



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Nicholas, A;Ovenden, G;Vlais, R

Title:

A Practical Evaluation Guide for Behaviour Change Programs Involving Perpetrators of Domestic and Family Violence

Date:

2020

Citation:

Nicholas, A., Ovenden, G. & Vlais, R. (2020). A Practical Evaluation Guide for Behaviour Change Programs Involving Perpetrators of Domestic and Family Violence. ANROWS.

Persistent Link:

<https://hdl.handle.net/11343/242023>



Developing a practical evaluation guide for behaviour change programs involving perpetrators of domestic and family violence

ANGELA NICHOLAS
GEORGIA OVENDEN
RODNEY VLAIS

ANROWS

AUSTRALIA'S NATIONAL RESEARCH
ORGANISATION FOR WOMEN'S SAFETY
to Reduce Violence against Women & their Children

RESEARCH REPORT
ISSUE 17 | JUNE 2020

ANROWS acknowledgement

This material was produced with funding from the Australian Government Department of Social Services. Australia's National Research Organisation for Women's Safety (ANROWS) gratefully acknowledges the financial and other support it has received from the government, without which this work would not have been possible. The findings and views reported in this paper are those of the authors and cannot be attributed to the Australian Government Department of Social Services.

Acknowledgement of Country

ANROWS acknowledges the Traditional Owners of the land across Australia on which we work and live. We pay our respects to Aboriginal and Torres Strait Islander Elders past, present, and future, and we value Aboriginal and Torres Strait Islander histories, cultures, and knowledge. We are committed to standing and working with Aboriginal and Torres Strait Islander peoples, honouring the truths set out in the [Warawarni-gu Guma Statement](#).

© ANROWS 2020

Published by

Australia's National Research Organisation for Women's Safety Limited (ANROWS)
PO Box Q389, Queen Victoria Building, NSW 1230 | www.anrows.org.au | Phone +61 2 8374 4000
ABN 67 162 349 171

Developing a practical evaluation guide for behaviour change programs involving perpetrators of domestic and family violence (Research report) / Nicholas et al.

Sydney : ANROWS, 2020

Pages ; 30 cm. (Research report, Issue 17/2020)

I. Family violence -- Australia -- Prevention. II. Abusive men -- Behavior modification -- Australia. IV. Abusive men -- Services for -- Evaluation.

I. Nicholas, Angela. II. Ovenden, Georgia. III. Vlasis, Rodney.

ISBN: 978-1-925925-49-4 (print) | 978-1-925925-50-0 (online)

Creative Commons Licence

Attribution-Non Commercial



CC BY-NC

This licence lets others distribute, remix and build upon the work, but only if it is for non-commercial purposes and they credit the original creator/s (and any other nominated parties). They do not have to license their Derivative Works on the same terms. Version 3.0 (CC Australia ported licence): [View CC BY-NC Australia Licence Deed](#) | [View CC BY-NC 3.0 Australia Legal Code](#) Version 4.0 (international licence): [View CC BY-NC 4.0 Licence Deed](#) | [View CC BY-NC 4.0 Legal Code](#)

Please note that there is the potential for minor revisions of this report.
Please check the online version at www.anrows.org.au for any amendment.

Developing a practical evaluation guide for behaviour change programs involving perpetrators of domestic and family violence

DR ANGELA NICHOLAS

Centre for Mental Health, School of Population and Global Health, University of Melbourne

DR GEORGIA OVENDEN

Centre for Mental Health, School of Population and Global Health, University of Melbourne

RODNEY VLAIS

Consultant

This report addresses work covered in the ANROWS research project BW.19.01 "Development of a best practice guide to perpetrator program evaluation". Please consult the ANROWS website for more information on this project.

ANROWS research contributes to the six National Outcomes of the *National Plan to Reduce Violence against Women and their Children 2010-2022*. This research addresses National Plan Outcome 4 –Services meet the needs of women and their children experiencing violence.

Suggested citation:

Nicholas, A., Ovenden, G., & Vlasis, R. (2020). *Developing a practical evaluation guide for behaviour change programs involving perpetrators of domestic and family violence* (Research report, 17/2020). Sydney: ANROWS.



Centre for Mental Health

Melbourne School of Population and Global Health, University of Melbourne

Level 4, 207 Bouverie Street, Carlton, Victoria 3053

Author acknowledgement

We are grateful to our review panel for their contribution to this guide: Dr Georgina Sutherland (Respect Victoria, Centre for Health Equity and Centre for Mental Health, University of Melbourne), Mr Michael Brandenburg (No to Violence), Dr Helen Jordan (Centre for Health Policy, University of Melbourne), Professor Emerita Patricia Easteal (University of Canberra, Legal Light Bulbs), Mr Jacob Pegg (Family Safety Victoria), Ilana Jaffe (Family Safety Victoria), and Dr Cathy Vaughan (Centre for Health Policy, University of Melbourne).

Acknowledgement of lived experiences of violence

ANROWS acknowledges the lives and experiences of the women and children affected by domestic, family and sexual violence who are represented in this report. We recognise the individual stories of courage, hope and resilience that form the basis of ANROWS research.

Caution: Some people may find parts of this content confronting or distressing. Recommended support services include 1800 RESPECT–1800 737 732 and Lifeline–13 11 14.

Contents

List of figures	3
List of tables	4
List of boxes	5
Acronyms	6
Executive summary	7
Background	7
Aims of the project	7
Project overview	8
Methodology	8
Recommendations for practitioners and policymakers	9
Introduction	11
Evaluation of men’s behaviour change programs	11
Background to the project	12
Structure of this report	12
State of knowledge review	13
Aims of this review	13
Methods used for this review	13
Part 1: Review of standards of practice for MBCPs in Australia	14
Part 2: Review of outcome measures	23
Conclusion	42
Methodology	44
Real-world examples	44
Consolidating information for the Evaluation guide	45
Expert review panel	45
Ethical considerations	45
Project limitations	46

THE EVALUATION GUIDE: A guide for evaluating behaviour change programs for men who use domestic and family violence	47
Introduction	47
Scoping the evaluation	51
Answering the evaluation questions	69
Ethical considerations	88
Recommendations	94
Conclusion	95
References	96
APPENDIX A: Data collection guide for real-world examples	112
APPENDIX B: Validated scales for use in MBCP evaluation	115

List of figures

Figure 1:	Components of the program logic model	55
Figure 2:	A hypothetical simple logic diagram for an MBCP adapted from a program logic by No to Violence	57
Figure 3:	A hypothetical “if ... then” program logic for an MBCP adapted from a program logic by No to Violence	58
Figure 4:	Steps in developing evaluation questions	59
Figure 5 :	Examples of evaluation questions that can be linked to the program logic in Figure 3	61
Figure 6:	Indicators to answer evaluation questions arising from the program logic in Figure 3	70
Figure 7:	Possible data collection methods used for indicators outlined in Figure 6	76

List of tables

Table 1:	Reviews of outcome measures from which we identified validated outcome measures	26
Table 2:	Examples of process evaluation questions arising from the program logic in Figure 3	62
Table 3:	Examples of outcomes evaluation questions devised from the program logic in Figure 3	64
Table 4:	Examples of questions and indicators for assessing unintended consequences of MBCPs	68
Table 5:	How to assess indicators against the SMART criteria	71
Table 6:	Advantages and disadvantages of various data collection methods	74
Table 7:	Examples of types of study design	80
Table 8:	Advantages and disadvantages of internal and external evaluators	87

List of boxes

Box 1:	A note on “outputs”	55
Box 2:	Process evaluation example	63
Box 3:	Assessing transferability of MBCPs	68
Box 4:	Non-equivalent control group design: Domestic violence perpetrator programs—Steps towards change (Kelly & Westmarland, 2015)	82
Box 5:	Longitudinal evaluation: European model for MBPC evaluations	83
Box 6:	Challenges of longitudinal evaluation: A study of the impact on men and their partners in the short term and long term of attending MBPCs (Brown et al., 2016)	84

Acronyms

ANROWS	Australia's National Research Organisation for Women's Safety
CALD	Culturally and linguistically diverse
COAG	Council of Australian Governments
DFV	Domestic and family violence
IPV	Intimate partner violence
MBCP	Men's behaviour change program
NOSPI	National Outcome Standards for Perpetrator Interventions
RCFV	Royal Commission into Family Violence (Vic)

Executive summary

Background

Men's behaviour change programs

Men's behaviour change programs (MBCPs) are largely group-based interventions that work with men to change their abusive and controlling behaviours against their partners, ex-partners or family members and to build healthy and respectful relationships (Barnett, Martinez, & Keyson, 1996; No to Violence, 2006, p. 28). MBCPs also offer advocacy and safety support for adult and (at least indirectly) child victims of men's use of domestic and family violence (DFV), and ideally operate within the context of integrated system responses.

MBCPs are now considered an essential component of long-term strategies to stop violence against women (Council of Australian Governments [COAG], 2015, p. 29); however, evidence of their success is varied, raising concerns about how change occurs, and is measured, over time (Gondolf, 2012). There is some doubt as to whether MBCPs lead to significant change in perpetrators' violent attitudes and behaviours (Corvo, Dutton, & Chen, 2008; Stover, Meadows, & Kaufman, 2009).

Evaluation of men's behaviour change programs

The importance of quality evaluation of programs implemented in response to DFV has most recently been illuminated by the Victorian Royal Commission into Family Violence (RCFV) (State of Victoria, 2016). The RCFV makes several recommendations regarding the conduct of quality, long-term, comprehensive and funded evaluations to determine the effectiveness of various programs in combating DFV in Victoria. Though these recommendations are Victoria-specific, the need for quality evaluation of MBCPs is nationwide.

To provide useful findings, evaluation of MBCPs requires not only assessment of whether desired outcomes are achieved, but also which components of the program assist in achieving those outcomes, why, and for whom. Responsibility for conducting evaluations of MBCPs often falls to personnel involved in running such programs, who frequently have

limited technical knowledge of program evaluation design, methodologies and measures, and are required to undertake evaluation activities with limited time and funding. Mindful of their needs, we aim to produce a practical guide for MBCP evaluation.

Aims of the project

We aimed to develop an easy-to-understand evaluation guide for personnel involved in the implementation of MBCPs in Australia. The purpose of the Evaluation guide is to provide information, punctuated with current, real-world examples, on how to scope an evaluation for an MBCP. This Evaluation guide will allow those commissioning an external evaluation of their program and those conducting their own in-house evaluation components to better understand the requirements of a quality evaluation. The learning objectives for users of the guide are the following:

- understand the purposes of conducting comprehensive evaluations of MBCPs
- be able to develop a program logic and articulate appropriate evaluation questions
- be aware of a range of methodologies available to answer the evaluation questions
- gain an awareness of the complexities of designing an evaluation of MBCPs within the real-world context and given constraints (e.g. time, funding or ethical considerations)
- be aware of the ethical issues that need to be considered when commissioning or conducting an evaluation of an MBCP.

The evaluation information contained in the guide is illustrated by excerpts from interviews with evaluators who have conducted evaluations of an MBCP in Australia and from a group consultation with the specialist No to Violence members' forum. This forum consisted of facilitators and program managers of MBCPs in Melbourne, Victoria.

Project overview

In order to deliver the Evaluation guide, we have conducted three main activities:

1. A state of knowledge review: the aim of this review was to summarise the Australian standards relating to MBCPs for perpetrators of DFV and the academic literature relating to psychometric properties of measures commonly used to assess the outcomes of MBCPs.
2. Interviews with evaluators of Australian MBCPs and a group consultation with the specialist No to Violence members' forum: the purpose of these interviews and consultation were to detail real-world examples of MBCP evaluation that could be used in the Evaluation guide and to assess the suitability of the planned content of the Evaluation guide for its target audience.
3. Collating and summarising key evaluation information to include in the Evaluation guide.

Each of these activities was overseen by a review panel comprising experts in the delivery of MBCPs or in evaluation of complex behaviour change programs. The purpose of this panel was to ensure the content of the Evaluation guide was correct and usable for the intended audience.

Methodology

State of knowledge review

A state of knowledge review was conducted to provide an update of the information summarised by Day, Vlasis, Chung, and Green (2019) in their Australia's National Research Organisation for Women's Safety (ANROWS) research report, *Evaluation readiness, program quality and outcomes in men's behaviour change programs*.

Review of the standards

We identified updates to the standards as they were outlined in Day et al. (2019) using a number of strategies:

- a desktop review to identify newly published standards
- existing knowledge of our research team, who were familiar

with some developments of the standards through their work in the DFV area

- contact with relevant government and other agency personnel (e.g. Stopping Family Violence in Western Australia and, in Queensland, the Office for Women and Violence Prevention, Department of Child Safety, Youth and Women) seeking any information that we were unable to find using the first two strategies.

For each state and territory where there had been updates to the standards, we provided brief contextual information; the process for review of the standards; a brief outline of the current standards and further detail regarding any specific evaluation standards; and a summary of any processes being undertaken to monitor compliance with the standards. We then highlighted similarities in standards of practice and areas that are emerging as significant in terms of regulation and practice.

Review of outcome measures

Through a broad review of the background literature, we identified three key outcome domains summarised in the United Kingdom's *Respect Outcomes Framework* (Respect, 2017) to be included in this review of outcomes measures:

- long-term changes in perpetrators' violent and controlling behaviour
- adult victims'/survivors' safety, wellbeing and freedom
- children's safety, wellbeing and family functioning.

We then conducted a scoping review (Pham et al., 2014) of large-scale and systematic reviews of measurement tools related to these domains that have been used in MBCPs, published between 2008–19. Only measures that met minimum criteria in terms of reliability and validity were included in the review (e.g. internal consistency between 0.70–0.95). We summarised the basic characteristics (purpose/construct, target group etc.), psychometric properties, and general acceptability and usability of each measure.

Real-world examples

Interviews with evaluators of MBCPs

To supplement the evaluation concepts outlined in the Evaluation guide, we included examples of current or recent evaluations of MBCPs across Australia. We conducted interviews with six evaluators from four Australian MBCPs. Broadly speaking, we gained information on how “success” for the MBCPs was determined; data collection methods; barriers and enablers to carrying out the evaluation; relationships with stakeholders, and how stakeholders were involved in the evaluation; and ethical considerations in designing the evaluation as well as ethical issues arising during the evaluation.

