

When the best laid plans go astray: a case study in pragmatic approaches to evaluation

Hayley Boxall

Australian Institute of Criminology

Hayley.Boxall@aic.gov.au

Keywords: Family group conferencing evaluation, pragmatic evaluation, reflexive evaluation, adaptive evaluation, administrative data use in evaluation

Abstract

Many researchers recognise the importance of conducting evaluations in accordance with clearly articulated plans and frameworks which are developed at the outset of the project and underpin all of the subsequently undertaken research activities. However, for reasons often beyond the control of the evaluators, research methods and data collection instruments identified in these plans and frameworks may become unfeasible or inappropriate at different stages of the evaluation period. Mitigating the potential impact of these events and circumstances can be incredibly challenging for even the most seasoned researcher, who at the end of the day is still required to answer key evaluation questions, one of which may be 'Did it work?' Using the evaluation of the Family Group Conferencing pilot project (NSW) for illustrative purposes, this paper highlights the benefits of reflexive, adaptive and pragmatic approaches to evaluation and of involving project stakeholders in the development of evaluation designs and research methods.

Introduction

It is well established that evaluation research should be undertaken in accordance with clearly articulated and agreed upon plans and frameworks (Bamberger, Rugh & Mabry, 2012; McDavid, Huse & Hawthorn, 2013; Mertens & Wilson, 2012; Morgan & Homel, 2013; White & Phillips, 2012). One of the benefits of developing an evaluation plan and framework during the early stages of the project period is that it encourages the researcher to develop a series of clear research questions and to identify which data are necessary to answer these questions. The process for developing evaluation plans and frameworks is ideally undertaken in consultation with relevant stakeholders and any established evaluation committees and working groups. This consultative process helps researchers to identify issues and challenges that may have implications for the

evaluation and to develop an understanding of the social and political context within which the project is being undertaken Compton & Baizerman, 2011; Mertens & Wilson, 2012).

However, for many researchers, particularly those who are brought onto projects on a consultancy/contractual basis, evaluation plans and frameworks may be developed during the very early stages of the research period. As a consequence, these evaluation plans and frameworks may be informed by partial information and often erroneous assumptions about how the program works 'on the ground', the quality and availability of evaluation data being collected by the contracting agency and program partners and the capacity and willingness of program staff to be involved in data collection processes. Evaluators also frequently encounter barriers and challenges throughout the research period that have implications for their ability to undertake the evaluation within the specified timeframes. This includes issues associated with evaluation data collection processes and the quality of the data itself (for example, the absence of baseline data, lack of appropriate and natural comparison group, small and non-representative samples and missing data; Bamberger, Rugh, Church & Fort, 2004; Eck, 2006).

Because of these factors, evaluators may be required to amend their plans and frameworks at various stages throughout the life of the project. In such situations, it is important that researchers adopt a pragmatic approach to evaluation – that they are able to adapt and amend their evaluation approach and research design and methodology so they are still able to collect the data necessary to answer key evaluation questions. A useful framework through which this flexible approach to evaluation may be understood and grounded is the aptly named 'Pragmatism' (for an overview see Morgan, 2007). Pragmatism is an attempt to reconcile what proponents view as the false dichotomy between positivist and interpretivist/constructionist research paradigms and their related emphasis on 'pure' qualitative or quantitative research methods (Feilzer, 2010; Johnson & Onwuegbuzie, 2004; Onwuegbuzie & Leech, 2005). Instead of single-method 'pure' research methodologies, pragmatists extol the benefits associated with mixed-methods research designs. For example, Onwuegbuzie and Leech (2005) suggest that mixed-methods research facilitates collaboration between researchers and the consideration of both micro and macro factors and analyses. Further, they suggest that the 'inclusion of quantitative data can help compensate for the fact that qualitative data typically cannot be generalized. Similarly, the inclusion of qualitative data can help explain relationships discovered by quantitative data' (Onwuegbuzie & Leech, 2005, p. 383).

