding from a modest number of people now must be weighed against the potential future benefits.

It is difficult to overstate the extent of our ignorance about ‘What Works’ in the absence of good quality evidence. Reviewing the unbiased, independent and transparent research commissioned by the Education Endowment Foundation, we find (Sanders et al., 2020), that average effects are small, and that most interventions do not work. Similar findings can be seen in children’s social care (Fitzsimons and McCracken, 2020). The Scared Straight programme, which was widely practised for decades before a high-quality evidence review, was thought to reduce re-offending, but actually increased it – by 60% (Petrosino et al., 2000). Set against this, we must ask ourselves not ‘is it ethical to test this with a randomised trial’, but instead, ‘is it ethical not to?’

‘What Works’ leads to intervention ‘fetishisation’

This criticism was raised in the context of probation in the workshop but has been repeated elsewhere. If something is to be tested in an RCT, we must know what it is that we’re testing. The need to know this means there is a need for manualisation – we need to write down what exactly must happen – and then to a fixation with fidelity – that things are doing things in the way that they are supposed to. These questions are important, but should not be the primary focus of large scale trials. If an intervention must be adhered to perfectly to be effective, it is not, in truth, flexible. It might be possible to master the ‘deployment’ of staff in fast food, of the manner designed by the McDonald Brothers in the 1960’s (Kroc and Anders, 1987); it
might be possible to create a production line in the manufacturing of cars, a la Henry Ford (Brinkley and Brinkley, 2003). It is not possible to precisely prescribe the behaviours of professionals interacting with complex individuals (Deaton and Cartwright, 2018). Instead, we can prescribe some of the inputs; things like the training that people receive, the support that they get, and the materials that they’re given. This mix of training, support, and materials may encourage professionals to behave in a particular way, or to interact with their clients in a specific way; but it cannot guarantee that they will.

A favourite example of ours is the ‘Visible Classroom’ intervention developed by the Australian Educationalist John Hattie (Hattie, 2012). Teachers receiving the intervention are instructed to wear headsets, of the style not-quite-popularised by Britney Spears in the early 21st Century, while teaching. Their lessons were recorded, and then analysed, with information about how much time the teacher had spent speaking compared to the students then presented back to the teachers after class. The intervention is the subject of an ongoing randomised trial funded by the Education Endowment Foundation (Sanders et al., 2017). What is the trial looking at? Well, ideally it would be looking at the effect of teachers wearing headsets and receiving feedback on their teaching approach. But teachers are not soldiers – they cannot be ordered what to do – and there is a very good chance that they might feel silly doing so – at least after a couple of lessons kitted up like an early 2000s popstar in front of their teenage students. In fact, what the trial is testing is a much more realistic question – what is the effect of providing teachers with headsets, and giving them training in how to use them to receive feedback? If this works, great; if it doesn’t work, it often matters little whether this was because the headset was ineffective, or because they were just too silly to wear.

**Opportunities**

Looking beyond the challenges that are posed, there are a clear number of opportunities ahead for a What Works Centre for Probation. As the sector undergoes a large-scale shift for the second time in a decade, a practical and cultural change could emerge to help the sector and the profession hold their head high as an evidence based, research active community, and a What Works Centre could sit at the core of this. Probation, as with other aspects of the justice system, begins with a crucial advantage over many other sectors – data. Although there are things that are measured which are vitally important to understanding ‘what works’ in a probationary context, many things, including re-offending and re-conviction rates, are already accessible to outside researchers and organisations through the Ministry of Justice ‘Data Lab’, which, combined with a commitment to greater transparency and rigour, could shoot Probation to the top of the leaderboard in terms of evidence-based policy (Lyon et al., 2015).