Group consultation with the specialist No to Violence members’ forum

We attended a specialist No to Violence members’ forum. The 13 attending members were largely MBCP facilitators and program managers. We provided a general overview of the Evaluation guide and asked for feedback. Members also discussed a range of issues related to conducting evaluation of their programs. We gained written consent to record this session, and we have included excerpts from this consultation in the Evaluation guide to illustrate relevant issues.

The Evaluation guide

To develop the content for the Evaluation guide, we drew on information from a range of sources. These included seminal texts on health and social program evaluation, academic papers on evaluation of MBCPs, general program evaluation information that is available in the public domain from reputable sources (e.g. Centers for Disease Control and Prevention’s, Program Performance and Evaluation Office webpages) published MBCP evaluation documentation, and the considerable evaluation expertise of members of the expert review panel and project team. This was supported by real-world examples (as described in the previous section).

Expert review panel

An expert review panel was formed for the duration of the project to ensure the practical guide was reviewed, and qualitative input provided, by a variety of stakeholders.

Our expert review panel comprised six representatives from organisations with specific expertise in DFV, MBCPs and/or evaluation. The panel met in person three times over the life of the project to provide input into development of the Evaluation guide and feedback on the three major deliverables of the project. They also provided other advice and feedback as requested.

Recommendations for practitioners and policymakers

Recommendations related to the Evaluation guide

- To ensure this guidance is useful and meaningful to the work of MBCP providers, we recommend a number of steps that position the outcome of this project as a first iteration of an ongoing piece of work. These may include, for example:
 - knowledge exchange and learning activities, such as a national webinar that introduces the guide and engages in dialogue about its content with a range of potential end users, including those with experience in design and evaluation of MBCPs (internal and external program providers), as well as evaluators with less experience with MBCP evaluation
 - active promotion of the guide by harnessing the networks of two main MBCP peakbodies in Australia—No to Violence (Victoria) and Stopping Family Violence (Western Australia)—and the practitioner network Services and Practitioners for the Elimination of Abuse Queensland (SPEAQ). These organisations are well positioned to promote the use of the guide among their networks of MBCP providers
 - evaluate uptake of the guide in practice using a pilot study design with a select group of MBCP providers.
- In terms of future work, we recommend broadening the current guide to include how to scope evaluations for

a range of perpetrator interventions and behavioural change programs. While this project limited its scope to MBCPs, the considerations and guidance required to support an evaluation for the range of innovative and emerging perpetrator interventions is likely to be similar to that required to support a standard MBCP evaluation.

- Despite attempts to include real-world examples of evaluations of programs that have been specifically designed for diverse communities (e.g. programs for LGBTIQ+ people and for culturally and linguistically diverse [CALD] groups), we were unable to do so. We recommend, therefore, that future iterations include broader consultation to include information relevant to evaluation of these programs.
- We recommend development of a specific guide on conducting evaluations of interventions that involve children and young people who use violence.
- While the broad framework of this guide has applicability for monitoring and evaluation activities across the whole of the perpetrator intervention system, our intended purpose was to guide approaches to evaluation at the programmatic level. We recommend, however, that this guide be used as a platform to develop a dedicated approach to support evaluation activity conducted at a systems level.
- Although the scope of this work did not extend to explore what “success” looks like in terms of outcomes from MBCPs, we note that this remains a highly contested space. There is significant disagreement and variability in terms of how outcomes are conceptualised and measured in MBCP evaluations, hindering the ability to build a comprehensive evidence base for “what works”. We therefore recommend that the Commonwealth Government commission work to develop a national outcomes framework for MBCPs and perpetrator intervention programs.

Broader recommendations

- This guide does not attempt to inform or support providers conducting evaluations of MBCPs or other perpetrator interventions specifically provided by and for Aboriginal and Torres Strait Islander men and communities. Rather, we recommend that evaluation of any MBCP program by Aboriginal and Torres Strait Islander community-controlled organisations be led by these organisations.
- Process evaluations are critical to determine if an MBCP is being implemented as intended and is “evaluation-ready” when an opportunity arises to assess the program outcomes. We recommend that state and territory governments fund MBCP provider peak bodies to support program providers to conduct periodic process evaluations, even if these providers do not have the capacity to conduct or commission outcome evaluations.

Introduction

Men's behaviour change programs (MBCPs) are largely group-based interventions that work with men who use abusive and controlling behaviours against their partners, ex-partners or family members to change their behaviour and build healthy and respectful relationships (Barnett, Martinez, & Keyson, 1996; No to Violence, 2006, p. 28). MBCPs also offer advocacy and safety support for adult and (at least indirectly) child victims of men's use of domestic and family violence (DFV), and ideally operate within the context of integrated systems responses. The primary purpose of MBCPs is to keep women and children safe by holding perpetrators accountable for their use of violence, working with them to take responsibility for their use of violence, and monitoring and responding to women's and children's risk of violence (Vlais, 2014).

MBCPs are now considered an essential component of long-term strategies to stop violence against women (Council of Australian Governments [COAG], 2015, p. 29). The growing number of MBCPs currently operating in Australia has been driven by a cultural shift in understandings of men's role in stopping violence, as well as substantial boosts to funding in some jurisdictions (Australia. Department of Social Services [DSS], 2019). Since Victoria's Royal Commission into Family Violence (RCFV), it has also become more evident that MBCPs are part of an integrated response to ending family violence, rather than a standalone "solution" (State of Victoria, 2016). MBCPs are now considered as one type of perpetrator intervention, but not the only one. Rollout of other types of perpetrator interventions, such as case management programs, and piloting of interventions targeted at specific cohorts means there are a number of interventions available to engage with men who use violence and abuse. While this state of knowledge review focuses specifically on MBCPs, we acknowledge the broader context of perpetrator interventions in which they take place.

There has been significant investment in research, policy and practice to protect women and children experiencing, and at risk of experiencing, DFV in Australia and internationally; however, the efficacy of MBCPs remains contentious (Mackay, Gibson, Lam, & Beecham, 2015). Presently, evidence of their success is varied, raising concerns about how change occurs, and is measured, over time (Gondolf, 2012). There is some doubt as to whether MBCPs lead to significant change

in perpetrators' violent attitudes and behaviours (Corvo, Dutton, & Chen, 2008; Stover, Meadows, & Kaufman, 2009). MBCPs also differ considerably across a number of dimensions, including conditions of participation (e.g. court-mandated or voluntary); referral sources; program duration and intensity; session structure (e.g. structured or open); monitoring and evaluation; and facilitator qualifications and experience (Price & Rosenbaum, 2009; Smith, Gendreau, & Swartz, 2009). This presents a range of issues in terms of evaluation of programs, and there is a growing need to identify which components of a program are most effective and/or contribute to positive outcomes for both victims/survivors and perpetrators (Eckhardt, Murphy, Black, & Suhr, 2006; Eckhardt et al., 2013).

Evaluation of men's behaviour change programs

In consideration of these complexities, the Victorian RCFV made several recommendations regarding the conduct of quality, long-term, comprehensive evaluations to determine the effectiveness of MBCPs. For example, Recommendation 88 states that

the Victorian Government [should] provide dedicated funding for future perpetrator programs. These should include evaluation studies to establish longer term effectiveness and assist in improving program design in the long term [i.e., within three years]. (State of Victoria, 2016, p. 297)

While MBCPs have a relatively short history in Australia, the increase in funding of programs at both state and national levels has been coupled with a greater impetus to measure their efficacy across a number of outcome domains (Chung, 2014; COAG, 2016). At present, there remains a strong consensus in the field that MBCPs often use inadequate outcome measures of change, and thus provide very little insight into "who", "what" and "how much" has changed following program completion (Mackay et al., 2015). The focus in MBCP evaluation on self-reported change and "short-term" risk reduction in perpetrators across a program cycle also fails to consider long-term impacts for victims/survivors and their children (Vlais, 2014).

Despite the known barriers to evaluation of MBCPs—including funding and time constraints across large- and small-scale program implementation—there is an increasing sense of responsibility for organisations to undertake “evidence-based practice” (Mackay et al., 2015, p. 30). The need for more robust evaluation, and more evaluation in general, is most evident when reviewing the dearth of Australian-based findings in this area. To date, very few MBCPs in Australia have undergone formal evaluation. As a result, there is a strong impetus to build consistency in evaluation practice.

Background to the project

In 2019, researchers from the Centre for Mental Health at the University of Melbourne were funded by Australia’s National Research Organisation for Women’s Safety (ANROWS) to develop an evaluation guide for MBCPs. The purpose of the Evaluation guide is to provide easy-to-understand information, punctuated with current real-world examples, on how to scope an evaluation for an MBCP. This Evaluation guide will allow those commissioning an external evaluation of their program and those conducting their own in-house evaluation to better understand the requirements of a high-quality evaluation. The learning objectives for users of the guide are the following:

- understand the purposes of conducting comprehensive evaluations of MBCPs
- be able to develop a program logic and articulate appropriate evaluation questions
- be aware of a range of methodologies available to answer the evaluation questions
- gain an awareness of the complexities of designing an evaluation of MBCPs within the real-world context and given constraints (e.g. time, funding and ethical constraints)
- be aware of the ethical issues that need to be considered when commissioning or conducting an evaluation of an MBCP.

The evaluation information contained in the guide will be illustrated by examples from current or recent evaluations of a range of MBCPs being conducted across Australia. To

provide context for this practical Evaluation guide, we have undertaken a state of knowledge review that precedes and informs it.

Structure of this report

This report consists of five main sections:

1. Introduction: provided above, the introduction provides an outline of the current state of MBCP evaluation in Australia and internationally. This context helps to highlight the need for the information provided in this report.
2. State of knowledge review: this review is divided into two sections: a review of the current minimum standards for MBCPs across Australia, and a review of the current academic literature relating to psychometric properties of measures commonly used to assess the outcomes of MBCPs. This section includes a short background and detailed methodology for each of these review components, as well as an overall conclusion regarding findings from both parts of the review.
3. The Evaluation guide: this section contains a detailed methodology for development of the guide, as well as the complete document.
4. Recommendations: we make recommendations regarding the next steps in terms of this Evaluation guide and future research opportunities in MBCPs and perpetrator intervention evaluation more broadly.
5. Conclusion: the conclusion summarises key learnings from each of the project components that can be used to further develop the guide, and to improve MBCP evaluation more broadly.

State of knowledge review

Aims of this review

This state of knowledge review aims to outline changes to Australian practice standards for MBCPs since Day et al.'s (2019) research, in order to:

- summarise the psychometric properties of outcome measures used recently in evaluations of MBCPs
- make a brief critical statement about each measure regarding its suitability for use in current evaluations of Australian MBCPs.

The state of the knowledge review provides an update to some of the information summarised by Day et al. (2019) in their ANROWS research report *Evaluation readiness, program quality and outcomes in men's behaviour change programs*. While this research was published in 2019, it reviewed the state of knowledge up to 2017. In their substantial review of the MBCP literature, Day et al. (2019) considered a range of issues relevant to understanding the effectiveness of MBCPs. In brief, the report reviewed current standards of practice, approaches to program evaluation, and qualitative findings from practitioners and partners regarding improvements for practice and evaluation. As the work within Australian states and territories related to MBCPs is steadily progressing, we provide an update on two key areas of the Day et al. (2019) report that are most relevant for development of the Evaluation guide. Specifically, this state of knowledge review will provide updates concerning:

- current Australian standards of practice relating to MBCPs for perpetrators of DFV (Part 1)
- a review of the academic literature relating to psychometric properties of measures commonly used to assess the outcomes of MBCPs, supplemented by a short critical review of each measure (Part 2).

What is not included in this review

It is not the purpose of this review to undertake an in-depth, critical analysis of the quality of the minimum practice standards for each state and territory, nor what should be considered as the key outcomes of MBCPs. Rather, we present a summary of changes to the standards since their review in Day et al. (2019), with a short commentary on common

themes and issues arising from the standards. In the review of psychometric properties of outcome measures, we undertake a critical appraisal of each measure, but make no overarching claims about what evaluators should or should not include as outcome domains. This review serves as a background to the Evaluation guide, and more guidance on outcome measures will be provided in that document.

Methods used for this review

To summarise the current state of the minimum standards for MBCPs (Part 1), we conducted a desktop review of all Australian states and territories, supplemented by the knowledge of our research team and review panel, and contact with relevant personnel in various states and territories. These methods identified changes to the standards since 2017, as the state of these standards in 2017 were outlined in detail in the review by Day et al. (2019). The desktop review of current minimum standards for MBCPs across Australia was last updated in June 2019.

To assess the usability and psychometric properties of outcomes measures used in MBCP evaluations (Part 2), we conducted a scoping review of the available measures. We identified measures from three primary outcome domains:

1. long-term changes in perpetrators' violent and controlling behaviour
2. adult victims'/survivors' safety, wellbeing and freedom
3. children's safety, wellbeing and family functioning. (Adapted from Respect, 2017)

This scoping review involved searches of academic and grey literature from January 2008 to May 2019 to identify reviews of outcome measures for MBCPs and for evaluations of MBCPs. We then used the identified measures to conduct further searches on individual measures, and extracted data relating to the content and use of each measure and any published data regarding their psychometric properties. We then made comment on the utility, validity and reliability of each measure.

Detailed methodologies for each of these review components are outlined further in the following sections.

Part 1: Review of standards of practice for MBCPs in Australia

Background

The purpose of Australian standards of practice for MBCPs is to specify minimum requirements for the programs and to provide guidance to ensure that programs reflect good practice and are safe and effective. In addition to a very broad set of national standards—the *National Outcome Standards for Perpetrator Interventions* (NOSPI) (COAG, 2015)—some Australian States (Victoria, Western Australia, New South Wales and Queensland) have developed their own sets of standards of practice for MBCPs. There is considerable variability in standards across states regarding their content and specificity, and whether they include specifications regarding monitoring and evaluation of the outcomes of the program.

The Day et al. review (2019) outlined the current state of the minimum standards across Australia at the time of their research and writing, which occurred in 2017. However, they noted that several states were in the process of updating or developing their standards at the time that review was undertaken, and therefore these minimum standards are likely to have undergone substantial change since then. Therefore, this section of our state of knowledge review aims to provide an overview of any changes made since 2017. A current understanding of the minimum requirements of MBCPs in Australia is important, as this has consequences for MBCP evaluation. For example, evaluation questions may include to what degree MBCPs are adhering to their relevant standards, and why or why not. Furthermore, this understanding of minimum standards provides important context as to why MBCPs may be implemented in the way they are and helps provide an understanding of any standard data that is collected across jurisdictions which could be utilised for evaluation purposes.