Importantly, under the pragmatism framework the selection of research methods is 'needs-based' and underpinned by a 'contingency approach' (Johnson & Onwuegbuzie, 2004, p. 18). As Feilzer (2010, p. 14) notes, 'pragmatists do not "care" which methods they use as long as the methods chosen have the potential of answering what it is one wants to know'. Consequently, pragmatic frameworks give evaluators the freedom to select research methods that are feasible and will provide them with the data they need to answer the research questions, rather than to satisfy an epistemological need. This in turn facilitates flexible approaches to evaluation and the evolution of research designs and methods throughout evaluation periods.

Aim of this paper

This paper aims to provide evaluators, particularly early-career researchers, with practical advice on how to undertake pragmatic evaluations in situations where there are concerns about the collection and quality of data. More specifically, using a real world case study for illustrative purposes, this paper:

- provides an overview of some of the challenges that researchers frequently encounter when undertaking an evaluation in such situations;
- describes a range of strategies identified by researchers and practitioners for minimising the impact of and addressing these concerns;
- describes how these strategies have been applied in practice; and
- outlines the subsequent impact of these strategies.

Case study—the evaluation of the Family Group Conferencing pilot project

The Family Group Conferencing (FGC) pilot project was a small-scale project implemented in New South Wales in response to recommendations made as part of the *Special Commission of Inquiry into Child Protection Services in NSW* (Wood, 2008). The project commenced operation in March 2011 and was piloted in 11 community services centres (CSCs) located across metropolitan and regional areas of NSW. The overarching aim of the FGC pilot project was to empower families to develop, implement and manage Family Plans to address the care and protection issues raised by the statutory child protection agency operating in NSW – the Department of Families and Community Services (FACS). Conferences held as part of the project were attended by parents, the children and young people (where appropriate), extended family members, service providers, Community Services Caseworkers and Managers Casework; and chaired by trained Facilitators who were independent of Community Services. Conferences were conducted in neutral community-based venues and were focused on developing strategies that could be implemented by the family to address the identified care and protection issues.

In June 2011, FACS commissioned the Australian Institute of Criminology (AIC) to undertake a process and outcome evaluation of the FGC pilot project. To assess the implementation and short-term impact of the project, the allocated research team developed a program logic model and evaluation framework that aligned with the implementation plan for the evaluation of Keep Them Safe – the NSW Government’s response to the Wood Commission (Urbis, 2011). This logic model and framework formed the basis of the AIC’s evaluation and informed the development of a comprehensive quasi-experimental methodology combining quantitative and qualitative research methods.

In addition to the evaluation plan for Keep Them Safe, the FGC pilot project evaluation framework was informed by two other key pieces of information – the original tender documents provided by FACS, and the AIC’s evaluation of another project that was implemented as part of Keep Them Safe (see Morgan, Boxall, Terer & Harris, 2012).

Unfortunately, due to time constraints the research team were not able to consult with the relevant project stakeholders or Evaluation Working Group (EWG) established by FACS, prior to drafting the evaluation plan. The EWG was comprised of individuals who were involved in the management and delivery of the FGC project in the ten sites, including the Project Coordinator.

As demonstrated in Table 1, over the course of the research period the original evaluation plan and framework was amended and adapted in response to a series of challenges (and opportunities) that were identified by the research team. Some of these challenges and opportunities – specifically, small sample sizes, unfeasible data collection methods and concerns regarding administrative data – will be discussed throughout the rest of this paper.

