bjcj

THOUGHT PIECE

'Thought Pieces' are papers which draw on the author's personal knowledge and experience to offer stimulating and thought provoking ideas relevant to the aims of the Journal. The ideas are located in an academic, research, and/or practice context and all papers are peer reviewed. Responses to them should be submitted to the Journal in the normal way.

EFFECTIVE PROJECT DESIGN: HOW CAN VOLUNTARY AGENCIES HELP TO DEMONSTRATE THEIR IMPACT

Dan Ellingworth, Senior Research Fellow, Sheffield Hallam University

Introduction

In the highly marketised landscape of criminal justice service delivery in England and Wales, the requirement for voluntary and community sector agencies to be able to demonstrate impact has attained greater urgency. This thought piece draws together learning from the author's experience of evaluating criminal justice projects designed and delivered by voluntary and community sector agencies. Specifically it examines the centrality of project design in enabling the demonstration of impact.

Policy context

Transforming Rehabilitation (TR) (MoJ, 2013a) was launched in the United Kingdom with the intention to enable market mechanisms to produce more effective management of low to medium risk offenders managed through the new Community Rehabilitation Companies (CRCs) across England and Wales.

The stated intention was to produce "increased efficiency and new ways of working" (MoJ, 2013a) and by so doing reduce re-offending and the associated costs to society, through the market incentivisation of public, private and voluntary sector organisations. This incentivisation has been delivered through a two part payment to CRCs (one part a fee for delivery, and a second part a Payment-by-Results (PbR) element, where the fee varies according to the reduction of reconvictions achieved). Organisations who are in turn commissioned by the CRCs (termed Tier 2 and Tier 3 organisations) receive payments from the CRCs either through a simple fee, or a combined delivery fee plus PbR element.

At the core of the TR landscape is the need for organisations to robustly demonstrate the impact that their services provide. This will be essential if they are subject to a formal PbR element, when the fee paid will be determined by the impact on reconvictions, or when services are to be (re-) commissioned in the future.

This requirement may in turn require a substantial change in routine for many organisations. As recently highlighted (Stephenson Dodd, 2016), the high profile collapse of Kids Company suggested a culture where the quality standard for impact evidence fell unacceptably low, relying on 'good stories' rather than measured, demonstrable impact.

As noted by Wong (2013) in the run up to the implementation of the TR changes, the TR landscape demands a higher level of objective evidence. In setting out evaluation-related considerations, the suggestion is not that this is the primary concern for service providers. Clearly the provision of a quality service should be the main concern, but quality evaluation evidence is obviously crucial to identifying what comprises a 'quality' service, and guiding efforts to improve services.

What follows is a checklist that might be considered by project designers and commissioners to ensure that the evidence of impact is as robust and defendable as possible.

Effective project design to assist impact evaluation

Making sure that projects are designed and delivered in the right way to enable robust evaluation is central to Dawson and Stanko's reflections captured in their aptly titled paper "Implementation, implementation, implementation: insights from offender management evaluations" (Dawson & Stanko 2013).

Impact evaluations can be characterised as an assessment of the change produced by a specific intervention. As noted above, 'good stories' do not, in themselves, make impact evaluation findings. Honest and robust evaluation requires findings to be representative of a project as a whole: individual quotes or case studies, without some indication of whether this is representative of the overall project cannot constitute robust findings. Qualitative findings can provide an important depth of understanding that purely quantitative findings lack, but on their own rarely provide much insight into the overall impact of a project.

What is being evaluated?

Too many projects being evaluated fail to clearly delineate the exact extent and dimensions of the project. The guiding principle of project design needs to delineate the project in question from other delivery: without the ability to do this, the impact of the project cannot be assessed.

Projects need to be clearly identified in terms of:

- a start and end date;

- what constitutes the 'in-scope' cohort, how it is selected, and to what extent groups with particular needs such as women and/or black and minority ethnic (BME) groups are included;
- the economic and other inputs that the project benefits from;
- the specific outcome(s) that the project is aiming to produce;
- the model of intervention being followed and how this may be tailored for particular sub-cohorts such as women and/or BME groups;
- relationships with other providers;
- criteria for the continuation/cessation of client engagement.