National standards

As outlined by Day et al. (2019), basic guidelines and support for delivery of MBCPs were first introduced in Australia in the mid-1980s. Most were framed in terms of essential education to be provided within the programs, such as ensuring that men understood the cycle of violence and their own use of power and control. Over the past two decades, the quality and evidence base for standards of practice in Australia has increased considerably.

In 2015, COAG developed the NOSPI (COAG, 2015) as part of the *National Plan to Reduce Violence against Women and their Children 2010–2022* (DSS, 2016b). Based on extensive consultation with government and non-government sector experts, the NOSPI outline core outcome areas with which interventions that work with perpetrators of DFV must ensure they comply.

The NOSPI include six headline standards:

1. Women's and their children's safety is the core priority of all perpetrator interventions.
2. Perpetrators get the right interventions at the right time.
3. Perpetrators face justice and legal consequences when they commit violence.
4. Perpetrators participate in programs and services that change their violent behaviours and attitudes.
5. Perpetrator interventions are driven by credible evidence to continuously improve.
6. People working in perpetrator intervention systems are skilled in responding to the dynamics and impacts of domestic, family and sexual violence. (COAG, 2015, p. 4)

The NOSPI reporting system assists in collecting information about overall national strategies in the area of perpetrator interventions, with the aim to monitor progress and plan future policy priorities. The NOSPI's baseline report (2015–16) included reported data from each of the states and territories against the above NOSPI headline standards (DSS, 2016).

There is little research into whether the intended outcomes of the NOSPI are achieved, and if they sufficiently influence the development and revision of program policies and

practice. In view of this, Day et al. (2019) recommended that the NOSPI be reviewed regularly to ensure there is some consistency of practice across states and territories. At the time of writing, DSS is collating feedback from states and territories undertaking data validation work. In March 2019, the Federal Government committed to further investment to reduce violence against women and children under the *Fourth Action Plan 2019–2022: National Plan to Reduce Violence against Women and their Children 2010–2022* (DSS, 2019).

While the NOSPI are important in terms of providing a reference point for current national activities, the extent to which each jurisdiction measures its progress against the NOSPI is unclear. For example, a number of jurisdictions developed their own operational standards for MBCPs well before the NOSPI were established. As outlined below, other states have also relied on their own advisory processes in developing overarching principles to guide perpetrator interventions.

Methodology

We identified any updates to the standards as they were outlined in Day et al. (2019) using a number of strategies similar to some of those used in their research. We initially conducted a desktop review to identify newly published standards. Given that one of our researchers (Rodney Vlasis) was also part of the research team who prepared the Day et al. (2019) report, we were also able to use his existing knowledge, and that of the members of our review panel who were familiar with some developments of the standards through their work in the DFV area, to identify updates to the standards. Where necessary, we also contacted relevant government and other agency personnel (e.g. Stopping Family Violence in Western Australia, and the Office for Women and Violence Prevention, Department of Child Safety, Youth and Women in Queensland) in relation to any information that we were unable to find using the other two strategies. Given that Day et al. (2019) did significant work in identifying the standards nationally, we were able to build on their work and focus on identifying changes since their review. Therefore, we did not find it necessary to replicate some components of the work conducted by Day et al. (2019).

For each state and territory where there have been updates to the standards, we provide the following information:

- **current context:** some brief contextual information regarding MBCPs and any known updates to the standards in that state
- **process for review of the standards:** a summary of the processes undertaken to revise the previous standards
- **the standards:** a brief update of changes to the standards since the Day et al. (2019) review (note that while the review was published in 2019, the research was conducted in 2017, and therefore we provide an update since that time)
- **evaluation standards:** an outline of the content of any standards relating to monitoring and evaluation
- **compliance monitoring:** how compliance with the standards is assessed and enforced, including whether these processes are outlined in a specific compliance and monitoring framework.

We also aim to highlight some of the similarities in standards of practice that exist nationally and to detail the areas that are emerging as significant in terms of regulation and practice.

Results of the review of standards

Northern Territory, Australian Capital Territory, Tasmania and South Australia

There are no updates regarding minimum standards or compliance frameworks for the Northern Territory, the Australian Capital Territory, Tasmania or South Australia as these states and territories do not currently, and have never had, minimum practice standards for MBCPs.

The Northern Territory, the Australian Capital Territory and Tasmania have not yet invested in the development of minimum standards. This has likely been influenced by their standing as small jurisdictions, particularly in terms of the number of community-based MBCP providers. These jurisdictions each have a very small number of non-government organisations (NGOs) that support the provision of MBCPs, although specific initiatives such as Marra'ka Mbarintja (Talking straight to make change), an MBCP run by the Tangentyere Council in the Northern Territory, have been

implemented and evaluated. At present, there is no specific peak body representing DFV perpetrator programs for the Northern Territory and the Australian Capital Territory, and no local training programs or professional development opportunities for practitioners specific to DFV perpetrator intervention work (beyond what might incidentally arise).

South Australia supported the development of the very first MBCP in 1983, which closely followed the commencement of programs and intellectual work in the United States (Mackay et al., 2015). This state has a long history of design and innovation in perpetrator interventions, including the support of several noteworthy academics in the field (Jenkins, 1990, 2009). Despite this, there are currently no South Australian practice standards.

While standards of practice for perpetrator intervention work in South Australia were implemented in the late 1990s (Colley, Hall, Jenkins, & Anderson, 1997), numerous setbacks have led to the gradual decline of MBCPs in the state. Most significant of these was the absence of a government agency or NGO to advocate for the specific needs of practitioners, and for the development of policy, which may have buoyed the sector (Vlais & Green, 2018). Presently, the South Australian Government funds two NGO providers to run early interventions for perpetrators who have been charged with a DFV offence through the Abuse Prevention Program, which falls under the responsibility of the Courts Administration Authority. Overall, 300 funded places are available, extending from metropolitan to regional areas.

Western Australia

Western Australia has a small number of program providers, with four key NGOs providing programs in multiple locations, as well as two smaller providers. MBCPs are largely run in Perth, with few available options for rural and remote areas. There has been no update to the 2015 minimum standards as reported in the Day et al. (2019) review. Therefore, a summary of the Western Australian standards can be found in that review. The 2015 standards, entitled *Practice Standards for Perpetrator Interventions: Engaging and Responding to Men who are Perpetrators of Family and Domestic Violence* (Western Australia. Department for Child Protection and

Family Support, 2015), are concise and share many similarities with the 2011 New South Wales standards (New South Wales. Department of Justice, 2012). The standards do not include reference to evaluation of MBCPs, although there is mention of “a commitment to evidence-based practice” (Department for Child Protection and Family Support, 2015, p. 8), including review and evaluation of the ethics principles that accompany the standards.

At present, Western Australia’s peak body for DFV perpetrator programs/interventions, Stopping Family Violence, is developing a range of policy, sector capacity-building, training and research projects. Stopping Family Violence also convenes the Western Australia Men’s Behaviour Change Network, a practitioner network for program providers to explore shared opportunities and concerns across the workforce. Stopping Family Violence, with the assistance of the Western Australia Men’s Behaviour Change Network, is delivering a training program for MBCP practitioners. This training includes online components and culminates with a 5-day, face-to-face training component. While it is not officially a competency-based training for practitioners, participants are assessed by means of role plays and other criteria. One intention of this training, which is in a pilot stage, is to develop a registration process for MBCP practitioners (Department for Child Protection and Family Support, 2015). Registration will also ensure that registered practitioners are monitored in terms of safe conduct and appropriate work.

Stopping Family Violence has written two unpublished documents outlining issues, considerations and the broad features of a desirable accreditation framework (at the program and organisational level) in Western Australia; however, the Western Australian Government has not, to date, committed to adopting an accreditation framework.

Queensland

The current context

The Queensland Government is currently implementing a 10-year reform program aimed at eliminating DFV in response to the *Not Now, Not Ever: Putting an End to Domestic and Family Violence in Queensland Report* (Not Now, Not Ever report) (Special Taskforce on Domestic and Family Violence

in Queensland, 2015). This reform is underpinned by the *Domestic and Family Violence Prevention Strategy 2016–2026* (Queensland. Department of Child Safety, Youth and Women [DCSYW], 2016a), which sets the direction for government action to deliver on the Not Now, Not Ever report.

The Queensland Department of Child Safety, Youth and Women (DCSYW) currently funds 17 services to deliver 28 MBCPs across the state (DCSYW, 2019). This includes a specific behaviour change program for Aboriginal and Torres Strait Islander men, Gatharr Weyebe Banabe (Man's Life Change).

The DFV sector in Queensland will continue to undergo significant reform as the recommendations of the Not Now, Not Ever report are progressively implemented. The development of high quality and accessible community-based perpetrator interventions complementing correctional and justice system interventions that keep perpetrators accountable for their actions and keep women safe are considered essential under these reforms.

Recent increases in the 2016–17 state budget to fund MBCPs has led to enhancements to existing programs, the establishment of additional programs to help respond to increases in demand for services, and several initiatives, including a pilot of the Walking with Dads initiative¹ and an externally funded review of Queensland's *Professional Practice Standards: Working with Men who Perpetrate Domestic and Family Violence* (DCSYW, 2016b).

The DCSYW is also developing a quality framework and monitoring process to ensure ongoing compliance with revised practice standards in line with Recommendation 82 from the Not Now, Not Ever report (see Special Taskforce on Domestic and Family Violence in Queensland, 2015, pp. 32–33).

1 This initiative is designed to improve the safety and wellbeing of family members experiencing or at risk of DFV who are currently in the statutory child protection system (with a specific focus on creating opportunities for the children of these families to return home, or to remain at home). See https://noviolence.org.au/wp-content/uploads/2020/05/WWD_Final-Summary-Report-w-ISBN.pdf

In addition, the DCSYW is exploring the use of alternative interventions while perpetrators wait to attend an MBCP, and undertaking analysis of Australian and international use of online interventions. Identified opportunities will be considered as part of future policy and planning for perpetrator intervention reforms in Queensland.

Integrated service responses have emerged as a way of strengthening and improving responses to DFV. They are intended to provide timely and holistic responses to victims and perpetrators, streamline referrals between service providers, and facilitate early interventions and support. An overwhelming theme of the Not Now, Not Ever report (Special Taskforce on Domestic and Family Violence in Queensland, 2015) was that an integrated response was essential for achieving best practice.

Process for review of the standards

Queensland's *Professional Practice Standards: Working with Men who Perpetrate Domestic and Family Violence* (DCSYW, 2016b) were largely developed by service staff and court assistance workers. The standards document is currently being updated. As outlined in the Queensland Government's *Second Action Plan of the Domestic and Family Violence Prevention Strategy 2016–17 to 2018–19* (DCSYW, 2016c), Recommendation 82 includes broadening the scope of men's behaviour change interventions in the state to include individual counselling, culturally appropriate approaches to Aboriginal and Torres Strait Islander clients, young offenders, and provision of information to respondents appearing at court (p. 19).

The standards

The Queensland *Professional Practice Standards: Working with Men who Perpetrate Domestic and Family Violence* (DCSYW, 2016b) are currently being updated.

The current practice standards relate to four primary domains of practice:

1. Coordinated responses and referral pathways: this standard states, "The service will proactively engage with government and non-government services in the community at the

local and regional level” (DCSYW, 2016b, p. 12) which includes, for example, suggestions for optimal practice around working with advocacy services and with those referring into the program.

2. Program staff: the standards include those for staff recruitment and selection, staff roles and responsibilities, safety, supervision and professional development.
3. Overall program structure and operation: the standards include those related to program accountability (clients’ rights, safety and risk assessment, advocacy work and reporting) and individual and group program practice (intake and assessment, post-intake individual practice and group practice).
4. Internal–external review and evaluation: this standard states that evaluation should be “an integral part of service delivery through ongoing practices of planning, monitoring and review” (DCSYW, 2016b, p. 33).

There is detailed guidance included under each of the headline standard indicators. Each standard includes practices that are mapped as either unacceptable, essential or optimal.

Evaluation standards

The current and soon-to-be-superseded standards include reference to evaluation, which specifically includes as essential practice “planning, monitoring and evaluation” (DCSYW, 2016b p. 33). In addition, there is a strong emphasis on performance measures to include outcomes that reflect the program’s objectives; outcomes that can be regarded as “sustainable” beyond the life of the program; specific measures that address the accessibility of service provision to the target groups, including Aboriginal and Torres Strait Islander peoples and CALD populations; and evaluation of cross-cultural competencies of staff. The standards also outline as optimal practice that partnerships be made with tertiary institutions with appropriate expertise to evaluate the outcomes of MBCPs, and that external evaluation of program effectiveness be sought.

Compliance monitoring

The development of the quality framework and revised standards will assist the Queensland Government in

implementing the first step of a more formal compliance monitoring process into the future.

Victoria

The current context

The original, and since revised, documentation of minimum standards for MBCPs was produced in Victoria by No to Violence (the peak body for organisations and individuals working with men to end family violence in Victoria and New South Wales) between 1994–96. At this time, the Victorian-based MBCP field already had a large number of programs in operation. While No to Violence’s 1996 standards have had considerable influence nationally, they were largely generated to guide practitioners and thus focus chiefly on requirements for safe practice and key concepts required to be delivered within an MBCP framework. The original standards were updated by No to Violence in 2006, and this update incorporated some evidence-based literature. However, given the infancy of this field of work, it largely focused on consultations with program providers and practitioners in the field. In this way, it neglected consultation with other sectors (e.g. child protection, community corrections and police), which is integral to the development of minimum standards concerning interagency collaboration and risk management.

In 2017, Family Safety Victoria, Monash University, No to Violence and No to Violence’s MBCP members reviewed the standards to align them with Victoria’s current practice environment. This included more detail in relation to program length (particularly as it related to long-term outcomes), information sharing, the inclusion of a standalone family safety contact worker, and extending the standards to respond to the increasing diversity of men in MBCPs. The final and current standards, *Men’s Behaviour Change Minimum Standards* (Family Safety Victoria, 2017), were implemented in July 2018.