Table 1: Overview of the evaluation methodology for the FGC pilot project at the beginning and end of the evaluation period

<i>Original evaluation methodology</i>	<i>Final evaluation methodology</i>
Literature review to identify good practice	Literature review to identify good practice
Observation of a small number of conferences	Observation of a small number of conferences
Semi-structured interviews with parents and family members who participated in conferences held as part of the project.	Semi-structured interviews with parents and family members who participated in conferences held as part of the project.
Interviews, focus groups and a qualitative survey to seek feedback from stakeholders involved in the project (including FACS Case Managers).	Interviews, focus groups and a qualitative survey to seek feedback from stakeholders involved in the project (including FACS Case Managers).
Analysis of administrative data collected by FACS in relation to families who: <ul style="list-style-type: none"> participated in a conference held as part of the project; and were eligible for participation in the conference but were not asked. 	Analysis of administrative data collected by FACS in relation to families who: <ul style="list-style-type: none"> participated in a conference held as part of the project; consented to participate in the project but did not participate in the conference; and were eligible for participation in the conference but were not asked.
Survey of parents and family members: <ul style="list-style-type: none"> at the end of conferences held as part of the project; and at the end of a case meeting held between the family and the FACS Case Manager. 	Case studies

Challenge 1: Small sample sizes

Many projects and programs struggle to identify and maintain a consistent and sufficient flow of referrals and participating clients throughout their lifespan (for FGC specific examples see Berzin, Cohen, Thomas, & Dawson, 2008; Brady & Miller, 2009; O'Brien, 2002; Shore, Wirth, Cahn, Yancey & Gunderson, 2002). This is particularly the case for projects:

- that are new – referring agencies may be unaware of the project and the referral processes involved, or sceptical of its merits and benefits;
- that are non-mandated processes – in other words, participation is voluntary; or
- constitute a significant departure from pre-existing processes – projects such as these may require referring agencies to change the way they approach certain issues and matters and there may be some resistance to this, particularly among long-term and senior staff (Boxall, Morgan & Terer, 2012).

Lower than anticipated project referrals numbers – and subsequently small sample sizes – can have significant implications for evaluation. In particular, small sample sizes have implications for the types of analyses that may be undertaken and the subsequent rigour and external validity of the findings¹ (Bamberger, Rugh, Church & Fort, 2004). Even so, 'small-n' evaluations are frequently undertaken and have benefit insofar as they provide useful information about the project or program being examined that may not be available otherwise (Eck, 2006; White & Phillips, 2012). Further, a number of existing research methods (e.g. success case methods, general elimination methodologies) and theoretical frameworks (e.g. realism) are compatible with small-n evaluations (Brinkerhoff, 2005; White & Phillips, 2012). The point of commonality among these methods and frameworks is that they aim to identify and understand the mechanisms and processes underlying projects and programs that bring about observable change.

Other strategies for minimising the impact of small sample sizes include the use of case studies and/or multiple sources of data to examine issues from various viewpoints and angles. This strategy is often referred to as triangulation and is commonly used in mixed-methods and qualitative studies (Bamberger, Rugh & Mabry, 2012; Mertens & Wilson, 2012; White & Phillips, 2012).

The evaluation of the FGC pilot project and small sample sizes

The original tender documentation provided to the research team indicated that 60 families would be invited to participate in the FGC pilot project and proceed to a conference. Consequently it was always anticipated that the evaluation would involve the analysis of data collected and extracted in relation to only a small number of families. However, during the early stages of the evaluation it became evident that referrals to the project were lower than expected². To maximise the number of families that could participate in the FGC pilot (and therefore the potential sample size), FACS decided to extend the project period for an additional few months. As a consequence, by the end

of the evaluation period 29 families had been referred to the project and participated in a conference. Another 31 families were referred to the project but had not proceeded to a conference for a range of reasons described in the final evaluation report (Boxall et al. 2012).

To partially mitigate the impact of the small sample size on the evaluation, the research team developed a series of case studies that were selected using success case sampling methods. This resulted in the identification of families who did and did not appear to benefit from the project, and an analysis of the factors that may have led to their success or lack thereof (for example, the number and identity of family members who participated in the conference). These case studies were developed using a range of information collected through the evaluation, including the analysis of administrative data, and were included in the final report to highlight key findings (Boxall, Morgan & Terer, 2012).