At the heart of all these considerations is the concept of 'additionality': what is the project doing that is *additional* to existing delivery, and what additional outcomes are anticipated. Funding for innovation can often in practice represent 'continued funding for existing services'. While it is understandable that agencies will seek to secure funding (from whatever sources are available) to extend the life of their projects, under such circumstances, the *additional* impact is very difficult to identify and demonstrate.

The counterfactual

Closely related to the idea of 'additionality' is the concept of the counterfactual, usually defined as the situation that would have existed had the project not been in place. Effective impact evaluation is greatly assisted by a clearly identified counterfactual situation to provide a baseline condition against which performance can be compared. Different estimates of the counterfactual conditions include:

- a geographical counterfactual: a similar cohort in another area, including the possibility of a wider geographical area (e.g. national);
- a historic counterfactual: the same cohort in the period preceding the project;
- randomly allocated counterfactual: in some circumstances it may be possible and desirable for 'in-scope' clients to be allocated to either an intervention group, who engage with the project, or a counterfactual group who do not engage;
- a statistically identified comparison group: if a large population of similar potential clients can be identified, a multivariate technique, Propensity Score Matching, can identify close matches from the group to act as a comparison group. This is the methodology adopted by the Justice Data Lab (MoJ, 2013b): a service developed within the Ministry of Justice. An organisation sends details of their client group to the Data Lab, which statistically identifies a matched comparison group allowing reconviction figures to be compared[33].

The principle is that the counterfactual conditions are, as far as possible, identical to the project conditions, differing only in receiving an intervention from the project being evaluated. In such circumstances, differences in measured outcomes can be robustly ascribed to the impact of the project. The reality of real-life research, though, usually

---

[33] Certain criteria need to be met before the match is carried out: for example, a sufficient cohort size needs to be provided, and offenders who have committed certain offence categories will not be matched. Some concerns also exist about what interventions the matched comparison group have received.

means that none of the counterfactuals are ideal. A randomly allocated counterfactual, for example, offers the greatest likelihood of an otherwise identical group, but random allocation is often considered unethical or impractical. Historical and geographical counterfactuals are limited by the possibility that there are systematic differences in the circumstances underpinning the project.

Counterfactual conditions are often ignored by projects, particularly working with people whose lives are in flux, and particularly young people. It is a well-established empirical finding (McVie, 2005) that the prevalence of offending peaks in the late teenage years, and thereafter declines substantially. Although the subject of much debate, this pattern is not strongly related to any particular intervention, but is more to do with general maturation, and other signifiers of maturation (for example, the establishment of a significant relationship, marriage, or getting and maintaining employment). Projects, therefore, that are aiming to reduce offending with a cohort of young adults need to demonstrate an additional effect over and above this underlying pattern.

A consideration of the counterfactual condition is therefore vital, and therefore needs to address:
- what is the most appropriate counterfactual condition to consider?
- what is the availability and quality of the data for the counterfactual?
- how different is the counterfactual condition from the project condition?

Research evaluation teams may be able to advise here, but project staff with the experience of the organisation and client group are likely to be in a better position to answer these questions.

Data considerations
Impact evaluations will require a range of good quality data to evidence change in the project outcomes. It is essential that data is available, as a minimum:

- Initial assessments, made at the point of initial engagement;
- Subsequent assessments, made at the point of disengagement, or the end of the project;
- For both the project and counterfactual conditions.

The data considerations for the counterfactual condition can provide a potential stumbling block: geographical counterfactuals will need to be identified and matched carefully, and if a historical counterfactual is selected, it may be difficult to access data retrospectively. The most suitable counterfactual condition will be different for each project, and strategic and delivery staff are often the best people to identify these.

Many outcomes may be easily measured (such as binary measures such as 'Got a job/Didn't get a job') but careful consideration should be made of the precise form the data is collected in. These considerations should include:

- timescales: is it sufficient for project purposes for a client to start a job or training opportunity, or does this need to be sustained for a particular length of time?
- binary or other level of measurement: is it necessary for a client to cease offending, or would a certain reduction in the volume of offending be sufficient?