Process for review of the standards

The 2015 Victorian RCFV was instrumental in raising the issue of standards of practice for MBCP. Importantly, the RCFV highlighted that the implementation of practice standards

in this area is inconsistent, particularly in terms of program content evaluation (State of Victoria, 2016). Recommendation 91 of the RCFV recommends that:

The Victorian Government, in consultation with No to Violence [within 12 months]:

- review and update the *Men's Behaviour Change Programs Minimum Standards* to reflect research findings, national and international best practice, and the central importance of partner contact work
- develop a compliance framework, incorporating an accreditation process, for providers of men's behaviour change programs. (State of Victoria, 2016, p. 300)

In line with this recommendation, the recent review of the minimum standards by Family Safety Victoria has resulted in considerable changes to the structure of the previous standards.

The standards

The revised minimum standards, outlined under ten key principles, have been organised according to the *Victorian Principles for Perpetrator Interventions* developed by the Expert Advisory Committee on Perpetrator Interventions (standards 1–8) (State of Victoria, 2018) and two standards taken from the NOSPI (9–10) (Family Safety Victoria, 2017). Family Safety Victoria is a Victorian Government agency established in July 2017 to ensure delivery of family violence response reforms following the Victorian RCFV. As a result, the standards represent a combination of high-level statements and more detailed prescriptions.

The key principles are as follows:

1. Victims', including children's, safety and freedom underpins all interventions with perpetrators of family violence.
2. Interventions with perpetrators are informed by victims and the needs of family members.
3. Perpetrators take responsibility for their actions and are offered support to choose to end their violent behaviour and coercive control.
4. Interagency risk assessment and risk management processes are consistent, robust and strong, and any risk associated with intervention is minimised.

5. Perpetrators are kept in view through integrated interventions that build upon each other over time, are mutually reinforcing, and identify and respond to dynamic risk.
6. Responses are tailored to meet the individual risk levels and patterns of coercive control by perpetrators, and address their diverse circumstances and backgrounds which may require a unique response.
7. Perpetrators face a range of timely system responses for using family violence.
8. A systems-wide approach collectively creates opportunities for perpetrator accountability, both as a partner and a parent. Actions across the system work together, share information where relevant, and demonstrate understanding of the dynamics of family violence.
9. People working in perpetrator intervention systems are skilled in responding to the dynamics and impacts of domestic, family and sexual violence.
10. Perpetrator interventions are driven by credible evidence to continuously improve. (Family Safety Victoria, 2017, pp. 4–5)

Evaluation standards

Standard 10 of the revised standards, which addresses the issue of gathering evidence about program effectiveness, is broad rather than directive. The key message is that providers conduct operational reviews (every 12 months) and include quantitative and qualitative data from perpetrators, partners and children, and other stakeholders (Family Safety Victoria, 2017, p. 14). The practice guide does list several examples of "assessments" that may be used in evaluation, including feedback from family members regarding perpetrator attitudes/acceptance of responsibility. However, this information is general rather than operational, and does not include detail regarding specific measures, or what measures may be useful in particular evaluation contexts.

The *Implementation Guide*, developed by No to Violence and released in August 2018, accompanies the Family Safety Victoria standards and outlines that "program providers should regularly collect and analyse information related to the safety of women and children and other women in intimate relationships with perpetrators and children

associated with that relationship” (No to Violence, 2018, p. 81). However, this information is framed in relation to basic feedback from group participants, in line with the *Family Violence Risk Assessment and Risk Management Framework* (Victoria. Department of Human Services, 2012).

Compliance monitoring

It is a requirement within funding agreements with the Victorian Government that providers deliver programs in line with the minimum standards. Compliance with the standards is managed by Victorian Government purchasing agencies, including the Department of Health and Human Services, Corrections Victoria and Court Services Victoria.

The minimum standards have also supported the work of the following programs: Kornar Winmil Yunti, in South Australia; Cross Border Indigenous Family Violence Program, evaluated multiple times on its operations in South Australia, Western Australia and the Northern Territory; and the Tangentyere Family Violence Prevention Program in Alice Springs, Northern Territory.

New South Wales

Current context

The New South Wales Department of Justice released its MBCP minimum standards (updated from the 2012 version), *Practice Standards for Men’s Domestic Violence Behaviour Change Programs*, in 2017 (New South Wales. Department of Justice, 2017), but they are technically still not in effect. A compliance framework, which was finalised in late 2019, will accompany the standards (New South Wales. Department of Justice, 2018). However, the 2012 standards still apply until the compliance framework comes into effect. A practice guide was commissioned by the New South Wales Government, entitled *Towards Safe Families: A Practice Guide for Men’s Domestic Violence Behaviour Change Programs* (No to Violence & Red Tree Consulting, 2012), to be used by program providers as a guide towards implementing the standards. There are currently no plans to update this document.

Process for review of the standards

The New South Wales practice standards were reviewed in 2016 using a consultation process that involved providers and facilitators of MBCPs, peak bodies, advocacy groups and relevant government agencies. As a result of this review, and to align with current policy, there have been some changes to the standards initially created in 2012. They do, however, retain their “focus on safety” (New South Wales. Department of Justice, 2017, p. 5).

The standards

The revised minimum standards are only moderately different from the 2012 version. While some of the standards are streamlined, they essentially remain unchanged. Like in the new Victorian minimum standards, there is a stronger focus on addressing the needs and experiences of children. In the revised standards there is also the expectation that program providers will take into account the risk–need–responsivity (RNR) principles when designing and delivering programs (see more information below).

In addition, the revised standards now have no minimum in terms of program length. This revision is intended to allow program providers to deliver programs more flexibly.

The standards document states six principles, each with 1–8 accompanying standards. The principles guiding the practice standards are:

1. The safety of victims, including children, must be given the highest priority.
2. Victim safety and perpetrator accountability and behaviour change are best achieved through an integrated service response.
3. Effective programs must be informed by a sound evidence base and subject to ongoing evaluation.
4. Challenging domestic and family violence requires a sustained commitment to professional practice.
5. Men responsible for domestic and family violence must be held accountable for their behaviour.
6. Programs will respond to the diverse needs of the participants. (Department of Justice, 2017, p. 6)

Evaluation standards

The New South Wales standards include a focus on evaluation: “Effective programs must be informed by a sound evidence base and subject to ongoing evaluation” (Principle 3) (Department of Justice, 2017, p. 13). One of the practice standards of this principle (Standard 3.2) specifies that “MBCP providers will apply the risk, needs, responsivity (RNR) principles to the program design” (Department of Justice, 2017, p. 13).

The RNR model (Andrews, Bonta, & Hoge, 1990), traditionally used in corrections, is currently being adapted for use in MBCPs. The RNR is a model of offender management that aims to reduce re-offending. The model encourages use of “actuarial” risk assessment in combination with professional opinion to determine risk of re-offending. To reduce re-offending, the model states that program intensity should be matched to risk of re-offending (risk); the program should focus on factors directly related to offending behaviour (need); the program should match the offender’s learning style and needs (responsivity); and programs should be delivered as intended (program integrity) (New South Wales. Department of Justice, 2017).

Compliance monitoring

The New South Wales Government has implemented a “registration” process. The Day et al. (2019) review describes briefly how this registration process operates. In summary, program providers submit documentation to New South Wales Justice to demonstrate that they are compliant with the minimum standards. It is envisioned that the paperwork will be reviewed, but there is little detail regarding how providers will be impacted in the long term.

The New South Wales Government compliance framework (which did not accompany the 2012 standards) provides some direction to program providers in relation to the types of evidence they will need to supply to demonstrate compliance with the minimum standards. Once the compliance framework is made operational, program providers will have a period of time to reconfigure their programs (if necessary) to meet the revised standards and provide evidence. The compliance framework document outlines this intended process (see Department of Justice, 2018).

Emerging themes and issues

Commonalities across standards

The first standard of the NOSPI has an emphasis on the freedom and safety of victims, including children (COAG, 2015). This standard is echoed in the headline standards in New South Wales and Victoria in particular, although reference is also made in other standards documents across Australian states. The focus on the safety of women and their children as the core priority of all perpetrator interventions is necessarily broad and encompasses wider issues of risk and safety. However, this overarching standard also speaks to the need to prioritise outcome measures that can capture “impacts” in terms of the lives of women and children specifically (Vlais & Green, 2018). This includes adequately addressing the intended outcomes of MBCPs in terms of content and approach (Family Safety Victoria, 2017, p. 8). At present, there are several issues pertaining to outcome frameworks and outcome measures that mean that women’s and children’s safety are not being prioritised in terms of measuring program impact.

Monitoring compliance to standards

There has been an effort in some Australian states to address compliance levels to specific standards, particularly as they relate to safety procedures. For example, the New South Wales Government’s implementation of a registration process for program providers was designed to monitor compliance with the minimum standards. However, as outlined in Day et al.’s (2019) review, compliance mechanisms that rely on funding bodies to obtain information from funded services represent an ineffective form of compliance-checking. Most funding bodies do not have the expertise to interpret information provided by program providers in relation to meeting the minimum standards. Further, funding bodies rarely have the time or capacity to liaise with each funded service to the depth required to interpret this information.

Broad versus prescriptive standards

As outlined in the Day et al. (2019) review, there is considerable variability in the detail included in standards across Australian jurisdictions. The primary aim of standards for MBCPs is to ensure consistency in terms of the quality of MBCPs, as well

as quality standards of practice. There is acknowledgement that, while increasing the likelihood of compliance to the standards, broad standards ultimately do little in terms of guiding program effectiveness, including guiding the conduct of quality evaluations (Vlais & Green, 2018).

On another level, Day et al. (2019) suggested that practice standards which propose broad principles for the conduct of MBCPs may be more useful and accessible to agencies than highly prescriptive mandatory standards, which may be difficult to apply to the necessarily broad range of MBCPs. From this perspective, the setting of high and prescriptive standards will inevitably mean that fewer programs meet the criteria required for them to be delivered.

Compliance and the need for capacity-building

Day et al. (2019) underlined that program providers need to be actively supported through a capacity-building process to meet minimum standards. Some jurisdictions have attempted to address this issue by differentiating between minimal safe practice, and optimal practices. This was the purpose of Victoria's *Implementation Guide* (No to Violence, 2018), which was developed to support providers to meet the new standards and included training across the state. Substantial funding increases in Victoria have also supported the implementation of the revised minimum standards.

Need for standards that address program- and system level-inputs

There has been a recent shift in the DFV sector in terms of widening the reach of minimum standards to target services (such as mental health services, alcohol and other drugs services, housing and homelessness services) that do not specialise in perpetrator interventions (nor even in responding to DFV) but are linked in terms of their contact, and provision of services to, perpetrators. This is the purpose of Victoria's Multi-Agency Risk Assessment and Management (MARAM) framework reforms (Family Safety Victoria, 2018). The framework will outline broad roles and responsibilities for non-specialist DFV agencies when identifying and responding to perpetrator-driven risk, ensuring consistent messaging and approaches.

The MARAM framework is unique in the Australian context (and possibly internationally) for including services that currently sit outside of mainstream MBCP practice. While the impact of this model of standard provision is yet to be evaluated, there may be strengths in increasing exposure of related services to minimum standards. Ostensibly, the broadening of programs and practice relevant to the MBCP standards may mean that services that have contact with perpetrators will be influenced by the standards and compliance frameworks. This message is consistent with a growing awareness both nationally and internationally that services that both directly and indirectly engage perpetrators also engage in practices to safeguard women and their children. There are also disadvantages to applying standards to such a wide range and volume of services. Monitoring compliance with the standards in any sort of statewide, comprehensive audit is difficult.

Allowing for innovation of non-traditional MBCPs

There has also been a shift in the DFV sector to accommodate a wider range of specialist perpetrator interventions in minimum standards to allow organisations to be more innovative in their provision of MBCP services, including to specific cultural cohorts. This format differs significantly from the usual presentation of standards, which often position cultural diversity practices and issues of intersectionality as a consideration rather than an overarching principle of MBCP practice. This supports the arguments presented in Day et al. (2019) regarding the need for standards to speak to both organisation and program levels.

Summary

The evolution of MBCP minimum standards in Australia, over the past 20 years, has followed a broad trajectory from prescriptive, detailed standards towards higher level principles and a focus on statements of intent, with an increasing focus on the importance of women's and children's safety as the primary outcome of MBCPs. This shift recognises the multiple ways in which some standards can be met, enabling a degree of creativity and adaptiveness to local contexts. This shift, however, places an even greater degree of onus on the development of compliance monitoring systems that assist program providers to put into practice, and demonstrate,

locally determined ways of meeting these standards. At the time of this review, jurisdiction-based compliance monitoring systems vary greatly, but largely do not reflect what Day et al. (2019) identified as quality accreditation system practice.

Part 2: Review of outcome measures

Background

Reviews of current MBCP evaluations have consistently revealed a shortfall in meeting requirements for evaluation best practice. One key area in which evaluations fail to meet requirements for high-quality evaluation of MBCPs is in the way “success” or the long-term, desired outcomes are measured (Axford, Elliott, & Little, 2012; Mihalic & Elliott, 2015; Westmarland, Kelly, & Chalder-Mills, 2010). As outlined in the Day et al. (2019) review, program providers often describe difficulties in identifying appropriate measures and tools for assessing the impacts and outcomes of MBCPs. In their survey of European perpetrator programs, Geldschläger, Ginés, Nax, and Ponce (2013) note that one significant barrier to measuring outcomes noted by MBCP personnel was a lack of “methodology” for doing so. Geldschläger et al. (2013) found 65 percent of MBCP personnel consequently stated their need for a “toolkit with methodologies” (p. 56) to improve their outcome measurement. Poor measurement of outcomes generally relates to two key issues:

1. use of process and short-term impact indicators without the measurement of longer-term outcomes (how outcomes are measured)
2. use of outcomes that fail to measure a range of factors imperative to the assessment of success of MBCPs (what outcomes are measured). (Vlais & Green, 2018)

Use of process and impact indicators without measures of long-term change

Many evaluations of MBCPs do not sufficiently attend to measuring long-term program outcomes, and barriers to long-term outcome measurement are many, including restrictive timeframes for reporting of evaluation outcomes as well as prohibitive costs and lack of staffing to do so (Geldschläger

et al., 2013). There has, however, been significant progress in outcome measurement in some jurisdictions, particularly toward the inclusion of ratings of women’s safety by partners and ex-partners of perpetrators of DFV. Notwithstanding these developments, when evaluations of MBCPs do measure long-term outcomes, they often use recidivism data to measure program effectiveness (Blatch, O’Sullivan, Delaney, van Doorn, & Sweller, 2016; Day et al., 2019; Day, Vlasis, Chung, & Green, 2018; Migliore, Ziersch, & Marshall, 2014; Vlasis, Ridley, Green, & Chung, 2017). Given the complexities involved in obtaining records of recidivism, however, program providers often focus on intermediate impacts including self-reported abusive behaviour and attitudes as indicators of success of the program (Day & Casey, 2010). Prior research suggests that self-reported, short-term change in behaviour is unlikely to be indicative of sustained attitudinal and behavioural change, or indeed, even of current behaviour. As noted by Flood (2019), there are often disproportionate differences in self-reports taken from perpetrators and female victims/survivors, raising concern about the validity of perpetrator self-reported data in evaluation contexts. Others have noted that self-report measures are also limited because of outdated and unvalidated psychological tests that have not been specifically designed for use with offenders or perpetrator populations (Wakeling & Barnett, 2014).