Further, because we had collected a range of information using different data collection methods (see Table 1), we were able to problematize and explore evaluation findings from various angles and viewpoints. This increased the validity and reliability of our findings. Finally, the research team provided members of the EWG and other relevant project stakeholders with opportunities to provide their interpretation of the data and preliminary findings at different stages during the evaluation. These ‘member checks’ again increased the reliability of the findings (Mertens & Wilson, 2012).

Challenge 2: Unfeasible/Inappropriate data collection methods

As mentioned earlier in this paper, during the early stages of an evaluation, a researcher’s understanding of the project may be limited and in some cases based on a series of erroneous assumptions. In situations such as this, research methods and data collection tools that are initially identified by evaluators may be unfeasible or even inappropriate. For this reason, it is beneficial for researchers to involve project staff and established evaluation working groups in the initial planning stages of the evaluation (Mertens & Wilson, 2012; O’Sullivan, 2012). There are numerous frameworks that explicitly encourage this process, including collaborative, empowerment, participatory and utilization evaluation. One of the functions of this engagement process is to ‘reality-test’ proposed research methods and advise the researchers on appropriate research approaches (American Evaluation Association, 2004). This is particularly important if the collection of data are reliant on the participation and support of individuals involved in the delivery and management of the project itself.

1. It is important to note that there are some qualitative research methods that involve the in-depth examination of one or a small number of observations. This includes case studies, which were developed as part of the evaluation of the FGC pilot project.

2. The process evaluation identified a number of factors that contributed to this phenomenon, including a lack of awareness of the project among FACS staff (who were responsible for referring matters in the first instance), the perceived cumbersome and time-consuming referral processes themselves and that participation in the project was voluntary (Boxall, Morgan & Terer, 2012).

Unfeasible/Inappropriate data collection methods and the FGC pilot project

Due to time constraints, the research team were unable to consult with the EWG prior to drafting the initial evaluation framework. However, shortly after submitting the evaluation framework we had an opportunity to meet with the EWG who were tasked with providing feedback on the document. Throughout this consultation process, it became apparent that there were potential issues administering surveys to; (1) parents and family members at the end of conferences held as part of the project (the intervention group); and (2) at the end of case planning meetings held with families not involved in the program (the comparison group) Because conferences held as part of the FGC pilot project could and occasionally did run for the entire day and were emotionally draining for all involved, the EWG suggested that asking parents and family members to complete a survey at the end of this process was inappropriate. Also, in order to attend conferences, family members often had to take time off work or make alternative care arrangements for their children. Consequently, family members frequently had to leave during or immediately after conferences and so did not have time to complete a survey. Similar issues were raised in relation to the administration of surveys to parents and family members at the end of case planning meetings held with FACS Case Managers³.

In light of the issues raised by members of the EWG, the research team decided not to administer surveys to family members and parents after conferences or case planning meetings. This decision was informed by the consideration of both the potential impact of removing these research methods from the evaluation plan and of not removing them. Although their non-inclusion limited our ability to explore evaluation findings from various angles (triangulation), we determined that it was feasible to answer the identified research questions without the survey data. Further, because of issues identified by the EWG, the usefulness and reliability of this data was potentially limited anyway (i.e., non-representative sample of family members). However, more importantly, if the research team forged ahead with their plans of administering the surveys we ran the risk of jeopardising our working relationship with the EWG who may have perceived that their opinions had not been acknowledged and valued and disengage from the evaluation as a result. This was of particular concern considering that many of the research methods identified in Table 1 were dependent on the support and involvement of FGC staff and FACS Case Managers. In light of these considerations, the research team were comfortable that the potential benefits associated with removing the surveys from the evaluation methodology outweighed the drawbacks.

Challenge 3: Concerns regarding administrative data

Many program evaluations involve the analysis of administrative data collected, stored and managed by staff and agencies involved in the delivery and implementation of the project. This data may include information relating to the processes involved in the project (for example, project participation numbers), outputs (e.g. treatment and intervention plans) and also outcomes. While administrative data can be an important

and integral element of any evaluation, there may be issues associated with this information. For example:

- the data may not be in a format suitable for analysis – information may only be available in hardcopy;
- there may be ‘gaps’ in the data – data collection protocols may have changed during the project period meaning the same information was not collected for all participants; and
- there may be some concerns regarding the quality of the data – due to inconsistent data collection procedures, multiple people inputting data into databases and spreadsheets etc (Bamberger, Rugh, Church & Fort, 2004).