Other outcome variables may be more complex to measure: where possible the use of existing, validated measures for such a concept strengthens evaluation findings, allowing them to be compared to other studies. A good example is the Warwick- Edinburgh Mental Wellbeing Scale (http://www2.warwick.ac.uk/fac/med/research/platform/wemwbs/).

One consideration which is often overlooked is the timing of the assessment of outcome measures, whether at the start or the end of the period of engagement. In the idealised world, 'before' assessments clearly need to be made as near as possible to the start of the period of engagement. Often this is not possible: assessments are often time-consuming and thorough, requiring a number of hours of discussion between clients and key-workers, and are often also reliant on good rapport having been established. In many cases, this is simply not possible at the earliest stage of engagement. Similarly, at the end of a period of engagement, clients may be less than keen to complete a final assessment, no matter how important this is for the evaluation. These problems are generally even more acute in counterfactual areas, where the propensity to complete assessments is likely to be lower still. Project design that recognises these issues could consider the most effective points that assessments should be carried out, and what incentives might be put in place.

Throughput, drop-out and bias
Effective project design needs to ensure that sufficient numbers of clients are engaged by the project, and are included in the counterfactual conditions. If too few clients are engaged, then comparisons between the outcomes achieved in the two conditions will not achieve statistical significance. The critical cohort sizes will come from the number of clients who have completed the project and have a final outcome level assessed. Design and resourcing of a project therefore needs to consider the potential number of in-scope clients who may potentially access the service, the resources required to deliver the service to them, and the likely drop-out rate.

Drop-out needs further consideration: the very strong probability is that those that drop-out are not representative of the evaluation cohorts as a whole. As such, comparison of outcomes between the project and counterfactual conditions may be significantly biased. For example, if an evaluation was focussing on re-offending as the key outcome for a project, those who disengaged from a project may be considerably more likely to re-offend, but because the comparison of outcomes can only be done on those clients who remain engaged throughout the project, this finding would not be accurately identified. In addition, if the patterns of disengagement are different between the test and counterfactual conditions, perhaps as a result of a particularly challenging model of delivery, then again the evaluation is undermined. In terms of project delivery, the challenge often comes down to the maintenance of the cohort numbers: whilst financial incentivisation may be considered to boost numbers, this can often be contentious with some client groups, raising the possibility that incentives could affect the actual delivery of the project, and clearly adds to the financial costs incurred.

Inclusion bias may also be a problem for some projects. If recruitment to the evaluation cohort requires a client to demonstrate their willingness to participate, then this can bias the evaluation: in effect, this willingness may be the critical factor in project engagement ending in a positive outcome. It does not, though, reflect an impact genuinely produced by the project, and if a similar process is not present in the counterfactual condition, then a bias is introduced in to the evaluation.

Issues related to throughput, drop-out and bias are generally not possible to remove from impact evaluations, but need to be covered in evaluation reports. Evaluations need to report on case numbers, and not just percentages (as recognised by Stephenson Dodd 2016). Drop-out numbers and rates for both intervention and counterfactual cohorts also need to be included in reports, and impact evaluations are strengthened considerably if disengaged clients can be contacted and the reasons for disengagement explored. This is, however, often extremely difficult and time-consuming to achieve.

Do we know why outcomes are produced?

The central message of the leading evaluation text 'Realistic Evaluation' (Pawson & Tilley, 1997) is that effective evaluation requires more than a simple 'What Works?' question, but rather asks " What works for whom in what circumstances and in what respects, and how?". Impact evaluations that rely on a black box approach (i.e. Intervention 'X' happened at this point, and this is the effect, but we don't know, or try to find out, what Intervention 'X' is) are of limited practical use if we cannot discern what contexts programmes work in, which elements of the project have the greatest impact, and what causal links can be identified between project activity and the outcomes produced.