Measurement of a range of key outcomes

Traditionally, outcome domains have focused on “evidence” and “change” in relation to a perpetrator’s use of violence (Chung, 2014). However, there has been a significant shift in understandings of how to best measure “success”, centred on the promotion of women’s and children’s safety (Kelly & Westmarland, 2015). The use of recidivism alone as a measure of long-term behaviour change is no longer considered a good indicator of men’s behaviour change following MBCPs or of the safety of women and children (Vlais & Green, 2018). Recidivism fails to capture perpetrators’ use of a range of other tactics to control women and their children, such as emotional and financial controls and sabotaging women’s relationships with their support networks, which MBCPs ultimately aim to address (Vlais & Green, 2018; Walby et al., 2017).

Perhaps the most recognised evaluation of “success” in MBCPs is Project Mirabal (Kelly & Westmarland, 2015). Project Mirabal was a 6-year, United Kingdom-based project that investigated

the extent to which perpetrator programmes reduce violence and increase safety for women and children, and the routes by which they contribute to coordinated community responses to domestic violence. (Kelly & Westmarland, 2015, p. 7)

The project included interviews with female partners and ex-partners, male participants, practitioners and funders to identify what success in their MBCP meant to them. The results revealed that women were focused on six outcomes of success, which moved beyond stopping the violence:

1. respectful communication
2. expanded space for action for women, which restores their voice and ability to make choices, while improving their wellbeing
3. safety and freedom from violence
4. safe, positive and shared parenting
5. enhanced awareness of men in MBCPs about the impact of violence on others
6. safer lives for their children.

The key outcome measures that emerged from the Project Mirabal study have been integrated into a number of practices and frameworks. Most significant of these is the United Kingdom’s *Respect Outcomes Framework* (Respect, 2017), developed in recognition of the need for organisations to demonstrate that their interventions have a positive benefit. The framework includes reference to all six of the outcomes of success from the Project Mirabal study, condensed into the following primary outcome domains:

- reduction in perpetrators’ violent and abusive behaviour
- increase in victims’/survivors’ safety, wellbeing and freedom
- improvement in children’s wellbeing and safety.

Two additional indicators were included in the framework to capture:

- improvement in multiagency work
- effective targeting of interventions. (Respect, 2017, p. 1)

We chose the three outcome domains from Project Mirabal to guide our choice of content for the review of outcome measures. To our knowledge, the Respect framework is the only existing MBCP-focused outcomes framework in either the grey or academic literatures. The United Kingdom MBCP sector has some similarities to the Australian context, so we considered it apt to use this framework.

Aims of this review of outcome measures

The aim of this part of our state of knowledge review is to identify validated outcome measures of constructs related to the three outcome domains of the United Kingdom’s *Respect Outcomes Framework* (Respect, 2017) and to briefly assess their usefulness for Australian MBCPs. The three outcome domains to be addressed in this review are therefore as follows:

- long-term changes in perpetrators’ violent and controlling behaviour
- adult victims’/survivors’ safety, wellbeing and freedom
- children’s safety, wellbeing and family functioning. (Adapted from Respect, 2017)

Methodology

To identify validated outcome measures fitting within our three target domains, we began with a search of large-scale and systematic reviews of measurement tools used in MBCPs, and for DFV and intimate partner violence (IPV). The term IPV has a narrower definition than DFV, and is restricted to violence within intimate relationships, while DFV includes abuse of any member of a household, including children (World Health Organization & Pan American Health Organization, 2012).

Key search terms used were: men’s behaviour change program OR batterer intervention program OR perpetrator intervention OR perpetrator treatment OR domestic violence program OR perpetrator treatment program; AND measurement tool* OR measurement outcome* OR measurement instrument OR outcome measure* OR outcome scale* OR evaluation measure OR evaluation outcome*. Two search levels were carried out: 1) title and 2) abstract.

We included only reviews that met the following inclusion criteria:

- referred to specific outcome measures used (rather than to broad categories of measures without naming an individual measure, e.g. “self-reported data on repeated perpetration and substance use”)
- differentiated between measures or tools that had been validated versus those that had not
- included assessments of psychometric characteristics of the identified measures and tools (reliability and validity)
- included measures or tools that had been used previously to assess MBCPs, or aspects of DFV and IPV more broadly.

The scoping review included academic and grey literature concerning evaluations of MBCPs published between 2008 and 2019. All searches of the academic literature were conducted using the following databases: Scopus, PsycINFO, ProQuest, the Cumulative Index to Nursing and Allied Health Literature (CINAHL), and Google Scholar. The search for grey literature was conducted using the Google search engine.

Using the identified literature, we then conducted the following processes:

- identified the validated outcome measures most frequently used in evaluations of MBCPs (or “perpetrator interventions” or “batterer programs” as they are referred to in Europe and the United States)
- summarised basic characteristics of each measure (purpose/construct, target group etc.)
- extracted specific psychometric data (e.g. Cronbach’s alpha) and described other psychometric properties.

Given the vast number of outcome measures identified in the literature, we present here only those measures that met minimum criteria in terms of reliability and validity. However, we acknowledge that some measures which have not been subject to reliability and validity testing are used in MBCP evaluation studies, and that some such measures may be worthy of consideration for use in MBCP evaluation in some circumstances.

Reliability refers to the consistency of a measure over time (test–retest reliability), across items (internal consistency), and across different researchers (inter-rater reliability) (Salkind, 2010, p. 117). That is, “a measure is considered reliable if it would give you the same observation over and over again” (Trochim, Donnelly, & Arora, 2016, p. 119). Convention in academic research states that reliability, most commonly reported as a Cronbach’s alpha statistic (internal consistency) or test–retest reliability, must be between 0.70–0.95 for a scale to be considered a reliable measure of the construct of interest (Bland & Altman, 1997; Graham, 2006). This convention was adopted here, and we only report on those outcome measures that have reported evidence of good reliability.

Validity is “the extent to which the scores from a measure represent the variable they are intended to” (Price, Jhangiani, & Chiang, 2015, p. 121). There are four types of reported validity: content, construct, criterion and factorial. Content validity is the appropriateness of the items used to measure a construct, and this is usually determined by the opinions of experts in the area (Litwin, 1995). Construct validity is the extent to which the measure assesses the theory or construct it purports to measure (Groth-Marnat, 2009). This is often determined by seeing how well it relates to other measures of the same construct and how poorly it relates to measures of unrelated constructs (Bruce, Pope, & Stanistreet, 2018). Criterion validity relates to how well the measure predicts future behaviour and how well it performs against other measures of the same construct (Goodwin, 2010). Factorial validity is the “degree to which the measure of a construct conforms to the theoretical definition of that construct” (Miles & Gilbert, 2005, p. 266). This is often determined by factor analysis, which assesses whether the items of a measure assess a single construct, and any “sub-sets” of constructs it is designed to measure. Validity is often described qualitatively (e.g. authors of a review stating that a measure shows “evidence of content, construct, criterion and factorial validity”) and where available, we also report on the validity of the identified measures.

Table 1: Reviews of outcome measures from which we identified validated outcome measures

Author (year)	Title and publication
Bates, Graham-Kevan, Bolam, & Thornton (2017)	A review of domestic violence perpetrator programs in the United Kingdom. <i>Partner Abuse</i> , 8(1), 3-46. https://doi.org/10.1891/1946-6560.8.1.3
Bloom (2008)	<i>Violence against women and girls: A compendium of monitoring and evaluation indicators</i> . Chapel Hill, NC: MEASURE Evaluation
Canales, Geldschläger, Nax, & Ponce (2015)	European perpetrator programmes: A survey on day-to-day outcome measurement. <i>Studia Humanistyczne AGH</i> , 14(2), 33-52. https://doi.org/10.7494/human.2015.14.2.33
Dwyer (1999)	Measuring domestic violence. <i>Journal of Offender Rehabilitation</i> , 29(1-2), 23-33. https://doi.org/10.1300/J076v29n01_02
Geldschläger, Ginés, Nax, & Ponce (2013)	<i>Outcome measurement in European perpetrator programmes: A survey</i> . Berlin: Dissens Institute for Education and Research
Graham, Sahay, Rizo, Messing, & Macy (2019)	The validity and reliability of available intimate partner homicide and reassault risk assessment tools: A systematic review. <i>Trauma, Violence, & Abuse</i> , 20(1), 1-23. https://doi.org/10.1177/1524838018821952
Murray & Graves (2013)	<i>Responding to family violence: A comprehensive, research-based guide for therapists</i> . New York: Routledge
Ronan, Dreer, Maurelli, Ronan, & Gerhart (2013)	<i>Practitioner's guide to empirically supported measures of anger, aggression, and violence</i> . Cham, Switzerland: Springer International Publishing
Sprague et al. (2016)	Outcome measures for evaluating intimate partner violence programs within clinical settings: A systematic review of the literature. <i>Trauma, Violence, & Abuse</i> , 18(5), 508-522. https://doi.org/10.1177/1524838016641667
Thompson, Basile, Hertz, & Sitterle (2006)	<i>Measuring intimate partner violence victimization and perpetration: A compendium of assessment tools</i> . Atlanta: Centers for Disease Control and Prevention, National Center for Injury Prevention and Control
Walby et al. (2017)	<i>The concept and measurement of violence against women and men</i> . Bristol: Policy Press.

Results of the review of outcome measures

Reviews of outcome measures

Table 1 shows the reviews of outcome measures and tools from which we identified validated outcome measures.

Outcome measures and tools

In the following section, we outline the main features of the outcome measures and their psychometric properties that we identified in the three domains of interest:

1. long-term changes in perpetrators' violent and controlling behaviour
2. adult victims'/survivors' safety, wellbeing and freedom
3. children's safety, wellbeing and family functioning. (Adapted from Respect, 2017)

A number of measures have been validated but are not appropriate for use in the context of MBCPs. Within the

critical review section, we therefore provide an overview of the various strengths and weaknesses of each measure.

Risk assessment measurement tools

Risk assessment is a comprehensive appraisal that involves gathering a range of information to assess levels of risk. This information includes patterns of perpetrator behaviour, violence and use of coercive control; and the presence of protective factors for the woman and any child exposed to DFV (Toivonen & Backhouse, 2018). Typically, a risk assessment will attempt to predict the likelihood of violence, and other factors such as the nature, severity, frequency and imminence of further violence (de Vogel, 2005). While risk assessment instruments are primarily used to assist victims/survivors with decisions about self-care and safety, they are increasingly being administered to determine the suitability of perpetrators for MBCPs (Graham et al., 2019) and as an outcome measure pre- and post-intervention. Geldschläger et

al.'s (2013) review found that of the 70 programs that indicated they used questionnaires to assess outcomes of interventions, more than half (56%) used a risk assessment measure, which they completed at program entry, then repeated at program completion and (in rare cases) at 6 months follow-up.

Geldschläger et al. (2013) note in their review that risk assessments can “potentially” be used to produce pre-intervention scores in evaluations, and to measure changes in the level of risk men pose to their partners at assessment and immediately post-MBCP. The risk assessment measures listed here have been detailed as they meet our inclusion criteria (i.e. they collect data regarding the existence of violence/violent behaviours and abuse and have some published psychometric properties). However, given that many risk assessments only examine the presence or absence of violence (rather than frequency and severity), the data captured in risk assessment need to be subsidised with other, more robust outcomes measures, including reports from victims/survivors. Ultimately, risk assessment tools do not adequately capture behavioural (and attitudinal) changes in perpetrators and/or the safety, wellbeing and freedom of women and children.

Spousal Assault Risk Assessment (SARA)

The Spousal Assault Risk Assessment (SARA) (Kropp, Hart, Webster, & Eaves, 1994) is a 20-item risk assessment tool for spousal violence to be completed by professionals. The assessment includes two parts: Part 1 (factors 1–10) relates to general violence risk (e.g. recent substance abuse/dependence, recent suicidal or homicidal ideation/intent), and Part 2 (factors 11–20) relates specifically to risk of spousal violence (e.g. past physical assault, or past sexual assault/sexual jealousy). This assessment includes a two-step process. First, each of the 20 items are coded as absent (0), subthreshold (1) or present (2). Second, the items are assessed as either “critical” or “not critical”, based on whether the individual poses a risk of harm. The final score of “low”, “moderate” or “high” is generated by the evaluator.

Psychometric properties

Reported Cronbach's alpha (α) coefficients range from .62 to .84. Kropp and Hart (2000) report moderate levels of internal consistency and item homogeneity, high inter-rater reliability

(.84), and good convergent and discriminant validity with respect to other measures related to risk for general and violent criminality (Kropp & Hart, 2000).

Critical review

The SARA is designed to be completed by a professional and is often not directly informed by a woman's perception of her own risk or safety (Kropp, 2008). When completion of the SARA is informed by the victim's/survivor's experience, there is concern that women at greatest risk may minimise their experiences out of fear of other consequences. As a result, completion of the SARA places a relatively heavy burden on users in terms of time, technical expertise, and case history information (Belfrage et al., 2012; Kropp, Hart, & Belfrage, 2005). The SARA is effective in terms of assessing risk of violence as part of initial intake; however, it is not an effective measure of men's behavioural change in the context of MBCPs.

The Revised Danger Assessment

The Revised Danger Assessment (DA) (Campbell, 2004) is a 21-item assessment scale, divided into two sections. The DA is designed to be completed by victims. The first section is a calendar that asks victims to record the frequency and severity of violence over the past 12 months. The second section includes a 20-item yes/no list of risk factors that are associated with high risks of partner lethality (e.g. “Has he ever forced you to have sex when you did not wish to do so?” and “Does he ever try to choke you?”).