There would certainly be a lot of support from the rest of the What Works network. What Works for Children’s Social Care, the Youth Endowment Fund, and What Works for Crime Reduction (located within the College of Policing), all work in similar areas, and could share lessons learned. It is also clear that probation is
current a missing piece of the What Works puzzle. Offending, and particularly re-offending, sits at the heart of many of the challenges facing policy, social workers, and all organisations focused on supporting young people – and yet the What Works network does not currently consider it at all. Beyond these overlaps, there is a growing desire – and a growing need – for evidence centres to be multi-sectoral in their approaches, reflecting the growth in multi-agency working across the public service. We are proud to support this through the Evidence Quarter, a physical home to seven Evidence Based policy organisations that share both space, resources, and some back office functions, located opposite the Ministry of Justice on Petty France in London. A What Works Centre for Probation would be very welcome to join us.

People working in probation services deserve a rich evidence landscape to turn to when making their decisions and shaping their practice. To see this happen, an expansion in lesser used methodologies is needed, as is the formation of a more coherent evidence journey in Probation. Professionals, and the people they serve, deserve to have their voices heard in decisions about policy and about how money is spent. A sympathetic What Works Centre, with democratisation of evidence production at its core, could help amplify these voices and, most importantly, improve outcomes.

Acknowledgement

We are grateful to the panellists and participants of the roundtable on What Works for Probation hosted by KSS CRC on 26/11/2020.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Michael Sanders https://orcid.org/0000-0002-6595-6375

Note

1. https://reducing-reoffending.uk/about.

References

Brinkley D and Brinkley D (2003) Wheels for the World: Henry Ford, His Company, and a Century of Progress, 1903–2003. New York, NY: Viking, p. xxii.
D’Agostino JV and Harmey SJ (2016) An international meta-analysis of reading recovery. Journal of Education for Students Placed at Risk (JESPAR) 21(1): 29–46.
Deaton A and Cartwright N (2018) Understanding and misunderstanding randomized controlled trials. Social Science & Medicine 210: 2–21.

Drucker PF (1995) People and Performance: The Best of Peter Drucker on Management. London: Routledge.

Fitzsimons A and McCracken K (2020) Children’s Social Care Innovative Programme Round 2 Final Report. The Department for Education.

Forsberg L, Ernst D and Farbring CÅ (2011) Learning motivational interviewing in a real-life setting: a randomised controlled trial in the Swedish Prison Service. Criminal Behaviour and Mental Health 21(3): 177–188.

Hattie J (2012) Visible Learning for Teachers: Maximizing Impact on Learning. London: Routledge.

Haynes L, Goldacre B and Torgerson D (2012) Test, Learn, Adapt: Developing Public Policy With Randomised Controlled Trials. London: Cabinet Office-Behavioural Insights Team.

Hollin C, Palmer E, McGuire J, et al. (2004) Pathfinder programmes in the probation service: a retrospective analysis. Home Office Online Report 66(04):

Kroc R and Anderson R (1987) Grinding It Out: The Making of McDonald’s. Basingstoke: Macmillan.

List JA (2011) Why economists should conduct field experiments and 14 tips for pulling one off. Journal of Economic Perspective 25(3): 3–16.

Lyon F, Gyateng T, Pritchard D, et al. (2015) Opening access to administrative data for evaluating public services: the case of the Justice Data Lab. Evaluation 21(2): 232–247.

Newbury-Birch D, Bland M, Cassidy P, et al. (2009) Screening and brief interventions for hazardous and harmful alcohol use in probation services: a cluster randomised controlled trial protocol. BMC Public Health 9(1): 418.

Pearson DA, McDougall C, Kanaan M, et al. (2016) Evaluation of the citizenship evidence-based probation supervision program using a stepped wedge cluster randomized controlled trial. Crime & Delinquency 62(7): 899–924.

Petrosino A, Turpin-Petrosino C and Finckenauer JO (2000) Well-meaning programs can have harmful effects! Lessons from experiments of programs such as Scared Straight. Crime & Delinquency 46(3): 354–379.

Sanders M, Carr D and Heal J (2017) Visible Classroom: Evaluation Protocol. Education Endowment Foundation.

Sanders M, Mitchell C and Ni Chonnaire A (2020) Effect sizes in education trials in England. Available at SSRN 3532325.

Shapland J, Atkinson A, Atkinson H, et al. (2008) Does Restorative Justice Affect Reconviction. The fourth report from the evaluation of three schemes. Ministry of Justice.