All of the above identified issues have implications for the use of administrative data in evaluation and need to be managed carefully. Consequently, it is advisable that researchers request to see a sample extract from the relevant databases and speak to the contracting agency’s data custodians and managers during the early stages of the evaluation. Alternatively, researchers may seek to conduct an ‘audit’ of existing databases and data collection processes early in the evaluation period. This process may be facilitated through the use of a quality assessment ‘checklist’ such as that developed by Statistics Sweden (Daas, Ossen & Arends-Toh, 2009). However, undertaking such auditing processes may be unfeasible or inappropriate, depending on the contracting agency, its data collection processes and the political context.

Concerns regarding administrative data and the FGC pilot project

The evaluation of the FGC pilot project involved the analysis of a range of administrative data collected by staff directly involved in the management and delivery of the FGC project and FACS. In particular, as demonstrated in Table 1, from the outset we planned to extract and analyse data stored in the Key information and Directory System (KiDS) for families included in the intervention and comparison groups to determine the impact of the project against a number of short-term and intermediate outcomes (e.g. re-contact with FACS).

During the early stages of the evaluation period, we were notified by the EWG and FACS data custodians that there were some issues with the KiDS data that may impact the evaluation. Specifically:

- KiDS had only been operating since the early 2000s and so did not include information in relation to families, parents and children prior to this point in time;

3. It is important to note that all research undertaken by the AIC involving human subjects requires clearance from the Institute’s Human Research Ethics Committee (HREC) which is comprised of experienced researchers in various fields who are not AIC employees. The evaluation of the FGC project involved the development of a comprehensive and detailed ethics application that acknowledged and addressed the perceived ethical issues associated with involving families not involved in the project in the evaluation.

- there was a three month time delay associated with data being entered into KiDS for children being managed by the OOHC teams; and
- information about Family and Children’s Courts legal orders made in relation to families managed by FACS was not included in KiDS – rather, this information was managed by FACS legal department.

All of these factors had potential implications for the evaluation. First the absence of historical information about families meant that contextual understandings of families involved in the project (and the factors that may have influenced their success or lack thereof) were limited. Second, the time lag associated with data being uploaded for Out-of-Home-Care (OOHC) matters meant that our ‘follow-up’ period for these families could have been three months shorter than for non-OOHC families. Finally, the lack of information about legal orders meant that our ability to identify the impact of the project on legal outcomes (e.g. restoration orders) was potentially limited.

Because the research team had consulted with the EWG and FACS data custodians during the early stages of the evaluation, we were able to anticipate and where possible, implement strategies to manage these issues. While there was nothing we could do about the loss of historical data, we discovered that this information was, to a limited extent, being included in the referral reports completed by families and FACS Case Workers during the early stages of their engagement on the project. Consequently, the research team extracted this information from the referral reports and included it in the analysis.

Similarly, there was not much that could be done to minimise the impact of the three month delay in data being uploaded into the KiDS for OOHC matters. This said, because we identified this as an issue from the outset, we were able to stagger the data extraction dates. This meant that data was extracted for all non-OOHC matters at the end of 2013 while the same data was extracted for OOHC matters three months later. This ensured that the data extraction period for all matters was approximately the period of time. Finally, although the legal data we needed was not available through KiDS, we consulted with FACS’ legal department and were able to obtain the information we needed. This data extract was subsequently ‘linked’ with our other data extracts using the unique case plan and child identifiers generated by FACS.