Psychometric properties

Initial evaluations of the DA using women victims/survivors of violence show an alpha coefficient of 0.71 and test-retest reliability ranging from 0.89–0.94 (Campbell, 1986; Campbell et al., 2003).

Critical review

The DA is a risk assessment measure and is not designed for measuring behaviour change in the context of MBCPs. The DA collects only “yes/no” data. The instrument includes numerous United States terms and assumptions regarding

violence (e.g. use of the term “homicide” and frequent reference to gun use) that may not be suitable for use in Australia.

Violence Risk Assessment Scheme (HCR-20)

The Violence Risk Assessment Scheme (HCR-20) (Douglas, Webster, Hart, Eaves, & Ogloff, 2001; Douglas, Hart, Webster, & Belfrage, 2013) is a comprehensive assessment of violence risk. It consists of 20 probing questions about the person being evaluated for violence risk and is completed by the perpetrator. The HCR-20 assesses three main domains—historical, clinical, and risk management—coded with a rating of 0 (not present), 1 (possible/less serious), or 2 (definite/serious). The HCR-20 has been subjected to more than 200 empirical evaluations based on more than 150 independent data sets (Nicholls, Pritchard, Reeves, & Hilterman, 2013).

Psychometric properties

Although there is some variability in the findings, the HCR-20 demonstrates acceptable inter-rater reliability of both the risk factors and the final summary risk judgment (Cutler, 2008, p. 354). Meta-analytic evidence has found that it performs as well or better than other approaches to risk assessment (Nicholls et al., 2013).

Critical review

While the HCR-20 is a flexible violence risk assessment tool designed to meet the needs of various populations and settings, it has not been specifically designed for those affected by DFV (Douglas & Reeves, 2010). The usefulness of the HCR-20 as a tool to measure outcomes in MBCPs is questionable. Validated tools for samples of DFV victims/survivors should be sought as they offer the best measure of change.

Readiness to change measurement tools

The importance of men’s readiness to change is now well established in the MBCP literature as it has been shown to impact rates of program attrition and the likelihood of men reoffending (Eckhardt, Holtzworth-Munroe, Norlander, Sibley, & Cahill, 2008; Scott & Wolfe, 2003). Readiness may be viewed at least in part as a perpetrator’s willingness to accept responsibility for violence (Day et al., 2019). Men who

attend MBCPs may have wide-ranging attitudes in relation to their perceived necessity to change (Day & Casey, 2010; Levesque, Gelles, & Velicer, 2000). Perpetrators in the more advanced stages of readiness to change are most likely to use strategies learned in the MBCP and less likely to blame their partners for their violence (Eckhardt, Babcock, & Homack, 2004; Levesque et al., 2000; Scott & Wolfe, 2003).

While measures of men’s perceived readiness to change have been used to evaluate the appropriateness of MBCPs at intake, the usefulness of readiness to change measures in the context of men’s long-term behaviour change is not well established. For this reason, readiness to change measurement tools would not be appropriate for use as an exclusive measure of outcomes in MBCPs.

Safe at Home Instrument (SAHI)

The Safe at Home Instrument (SAHI) (Begun et al., 2003) is a 23-item scale, with three subscales: pre-contemplation, contemplation and preparation/action. The scale is grounded in Prochaska and DiClemente’s (1983) Transtheoretical Model of change. It is designed to measure the extent to which perpetrators of IPV are ready to change their IPV behaviours. Pre-contemplation items assess men’s reluctance, including denial (e.g. “It’s no big deal if I lose my temper from time to time”), partner blaming (e.g. “It’s her fault that I act this way when we disagree”), and motivation to attend intervention (e.g. “I’ll come to groups, but I won’t talk”). Contemplation items assess client motivation to change ways of dealing with anger and conflict (e.g. “It’s time for me to listen to the people telling me I need help”). Preparation/action items assess clients’ use of a variety of change strategies (e.g. “I try to listen carefully to others so that I don’t get into conflicts anymore”).

Psychometric properties

Internal consistency for the three subscales vary, and was reported as: contemplation ($\alpha=.91$); pre-contemplation ($\alpha=.59$); preparation/action ($\alpha=.79$) (Begun et al., 2003). The original SAHI was revised in 2009, and internal consistency was reported as: contemplation subscale, $\alpha=.90$; preparation/action subscale, $\alpha=.77$; maintenance factor, $\alpha=.79$ (Begun et al., 2009).

Critical review

In the earlier version of this scale, Begun et al. (2003) found a relationship between preparation/action scores and social desirability, which may mean perpetrators misrepresent their desire for change in their entry to an MBCP. This instrument was developed and evaluated with perpetrators who were referred to an MBCP after arrest. Thus, little is known about the responses that would be generated from men who are never arrested, those who are arrested but not referred, or those referred who do not show up for intake (Begun et al., 2009). The scale may be used over the course of an MBCP to assess changes in readiness to change; however, it needs to be used in conjunction with other validated measures, including victim/survivor reports of perceived changes in their partner's attitudes and behaviours.

University of Rhode Island Change Assessment Scale-Domestic Violence (URICA-DV)

The University of Rhode Island Change Assessment Scale-Domestic Violence (URICA-DV) (Levesque, Gelles, & Velicer, 2000) is a 20-item, self-report scale designed to specifically measure men's readiness to change their physical violence. The URICA-DV was modified from the original URICA (Cohen, Glaser, Calhoun, Bradshaw, & Petrocelli, 2005) for men who are abusive in their intimate relationships. The URICA-DV makes specific mention of violence across the four stages of change: pre-contemplation (e.g. "The violence in my relationship isn't a big deal"); contemplation (e.g. "I'm beginning to see that the violence in my relationship is a problem"); action (e.g. "I'm finally doing something to end my violent behaviour"); and maintenance (e.g. "Although I haven't been violent in a while, I know it's possible for me to be violent again"). There are five items per stage or subscale, each answered on a Likert scale ranging from 1-5, with higher scores indicating greater endorsement of particular attitudes or behaviours.

Psychometric properties

Internal consistency estimates range from 0.68-0.81 (Levesque et al., 2000) and 0.63-0.81 (Eckhardt & Utschig, 2007).

Critical review

The instrument can potentially provide important information about perpetrators' readiness to change prior to their engagement in an MBCP. However, the authors suggest more evidence is needed before using the URICA-DV to match perpetrators to treatment (Levesque et al., 2000). There are some added complexities in using the URICA-DV to admit men into, or exclude them from, an MBCP, particularly given that the assessment excludes self-reports of risk from victims/survivors. The URICA-DV also only focuses on physical violence and does not address psychological, or other, tactics of coercive control.

Perpetrator self-report measures

It is well documented in the literature (Day et al., 2019) that there is an absence of scales that have undergone rigorous psychometric testing for completion by perpetrators of violence. Scales that have been developed specifically for perpetrators often rely on self-reported data from men regarding their own violent behaviours. As a result, they are unlikely to be an adequate measure of change in perpetrator attitudes and behaviour across pre- and post-intervention, as perpetrators may be motivated to provide socially desirable responses or to under-report aggression and violent behaviours.

Measures of violent attitudes and behaviour

While recidivism of violent behaviour is no longer considered adequate as a singular measure of success of MBCPs (see the section Measurement of a range of key outcomes), long-term reductions in violent attitudes and behaviours are still key outcomes of MBCPs. However, outcome measures must also include assessment of controlling behaviour, verbal abuse, and threats (Dobash, Dobash, Cavanagh, & Lewis, 2000). Previous research suggests that effective measurement tools for assessing change in men's violent behaviour:

- are validated and psychometrically tested
- address both physical and psychological violence (or include a combination of measures across these domains)
- are completed as a self-report instrument by the victim/survivor (or include triangulation with a self-report measure from the victim/survivor)

- examine violent behaviour over a period of 6–12 months
- include specific rather than broad items
- measure both frequency and severity of violence. (Eckhardt et al., 2013; Gondolf, 2012; Koehler, Lösel, Akoensi, & Humphreys, 2013)

Attitudes toward women and violence

Ambivalent Sexism Inventory (ASI)

The Ambivalent Sexism Inventory (ASI) (Glick & Fiske, 1996) is a 22-item inventory grouped into two scales: hostile sexism (e.g. “Women are too easily offended”) and benevolent sexism (e.g. “Women should be cherished and protected by men”). Participants rate their agreement with each of the statements using a 7-point Likert scale ranging from 1 (totally disagree) to 7 (totally agree). The higher the score, the higher the endorsement of sexism. The ASI is a measure of attitudes only.

Psychometric properties

Internal consistency for the hostile and benevolent sexism subscales have been reported as $\alpha=.87$ and $\alpha=.84$, respectively (Glick & Fiske, 1996); and $\alpha=.93$ and $\alpha=.90$, respectively (Glick & Whitehead, 2010). The 12-item short form of this inventory (Glick & Whitehead, 2010) has also demonstrated good psychometric properties (internal consistency: $\alpha=.70$) (Rollero, Glick, & Tartaglia, 2014).

Critical review

The ASI has predominantly been used to measure community attitudes towards victims and perpetrators in large student samples (Masser, Viki, & Power, 2006; Riemer, Chaudoir, & Earnshaw, 2014); law enforcement samples (Gölge, Sanal, Yavuz, & Arslanoglu-Çetin, 2016; Lila, Gracia, & Garcia, 2013); and in violence prevention programs with young men (Stewart, 2014). One study used the ASI to examine sexism (psychosocial outcome) of offenders’ pre-/post-completion of an MBCP (Lila, Oliver, Catalá-Miñana, & Conchell, 2014). The ASI can only measure changes in beliefs/attitudes toward women and therefore might not be appropriate for some interventions with LGBTIQ+ peoples. The scale was developed through consultation with a predominantly white,

middle-class, student sample and has been criticised for cultural specificity (Hayes & Swim, 2013), as well as social desirability concerns (Good, Woodzicka, & Wingfield, 2010).

Attitudes toward Women Scale (AWS)

The Attitudes toward Women Scale (AWS) has a 55-item version (Spence & Helmreich, 1972), a 25-item version (Spence, Helmreich, & Stapp, 1973) and a 15-item version (Spence & Helmreich, 1978). The 15-item scale, which is highly correlated with the original version in both males and females, has been used most by investigators (Spence & Hahn, 1997). Both long and short versions include role (AS) items (e.g. “It is ridiculous for a woman to run a locomotive and for a man to darn socks”) and freedom (DS) items (e.g. “Women should be given equal opportunity with men for apprenticeship in the various trades”).

Psychometric properties

The reliability (internal consistency) of the scale has been reported to be .89 and correlates strongly ($r=.91$) with the original 55-item AWS (Spence & Helmreich, 1972). Cronbach’s alpha and test–retest reliabilities for the 15-item scale have been reported as .81 and .86, respectively (Daugherty & Dambrot, 1986). Daugherty and Dambrot (1986) also concluded that the 15-item scale possesses high test–retest reliability.

Critical review

This is a very dated scale. Since its development in the early 1970s, there have been significant changes in responses to the AWS to suggest an overall decline in the endorsement of “overt” sexist beliefs (Spence & Helmreich, 1978). Validity data suggest that age is negatively related (i.e. less liberal attitudes with greater age) and education is positively related to AWS scores (Dainbrot, Papp, & Whitmore, 1984). Given its use of traditional gendered language, the AWS would not be appropriate for use in MBCP evaluations today.

Inventory of

Beliefs about Intimate Partner Violence (IBIPV)

The Inventory of Beliefs about Intimate Partner Violence (IBIPV) (Saunders, Lynch, Grayson, & Linz, 1987) is a revision of the original scale, Inventory of Beliefs about Wife Beating

(IBWB). The new instrument consists of 22 items grouped into three subscales:

1. “Justifying partner violence” deals with general behaviours of victims and abusers that may be used to legitimise or justify IPV (6 items; e.g. “Sometimes abuse of the woman helps resolve conflicts between partners”).
2. “Victims responsible for violence” measures victim behaviours that may lead to women being blamed for interpersonal violence (9 items; e.g. “Battered women are responsible for battering, because they provoke it”).
3. “Abuser responsible for violence” deals with abusive behaviours such as intimidation and humiliation where the aim is to ensure the submission of the victim/survivor and achieve control over her (7 items; e.g. “Abusers are responsible for battering, because their intention is to intimidate and humiliate their partner”).

Psychometric properties

Saunders et al. (1987) reported Cronbach’s alpha coefficients between 0.67 and 0.89.

Critical review

This scale uses the terms “battering” and “battered woman” and this is not appropriate language for use in Australia. The scale is designed to measure attitudes/beliefs about abusers/perpetrators in community samples rather than from the perpetrator directly. The original and revised versions of the IBWB have failed to address empirical and theoretical developments in the field, particularly the significance of psychological, systematic and financial abuse in IPV and DFV (García-Ael, Recio, & Silvan-Ferrero, 2018).

Anger and aggression

Aggression Questionnaire (AGQ)

The Aggression Questionnaire (AGQ) (Buss & Perry, 1992) is a 29-item questionnaire that returns scores for four dimensions of aggression: physical aggression, verbal aggression, anger and hostility. Respondents rank statements (e.g. “If I have to resort to violence to protect my rights, I will”) on a 5-point Likert scale from “extremely uncharacteristic of me” to

“extremely characteristic of me”. The scores are normalised on a scale of 0 to 1, with 1 being the highest level of aggression.

Psychometric properties

Reported internal consistency coefficients are as follows: physical aggression, $\alpha=.85$; verbal aggression, $\alpha=.72$; anger, $\alpha=.83$; and hostility, $\alpha=.77$, with the overall internal consistency being reported as $\alpha=.89$. Test–retest reliability (9 weeks) for the subscales and total score ranged from $\alpha=.72$ to $\alpha=.80$ (Buss & Perry, 1992).

Critical review

The AGQ measures dimensions of generalised aggression rather than IPV or DFV specifically. Although relationships between anger/aggression and DFV have been reported where 1) anger and aggression increases the likelihood of DFV, and 2) anger and aggression are used as a justification for violent behaviour (see Birkley & Eckhardt, 2015), most MBCP theories of change do not target aggression. Therefore, many MBCP providers would not consider this measure to be contextually appropriate. However, the AGQ may be used in conjunction with other scales/inventories if an MBCP has targeted aggressive behaviour specifically and the provider is seeking a validated measure.