Two broad approaches can be taken to explore these further questions: an experimental approach would compare sub-cohorts that may have received specific elements of a project, and compare the impact relative to each other, and the counterfactual conditions. This approach offers the potential for a more accurate measurement of impacts, but does require a high degree of control over the administration of different project elements to specific clients, and ultimately can only move so far towards a fuller causal explanation. A qualitative approach, involving interviewing clients, delivery staff and other stakeholders offers a complimentary approach, though this may be limited by the interviewees' perceptions.

With a view to project design, the more complex the new intervention is, the more difficult it is to robustly ascribe cause. A neatly designed, narrowly focussed project is more likely to produce robust findings than a multi-faceted project addressing a range of contexts and cohorts.

Conclusion

The internal logic of Transforming Rehabilitation is that market mechanisms will produce innovation, and in time improved rehabilitation efforts. Whether this is the case is open to debate, but the rules of the game appear to have been established. It is now even more vital that voluntary organisations know the impact their services are producing, whether to fulfil the formal requirements of PbR contracts, to demonstrate their effectiveness

when negotiating for new contracts, or simply to review their own performance. With this in mind, this article has highlighted areas that should be considered when designing projects, in order to assist the evaluation process. The primary purpose of project design should be to deliver effective interventions to clients, and the needs of evaluators should never be prioritised ahead of clients' needs, but these two elements are far from contradictory: effective evaluation is central to the design and evolution of high quality service delivery.

References

Dawson, P. and Stanko, B. (2013) Implementation, Implementation, Implementation: Insights from Offender Management Evaluations, *Policing*, 7(3): 289-298.

McVie, S. (2005) Patterns of deviance underlying the age-crime curve: the long term evidence. *British Society of Criminology*, 7: 1-15.

Ministry of Justice (2013a) *Transforming Rehabilitation: A Strategy for Reform.*

Ministry of Justice (2013b) *Justice Data Lab: Methodology Paper.*

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

Stephenson-Dodd, M. (2016) 'Are charities doing as much good as they say or are they telling tales?' *The Guardian*, 7 November 2016.

Wong, K. (2013) Integrated Offender Management: Assessing the impact and benefits - Holy Grail or Fool's errand? *British Journal of Community Justice*, 11(2-3): 59-81.

bjcj

COMMUNITY JUSTICE FILES 40

Compiled by Ross Little, De Montfort University

Quality and Impact inspection - The effectiveness of probation work in the north of London (HMIP)

A Quality and Impact inspection report into the effectiveness of probation work in the north of London has been published by HM Inspectorate of Probation in mid-December. This is their first inspection of adult probation services in the capital since 2014, when services were managed by one organisation, the London Probation Trust. On this occasion, they inspected in eight boroughs in the north of London. The London CRC manages 11% of the total 258,748 people subject to probation supervision across England and Wales.

The foreword by Dame Glenys Stacey, the HM Chief Inspector of Probation, notes that "Probation services in London have long struggled with high workloads, and workload pressures have been a regular feature in the most notorious of cases where a supervised individual has committed a Serious Further Offence." Dame Stacey wastes no time in stating that they found the work of the Community Rehabilitation Company, owned by MTCnovo, to be poor.

Some CRC staff were found to be working "heroically" in difficult circumstances, sometimes working long hours. However they were often 'fire-fighting' rather than enabled to deliver a professional service consistently or sufficiently well. The National Probation Service (NPS) was found to be delivering services better, but with plenty of room for improvement.

With respect to overall recommendations, the report considers how well probation partners were in protecting the public, reducing re-offending and in relation to how well service users abided by their sentence. The work of the CRC was found to be poor in all respects, with fewer than half the service users in their sample complying with their sentence. In contrast, in this respect the performance of the NPS was found to be good. The NPS performance in the other two respects (protecting the public and reducing re-offending) was found to be mixed. The following recommendations for probation partners are made, the following being taken directly from the report (p.11):

The Community Rehabilitation Company and National Probation Service should:
1. Produce easily accessible information to enable all staff to make swift contact with relevant colleagues in each.