Implications for evaluation quality and rigour

Any changes that researchers make to their evaluation frameworks and plans have potential implications for the collection and analysis of data and, in turn, their ability to answer identified evaluation questions. The FGC pilot program was no exception. For example, extending the project delivery period had both positive and negative ramifications for the evaluation. On the one hand, extending the project period resulted in a larger number of families being provided with the opportunity to participate in the project, which in turn meant there was a larger population to draw from for the purpose of the evaluation. However, by extending the project period, the length of time that families could be ‘followed’ post-intervention was in turn limited. Consequently, the research team were only able to determine the short-term impact of the FGC pilot project.

Further, the removal of the post conference/case planning survey meant that our ability to assess the impact of the program against short-term outcomes (e.g. increased satisfaction of parents and family members with their role in decision-making processes) was limited; as was our ability to compare the outcomes of the FGC pilot against a 'standard practice' approach. Because of these identified limitations, the research team recommended that 'A future evaluation should be conducted to measure the longer term impact of FGC on care matters once the program has been fully established and data on a larger number of participants is available. Processes for monitoring outcomes from FGC therefore need to be established and/or maintained' (Boxall, Morgan & Terer, 2012, p. xvi).

However, it is important to note that some of the changes that were made to the evaluation framework and plan actually *increased* the rigour of the research and the subsequent validity of the findings. For example, as previously discussed the research team removed the post-conference/case planning survey from the evaluation methodology in part because the concerns raised by the EWG suggested that the data collected through this process may be of limited use and reliability (e.g. low response rate, meaning the external validity of the data would be low). By removing a potentially flawed and problematic research method, the research team improved the quality and reliability of the evaluation design and the subsequent findings, and also freed up additional resources for the collection and analysis of other data.

Conclusion

The FGC pilot case study described throughout this paper demonstrates that initial plans and frameworks developed by evaluators will rarely be 'perfect'. Consequently, while it is beneficial and advisable to conduct evaluation research in accordance with clearly articulated plans and frameworks, it is also necessary for evaluators to be flexible and adaptive so they can respond appropriately to challenges they encounter. This may involve removing specific research methods, including others, and amending data collection tools or processes. Consequently, these evaluation plans should be viewed as flexible working documents that will be changed, tweaked and amended as new challenges and issues emerge and the research team's understanding of the program improves. This said, decisions to amend these plans and frameworks should not be made lightly and should take a range of factors into consideration – one of the most important being, are we still going to be able to answer our identified evaluation questions if we make these proposed amendments/changes?

The case study also highlights the importance of engaging with relevant project stakeholders throughout the evaluation period, and in particular, seeking their input into the evaluation design and specific research methods. Project stakeholders may be asked to undertake a range of tasks throughout the evaluation process, such as collecting and brokering access to data and reality testing proposed research methods. Ensuring that project stakeholder's views are acknowledged and, where appropriate, integrated into the evaluation plan goes some way to ensuring that strong working relationships are maintained and potential issues identified in relation to data collection processes; early identification of data and strategies to minimise any identified issues implemented.