Proximal Antecedents to Violent Episode (PAVE)

The Proximal Antecedents to Violent Episode (PAVE) (Babcock, Costa, Green, & Eckhardt, 2004) is a 30-item scale designed to measure a man’s self-reported likelihood of perpetrating IPV (from 1 [not likely at all] to 6 [extremely likely]) in a range of situations. The PAVE includes three subscales: violence to control (e.g. “My partner threatens to leave me”); violence out of jealousy (e.g. “I walk in and catch my partner having sex with someone else”) and violence following verbal abuse (e.g. “My partner ridicules or makes fun of me”). The PAVE is largely used for safety planning interventions rather than for the purposes of measuring behavioural and attitudinal data.

Psychometric properties

Cronbach’s alpha coefficients for the whole scale, as well as for the violence to control and violence following verbal

abuse subscales, were above .90. Internal consistency for the jealousy subscale is lower ($\alpha=.70$) (Babcock et al., 2004).

Critical review

There is evidence of social desirability issues with the PAVE (Babcock et al., 2004). There are psychometric issues with the violence out of jealousy subscale, and the authors report limitations in terms of the validity of this subscale. The PAVE is most relevant for individuals who demonstrate higher levels of violence, rather than lower levels, and may not detect more insidious forms of potential violence in men (Babcock et al., 2004).

Emotional support

Interpersonal Competence Questionnaire (ICQ)

The Interpersonal Competence Questionnaire (ICQ) (Buhrmester, Furman, Wittenberg, & Reis, 1988) is a 40-item questionnaire designed to measure five domains of interpersonal competence:

1. initiation (8 items)
2. negative assertion (8 items)
3. disclosure (8 items)
4. emotional support (8 items)
5. conflict management (8 items).

The items are rated on a 5-point Likert-type scale ranging from 1 (“I am poor at this”) to 5 (“I’m extremely good at this”). Examples of items are the following: “Being able to say and do things to support a close companion when she/he is feeling down” (emotional support); “Being able to admit that you might be wrong when a disagreement with a close companion begins to build into a serious fight” (conflict management). For each scale, the higher the score, the higher the interpersonal competence.

Psychometric properties

In Buhrmester et al.’s (1988) study, internal consistency for the subscales ranged from $\alpha=.77$ (conflict management) to $\alpha=.87$ (emotional support). Four-week test–retest reliability

for the subscales ranged from $r=.69$ (conflict management) to $r=.89$ (initiation).

Critical review

This measure was chiefly developed for use with adolescents and has been used extensively in violence prevention programs with young people. The questions broadly examine competence in relationships but the measure is far removed from the program logic of most MBCPs because it doesn’t include any items that measure violent attitudes or behaviour.

Combined reports from perpetrators and victims/survivors

Research consistently indicates a disparity between reports of violence from victims and perpetrators, with victims likely to report higher levels of recidivism than male perpetrators (Dutton, 1988; Tjaden & Thoennes, 2000). While the below inventories have, in some ways, attempted to attend to this disparity of reports, there are added issues with collecting data from both the perpetrator and victims/survivors. For example, the absence of appropriate contextual information (e.g. responses of violence by victims in self-defence) may mean that users misconstrue the data. In addition, the absence of contextual information may render other tactics of violence (emotional, social and financial tactics, as well as sabotaging the mother’s parenting and access to support) invisible in combined reports (Vlais et al., 2017, p. 43).

Abuse of Partner Scales: Non-Physical Abuse of Partner Scale (NPAPS) and Physical Abuse of Partner Scale (PAPS)

This scale comes in two sets, one of which is completed by the perpetrator (Non-Physical Abuse of Partner Scale [NPAPS] and Physical Abuse of Partner Scale [PAPS]) (Garner & Hudson, 1992) and the other by the victim/survivor (see below).

The NPAPS is a 25-item scale designed to measure self-reported, non-physical abuse perpetrated against a partner (items include, for example, “I make fun of my partner’s ability to do things”). The PAPS is a 25-item scale designed to measure self-reported, physical abuse perpetrated against

a partner (e.g. “I have pushed and shoved my partner around violently”) (Garner & Hudson, 1992). Both scales rate the frequency of perpetrated non-physical and physical abusive acts on a 7-point Likert-type scale, ranging from 1 (never) to 7 (all of the time).

Partner Abuse Scale: Non-Physical (PASNP) and Partner Abuse Scale: Physical (PASPH)

The Partner Abuse Scales (Non-Physical [PASNP] and Physical [PASPH]) (Hudson, 1992) use the same items as the NPAPS and PAPS, but the questions are worded from a victim’s/survivor’s perspective.

The PASNP is a 25-item scale designed to measure non-physical abuse perpetrated against a partner, to be completed by the victim/survivor (e.g. “My partner makes fun of my ability to do things”). The PASPH is a 25-item scale designed to measure physical abuse perpetrated against a partner, to be completed by the victim/survivor (e.g. “My partner pushes and shoves me around violently”) (see Hudson, 1997).

Psychometric properties

Internal consistency: $>.90$. The PASPH and the PASNP produced internal consistency reliability coefficients of .90 to .95 respectively (Hudson, 1997). The PAPS and NPAPS have Cronbach’s alphas in excess of 0.90 (see Fischer & Corcoran, 2007). There is evidence of content and factorial validity (Hudson, 1997).

Critical review

Because the two forms of this scale are designed to assess accounts of physical and non-physical abuse from the perspective of the perpetrator (NPAPS; PAPS) (Garner & Hudson, 1992; Hudson, 1997) and the victim/survivor (PASNP; PASPH) (Hudson, 1992), they are cross-referenced with each other. The scales, in conjunction, have been used with couples participating in divorce mediation (see Beck, Walsh, & Weston, 2009). Given the known issues with perpetrator self-reports, the literature strongly suggests focusing on self-report measures completed by victims/survivors (PASNP; PASPH) (see Talley, Heitkemper, Chicz-Demet, & Sandman, 2006). The above scales have been used

to assess violence in particular “at-risk” populations (e.g. against pregnant women) rather than assess change observed by victims/survivors in the context of MBCPs. There may be some interest from researchers and evaluators to compare the reports from abuser and victims/survivors, but given the known limitations, the usefulness of this comparison is questionable in an MBCP evaluation context.

Conflict Tactic Scale (CTS) and Revised Conflict Tactics Scale-2 (CTS-2)

The Conflict Tactic Scale (CTS) (Straus, 1979) is a 15-item scale that measures violence (as “minor” or “severe”) in the context of conflict management within intimate relationships. The original CTS is comprised of three subscales that measure reasoning, verbal aggression, and violence between partners (Straus, 1979). The Revised CTS (CTS-2) (Straus, Hamby, Boney-McCoy, & Sugarman, 1996) is a 78-item scale that assesses both victimisation and perpetration over the previous 12 months. Respondents are asked to report on their own and their partners’ behaviours. Both scales are designed to measure tactics used by partners (dating, cohabiting, marital) to resolve conflict.

Psychometric properties

Internal consistency: physical $\alpha=.90$ (Mechanic, Uhlmansiek, Weaver, & Resick, 2000) and $\alpha=.94$ (Lucente, Fals-Stewart, Richards, & Goscha, 2001). There is evidence of convergent, discriminant, and factorial validity (Straus et al., 1996; Straus, Hamby, & Warren, 2003).

Critical review

The CTS is the most commonly used measure to assess the occurrence and frequency of DFV and exposure to DFV (Hamby & Finkelhor, 2001), but it has also been subject to considerable criticism. Both the CTS and CTS-2 are largely informed by a family conflict approach, which measures conflict between individuals; as such, it does not consider complexities of gender, power relations and social inequality. The CTS has also consistently found equal rates of violence for men and women, which is contrary to other measures, as well as reported criminal justice system data on DFV violence (Dobash et al., 1992). Critics argue it is an

ineffective measure of IPV because it a) positions violence as an “argumentative act” rather than a pattern of abusive behaviour; b) counts acts of violence but neglects the context of violence (including whether violence was used in self-defence); and c) asks frequency in the past 12 months and fails to account for ongoing systematic patterns of abuse (DeKeseredy & Schwartz, 1998; Kimmel, 2002). More broadly, the CTS-2 has been criticised for failing to address how gender and cultural context influence perceptions of violence, where men overwhelmingly under-report their use of violence and women minimise violence used against them. Given the substantial difference in data from the CTS-2 and from other data sources, critics argue the CTS should only be used in conjunction with other scales, as well as with qualitative measures (DeKeseredy, 2011).

Victim/survivor self-report measures

Given the disparity between perpetrator and victim/survivor reports of abuse, the victim/survivor is viewed as the best source of information regarding their partner’s continued abuse. One important consideration in measurement tools for victims/survivors is the comprehensiveness of the scale to adequately capture women’s interpretations of abusive situations. Comprehensiveness is imperative, as tools that use narrow definitions—for example, that only measure physical assault—ultimately exclude more insidious forms of abuse (e.g. name calling or threats) in the absence of physical assault (Bonomi et al., 2006, p. 121; Vlais & Green, 2018). In addition, evaluations of MBCPs should include measures that attend to the spectrum of men’s violence against women, and also include measures of women’s general wellbeing, psychological state, and/or quality of life (Stith, Smith, Penn, Ward, & Tritt, 2004). In this review of outcome measures, therefore, we also include measures of non-physical forms of abuse. As broader indicators of women’s perceptions of their safety and functioning, we also include measures of wellbeing and parenting.

Measures of violence and safety

Index of Spouse Abuse (ISA)

The Index of Spouse Abuse (ISA) (Hudson & McIntosh, 1981) is a 30-item scale designed to measure self-reported severity of

both physical and non-physical partner abuse of adult women in heterosexual relationships. The ISA measures the severity of physical violence (e.g. “My partner beats me so badly that I must seek medical help”); emotional abuse (e.g. “My partner screams and yells”); and controlling behaviours (e.g. “My partner acts like I am his personal servant”) experienced by women from their intimate partners. Each item is measured on a 5-point Likert scale. The ISA was initially evaluated with clinical samples of women living in protective shelters and college-aged or adult married women living with a male partner (Burgess & Tavakoli, 2005).

Psychometric properties

Internal consistency reported by Hudson and McIntosh (1981) ranged from $\alpha=.90$ to $\alpha=.94$ for the physical subscale and $\alpha=.91$ to $\alpha=.97$ for the non-physical subscale.

Critical review

The ISA has been used to assess abuse in community samples of women (Campbell & Soeken, 1999) as well as women who are victims of abuse (Cobb, Tedeschi, Calhoun, & Cann, 2006), incarcerated women (Eliason, Taylor, & Arndt, 2005), and pregnant women (McFarlane, Parker, & Soeken, 1995). The original validation study was conducted with a United States sample of students and victims of abuse. While used extensively internationally, the ISA has not shown a consistent factor structure across different cultural contexts (Aldarondo & Malhotra, 2014). The internal consistency reported above has not been replicated in later studies (see Torres et al., 2010). The scale has also been criticised for not measuring less severe forms of physical violence (such as pushing, grabbing, and shoving) which previous research suggests may occur most frequently (Aldarondo & Malhotra, 2014).

Composite Abuse Scale (CAS)

The Composite Abuse Scale (CAS) (Hegarty, Sheehan, & Schonfeld, 1999) is a 30-item scale with four subscales for females with current/former intimate partners with whom they have been in a relationship for 1 month. The four subscales include severe combined abuse (e.g. “I was locked in my bedroom”); emotional abuse (e.g. “Partner told me that I was crazy”); physical abuse (e.g. “Partner slapped me”); and harassment (e.g. “Partner followed me”). Developed in

Australia, the CAS is considered a preferred measure of IPV, and has been endorsed as a criterion standard for assessing women's self-reported experiences of violence (MacMillan et al., 2009; Wathen, Jamieson, MacMillan, & McMaster Violence Against Women Research Group, 2008). The CAS measures the frequency of experience in relation to 30 violent acts over a 12-month period on a six-point Likert scale ranging from "never" (0) to "daily" (5).

Psychometric properties

Internal consistency: physical abuse =.94. There is evidence of content, construct, criterion, and factorial validity (Hegarty, Bush, & Sheehan, 2005).

Critical review

The original CAS, and subsequent revisions, all focus specifically on women's experiences of violence. The CAS adopts a feminist lens, and positions men's and women's experiences, and patterns of violence, as dissimilar. This scale is highly regarded as it considers the spectrum of violent behaviours, including sexual violence, emotional and physical violence, and harassment, and also captures the severe combined category (see Kelly & Johnson, 2008; Laing & Humphreys, 2013). There is a validated short-form version of this scale. While the CAS is highly regarded, criticisms have been raised in regard to the wording and language used, the limited response options, and the use of cut-off scores for some items (Ford-Gilboe et al., 2016). Critics argue that this fails to address how women experience violence on a continuum (Ford-Gilboe et al., 2016).

Measure of Wife Abuse (MWA)

The Measure of Wife Abuse (MWA) (Rodenburg & Fantuzzo, 1993) is a 60-item scale with four subscales that measure the frequency (in the past 6 months) of physical (e.g. "Your partner punched you"), sexual (e.g. "Your partner tried to rape you"), psychological (e.g. "Your partner stole your possessions"), and verbal abuse behaviours (e.g. "Your partner told you that you weren't good enough"). The respondent then rates the extent to which the abusive action was hurtful or upsetting.

Psychometric properties

The MWA has recorded adequate levels of both internal consistency and concurrent validity with the CTS. Internal consistency: total scale $\alpha=.93$; physical abuse $\alpha=.81$. There is evidence of convergent and factorial validity (Rodenburg & Fantuzzo, 1993).

Critical review

The title uses the word "wife", but the scale was designed for all relationships, and gender-neutral language is used. The instrument aims to examine both frequency and severity of abuse. The MWA has been criticised for being too lengthy (60 items) and is often abbreviated. The explicit nature of some of the items (e.g. "Your partner called you a cunt") may be distressing for victims and inappropriate for use in some settings (Murray & Graves, 2013).

Severity of Violence Against Women Scale (SVAWS)

The Severity of Violence Against Women Scale (SVAWS) (Marshall, 1992) is a 49-item scale including nine factors that measure both frequency and severity of violent behaviours experienced by women in intimate relationships (e.g. "How often has your partner hit or kicked a wall, door or furniture?"). Items are rated on a Likert scale from "never" (1) to "many times" (4).