References

- American Evaluation Association. (2004). *American Evaluation Association guiding principles for evaluators*. Retrieved from <http://www.eval.org/p/cm/ld/fid=51>
- Bamberger, M., Rugh, J., Church, M., & Fort, L. (2004). Shoestring evaluation: Designing impact evaluations under budget, time and data constraints. *American Journal of Evaluation*, 25(1), 5–37.
- Bamberger, M., Rugh, J., & Mabry, L. (2012). *Real World evaluation: Working under budget, time, data, and political constraints*. London: Sage Publications.
- Berzin, S.C., Cohen, E., Thomas, K., & Dawson, W.C. (2008). Does family group decision making affect child welfare outcomes? Findings from a randomized control study. *Child Welfare* 87(4), pp. 35–54.
- Boxall, H., Morgan, A., & Terer, K. (2012). *Evaluation of the Family Group Conferencing pilot program*. Research & Public Policy Series no. 121, Canberra: AIC.
- Brady, B., & Miller, M. (2009). *Barnardos Family Welfare Conference Project, South Tipperary: Evaluation Report*. Retrieved from http://childandfamilyresearch.ie/sites/www.childandfamilyresearch.ie/files/barnardos_family_welfare_conference_service_south_tipperary_evaluation_report.pdf
- Brinkerhoff, R.O. (2005). The success case method: A strategic evaluation approach to increasing the value and effect of training. *Advances in Developing Human Resources*, 7(1), pp. 86–101.
- Compton, D.W. & Baizerman, M. (2011). Managing evaluation: responding to common problems with a 10-step process. *Canadian Journal of Program Evaluation*, 25(2), pp. 103–123.
- Daas, P.J., Ossen, S.J. & Arends-Tóth, J. (2009). *Framework of Quality Assurance for Administrative Data Sources*. Statistics Netherlands: The Hague, retrieved from http://pietdaas.nl/beta/pubs/pubs/ISI2009_paper.pdf
- Eck, J.E., 2006. When is a bologna sandwich better than sex? A defense of small-n case study evaluations. *Journal of Experimental Criminology*, 2(3), pp. 345–362.
- Feilzer, M.Y. (2010). Doing mixed methods research pragmatically: Implications for the rediscovery of pragmatism as a research paradigm. *Journal of Mixed Methods Research*, 4(1), pp. 6–16.
- Johnson, R.B. & Onwuegbuzie, A.J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), pp. 14–26.
- McDavid, J., Huse, I. & Hawthorn, L. (2013). Key concepts and issues in program evaluation and performance measurement. In McDavid, J., Huse, I. & Hawthorn, L. (Eds), *Program Evaluation and Performance Measurement: An Introduction to Practice*. London: Sage Publications.

- Mertens, D.M. & Wilson, A.T. (2012). *Program evaluation theory and practice: A comprehensive guide*. New York: The Guilford Press.
- Morgan, A., Boxall, H., Terer, K. & Harris, N. (2012). *Evaluation of alternative dispute resolution initiatives in the care and protection jurisdiction of the NSW Children's Court*. Research & Public Policy Series no. 118, Canberra: AIC.
- Morgan, A. & Homel, P. (2013). *Evaluating crime prevention: lessons from large scale community crime prevention programs*. Trends & Issues in Crime & Criminal Justice no. 458.
- Morgan, D.L. (2007). Paradigms lost and pragmatism regained methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research*, 1(1), pp. 48–76.
- O'Brien, V. (2002). *Family group conference pilot project: evaluation report*. Mid-Western Health Board. Retrieved from <http://irserver.ucd.ie/bitstream/handle/10197/3085/MID%20Western%20Health%20Board%20FGC%20Report%202002.pdf?sequence=1>
- O'Sullivan, R.G. (2012). Collaborative evaluation within a framework of stakeholder-oriented evaluation approaches. *Evaluation and Program Planning*, 35(4), pp. 518–522.
- Onwuegbuzie, A.J. & Leech, N.L. (2005). On becoming a pragmatic researcher: The importance of combining quantitative and qualitative research methodologies. *International Journal of Social Research Methodology*, 8(5), pp. 375–387.
- Shore, N., Wirth, J., Cahn, K., Yancey, B. & Gunderson, K. (2002). *Long term and immediate outcomes of family group conferencing in Washington State*. Washington State: International Institute for Restorative Practices. Retrieved from http://www.iirp.edu/iirpWebsites/web/uploads/article_pdfs/fgcwash.pdf
- Urbis. (2011). *Implementation plan for evaluation of Keep Them Safe*. Sydney: Urbis Pty Ltd. Retrieved from http://www.dpc.nsw.gov.au/_data/assets/pdf_file/0019/125146/Urbis_Final_KTS_Implementation_Plan_-_publication_version.pdf
- White, H. & Phillips, D. (2012). *Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework*. New Delhi: International Initiative for Impact Evaluation.
- Wood, J. (2008). *Report of the special commission of inquiry into child protection services in NSW: Volume 2*. Sydney: NSW Government. Retrieved from <http://www.dpc.nsw.gov.au/publications/news/stories/?a=33797>