Psychometric properties

Internal consistency of SVAWS subscales ranges from $\alpha=.89$ to $\alpha=.96$ (Marshall, 1992).

Critical review

This scale is largely based on empirical research regarding severity and frequency of violence and may be useful for understanding women's perceptions about violence severity. However, this scale was developed with university students and community-based samples and not clinical populations.

Women's Experiences with Battering (WEB)

The Women's Experiences with Battering (WEB) (Smith, Earp, & DeVellis, 1995) is a 10-item scale designed to measure

women's experiences of violent/abusive relationships. The scale is specific to females with current or former male intimate partners. The WEB measures a woman's perceptions of her vulnerability to physical danger and loss of power and control in her current/previous relationship (e.g. "He has a look that goes straight through me and terrifies me"). The scale was developed through qualitative interviews with victims about their lived experience of violent relationships.

Psychometric properties

Research has shown that the WEB may be a more sensitive and comprehensive screening tool for identifying IPV compared to other validated tools that focus largely on physical assault. Internal consistency: $\alpha=.91$ to $.99$. There is reported evidence of convergent, discriminant, and criterion validity (Smith et al., 1995; Smith, Smith, & Earp, 1999; Smith, Thornton, Devellis, Earp, & Coker, 2002).

Critical review

This scale uses the term "battering" and would need to be adapted for use in an Australian context. The scale doesn't attempt to capture specific behaviours, and as a result may not provide an accurate overview of women's experiences of violence (Murray & Graves, 2013).

Measures of psychological violence and abuse

Historically, scales have tended to focus on measuring both physical and psychological forms of abusive behaviour. Physical abuse has also been identified as a catalyst for help-seeking and ending violent relationships (Pape & Arias, 2000; Strohine & Robinson, 2003). However, there is increasing evidence that violent tactics of control by the perpetrator may be purposively non-physical. Sackett and Saunders (1999) found that psychological abuse has a significant negative impact on victims' self-esteem. This has been supported in numerous studies (Henning & Klesges, 2003; Street & Arias, 2001) which suggest that victims of manipulation, humiliation and intense psychological abuse suffer comparable (or more severe) long-term psychological consequences to victims of physical abuse. Studies also suggest that psychological abuse is often a precursor to physical violence (Hannem, Langan,

& Stewart, 2015). For these reasons, measures which attend to the complexities of psychological abuse are increasingly being used in self-report scales for victims/survivors.

The following three scales were designed to detect and measure psychological abuse and should be triangulated with other broad measures of DFV.

Psychological Maltreatment of Women Inventory (PMWI)

The Psychological Maltreatment of Women Inventory (PMWI) (Tolman, 2001) is a 58-item scale divided into two subscales: dominance–isolation and verbal–emotional abuse. The short version of the PMWI contains 14 items and includes items from each of the subscales. Women are asked to rate how frequently each of the 14 items occurred over the past 6 months (e.g. "My partner called me names" and "My partner told me my feelings were irrational or crazy"). The scale was developed with high endorsement from women who had previously experienced domestic violence and has been used extensively with victims of domestic violence.

Psychometric properties

The short versions of each subscale have successfully discriminated between abused and non-abused women, and its internal consistency has been reported as $\alpha=.87$ (Kasian & Painter, 1992). Tolman (2001) has reported coefficient alphas for the subscales as $.88$ (dominance and isolation) and $.92$ (emotional and verbal). The PMWI scores were also highly and significantly correlated with the non-physical abuse subscale of the ISA, providing further evidence for the validity of the PMWI subscales (Tolman, 2001).

Critical review

The PMWI has been endorsed as "the most comprehensive questionnaire to measure psychological abuse" (Murphy & Cascardi, 1993). One criticism of the PMWI is that it may not accurately reflect the four different forms of abuse (dominance, isolation, emotional and verbal) as intended (Murphy & Cascardi, 1993).

Profile of Psychological Abuse of Women (PPAW)

The original Profile of Psychological Abuse of Women (PPAW) scale (Sackett & Saunders, 1999) includes 42 items drawn from clinical work with perpetrators (i.e. descriptions of tactics) and the experiences of victims/survivors (Pence, Paymar, Ritmeester, & Shepard, 1993). The items ask how often a partner engages in particular behaviours (e.g. How often does your partner “put you down if you cry or ask for emotional support?”) on a 7-point Likert scale ranging from “never” to “daily”. The items address a wide range of psychological abuse: humiliation, threats, invalidation of experiences, isolation, trivial demands, occasional indulgences, and emotional distance. The PPAW scale was revised to include 21 items, across four major forms of abuse: criticise behaviour, ignore, ridicule traits and jealous/control (Sackett & Saunders, 1999).

Psychometric properties

The internal reliability scores of the revised scale for the major forms of abuse are as follows: criticise behaviour ($\alpha=0.75$), ignore ($\alpha=0.80$), ridicule traits ($\alpha=0.79$), and jealous/control ($\alpha=0.85$) (Sackett & Saunders, 1999).

Critical review

This scale exclusively measures psychological abuse. It has an advantage of using specific time referents (e.g. “once a month”, “once a week”, “2–3 times a week”). There is some concern that the sample used in the Sackett and Saunders (1999) validation study did not account for the diversity of potential responders.

Measures of victim/survivor wellbeing

Research suggests that victims of psychological, physical and sexual IPV are more likely to suffer from depression and anxiety, and use healthcare services more often, compared to women who do not have IPV experiences (Stylianou, 2018). Meta-analyses of mental health studies in this area also reveal a strong relationship between IPV and major depressive disorder, depressive symptoms, postpartum depression and suicide attempts (Beydoun, Beydoun, Kaufman, Lo, & Zonderman, 2012; Devries et al., 2013), with between 9

percent and 28 percent of major depressive disorder and postpartum depression diagnoses being attributed to IPV experiences (Dillon, Hussain, Loxton, & Rahman, 2013). Following research regarding the mental health difficulties faced by victims of DFV, there is increasing pressure for evaluation of MBCPs to include measures of women’s mental health and wellbeing. This shift toward evaluating women’s mental health ensures that interventions and intervention outcomes are focused on improvements in the wellbeing of victims, rather than a pure focus on reduction in perpetrators’ violent and controlling behaviours.

Warwick-Edinburgh Mental Well-being Scale (WEMWBS) and Short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS)

The Warwick-Edinburgh Mental Well-being Scale (WEMWBS) (Tennant et al., 2007) is a 14-item scale with five response categories. The items are all positively worded (e.g. “I feel positive about the future” and “I’ve been feeling good about myself”). “Mental wellbeing” encapsulates affective-emotional, cognitive-emotional and psychological functioning. Respondents are required to describe their experience of each statement over the past 2 weeks using a 5-point Likert scale. A higher WEMWBS score indicates a higher level of mental wellbeing (Tennant et al., 2007). The Short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS) consists of seven items.

Psychometric properties

The reliability in the original study (general population) yielded a Cronbach’s alpha of .91 (Tennant et al., 2007). The WEMWBS has been validated in various populations, including ethnic minority samples (Trousselard et al., 2016).

Critical review

Critiques of the WEMWBS have focused on issues regarding the nonspecific nature and understanding of the items. This includes the potential for items to be misinterpreted (Gremigni & Stewart-Brown, 2011; Tennant et al., 2007). Small-scale community studies have used the WEMWBS to measure victim/survivor mental wellbeing. Qualitative studies with community samples (Tennant et al., 2007) revealed concerns

about whether responses can adequately capture self-reflection and communicate high and low mental health scores. The WEMWB is a brief measure that takes less than 3 minutes to complete. It may be an effective scale to use across pre- and post- timeframes of MBCPs to measure changes in victim/survivor wellbeing.

Centre for Epidemiologic Studies Depression Scale (CES-D)

The Centre for Epidemiologic Studies Depression Scale (CES-D) (Radloff, 1977) is a 20-item self-report measure that assesses the frequency of depression symptomatology in the general population. Items were devised from previous validated measures (e.g. the Beck Depression Inventory [BDI] and the Mental Muscle Diagram Indicator [MMDI]). However, the items chiefly focus on measures of depressed mood in terms of dysphoria, wellbeing, somatic complaints and interpersonal difficulties.

Psychometric properties

Radloff (1977) recorded a coefficient alpha of .90 for the general population. Test-retest reliability over 6 months was $r=.54$. Similar findings have been reported across different cultural groups (Losada et al., 2012; Nguyen, Kitner-Triolo, Evans, & Zonderman, 2004; Ros et al., 2011; Zhang et al., 2011).

Critical review

The CES-D has been widely used in research involving victims of IPV and abuse. The CES-D has good psychometric properties for use with these populations (La Flair, Bradshaw, Mendelson, & Campbell, 2015).

Kessler Psychological Distress Scale (K6)

The Kessler Psychological Distress Scale (K)6 (Kessler et al., 2003) is a simple measure of psychological distress comprising six questions that ask about the following feelings during the past month: sad, nervous, restless or fidgety, hopeless, everything is an effort, worthless. Responses range from 0 (none of the time) to 4 (all of the time). It measures distress over a period of 4 weeks prior to administration of the test.

Psychometric properties

Cronbach's alpha for the full sample for this scale has been reported as .83 (Kessler et al., 2002).

Critical review

The K6 did not perform as well as the Kessler Psychological Distress Scale (K10) for screening of severe disorders. However, the K6 is preferred overall for assessing DSM-IV mood and anxiety disorders because it is brief and shows consistency across subsamples (Furukawa, Kessler, Slade, & Andrews, 2003).

General Self-efficacy Scale (GSE)

The General Self-efficacy Scale (GSE) (Schwarzer & Jerusalem, 1995) is a 10-item scale originally developed in 1979 which has been translated from the original German into 26 languages, including English. The GSE was designed to measure a general sense of perceived self-efficacy, including coping with daily hassles and adaptation after stressful life events (e.g. "I have a special person who is a real source of comfort to me" and "My friends really try to help me"). Each of the ten items is measured on a 4-point scale (range 10–40).

Psychometric properties

In samples from 23 nations, Cronbach's alphas ranged from .76 to .90, with the majority in the higher range of $\geq .80$ (Schwarzer & Jerusalem, 1995).

Critical review:

The GSE has been used extensively with victims/survivors of violence, as well as across different cultural groups (see Johansen, Wahl, Eilertsen, & Weisaeth, 2007). The GSE takes approximately 2 minutes to complete and may be an effective scale to use across pre- and post- timeframes of MBCPs to measure changes in victims'/survivors' self-efficacy.

Multidimensional Scale of Perceived Social Support (MSPSS)

The 12-item Multidimensional Scale of Perceived Social Support (MSPSS) (Zimet, Dahlem, Zimet, & Farley, 1988)

examines social support in three areas: family, friends and significant personal relationships (e.g. “There is a special person with whom I can share my joys and sorrows”).

Psychometric properties

The MSPSS has recorded Cronbach’s alphas between 0.93–0.98 for the measure’s total score, and between 0.81–0.91 for the subscales (see Hardan-Khalil & Mayo, 2015).

Critical review

The MSPSS has been used extensively with victims/survivors of violence, and has been adapted for different population groups (Barnett et al., 1996). The MSPSS has been shown to have good internal consistency and test–retest reliability, good validity, and a stable factorial structure across numerous studies (Zimet, Powell, Farley, Werkman, & Berkoff, 1990).

Rosenberg Self-esteem Scale

This 10-item scale (Rosenberg, 1965) is the most popular measure of global self-esteem. Respondents are asked to rate on 4-point Likert scale the degree to which each of ten statements aligns with their views (e.g. “I feel like I have a number of good qualities”).

Psychometric properties

Cronbach’s alphas of .88 and test–retest reliability of .82 have been reported (Fleming & Courtney, 1984).

Critical review

The Rosenberg Self-esteem Scale has been used extensively with victims/survivors of violence and abuse. The scale is very short (10 items) and may be an effective measure of changes in self-esteem for women across pre- and post- timeframes of MBCPs (Cascardi & O’Leary, 1992).

Parenting

Project Mirabal (Kelly & Westmarland, 2015) interviewed women about their expected outcomes of MBCPs. The findings suggest that parenting, particularly having a safe space to

engage in respectful co-parenting, is a key area that victims/survivors want to improve. Overall, women identified six desirable outcomes of MBCPs, the primary one of which was “more space for action to regain parenting capacity”, as well as three other outcomes that focused on improvements to men’s fathering.

A number of MBCPs have been developed for fathers who use violence (McConnell, Barnard, Holdsworth, & Taylor, 2014), with the aim of communicating to fathers the possible impacts of DFV on child development. The evidence base in this area is evolving rather than definitive, with some promising outcomes emerging (McConnell et al., 2014).

At present, there are few validated parenting scales that are able to target changes in parenting, and specifically changes in levels of co-parenting and support. The main limitation of scales in this area is that they tend to examine “types” of parenting and parental–child behaviours (e.g. dominant, authoritarian etc.), rather than co-parenting, co-parenting conflict, or capacity for safe parenting.

Parental Locus of Control Scale (PLOC) (Parental efficacy and parental control of a child’s behaviour factors only)

The Parental Locus of Control (PLOC) (Campis, Lyman, & Prentice-Dunn, 1986) is a 14-item scale developed to assess parenting locus of control; that is, beliefs about parents’ ability to control their children’s behaviour and development (e.g. “When something goes wrong between me and my child, there is little I can do to correct it”).

Psychometric properties

An internal parental locus of control has been associated with higher maternal self-esteem and parental satisfaction after family structure, background variables, and social support were controlled for. This finding provides support for the criterion validity of the scale. Internal reliability of the 14-item PLOC scale has been reported as $\alpha=.84$ (Koeske & Koeske, 1992).

Critical review

The PLOC has been criticised for not adequately addressing abusive behaviour, and for positioning “authoritarian techniques”, such as “punishment”, “threatening” and “verbal force”, as appropriate parenting behaviours. This scale is also not consistent with MBCP theories of change. Given the focus of this scale on parents’ beliefs about controlling their child’s behaviour, it is conceptually in opposition to the ultimate goals of MBCP, including a victim’s/survivor’s freedom and space for action to regain her parenting capacity.

Parenting Scale

The Parenting Scale (Arnold, O’Leary, Wolff, & Acker, 1993) is a 30-item, self-report scale that assesses parental discipline strategies in response to child behaviours. Parents’ discipline strategies are rated on 7-point Likert-type scales (including reverse items), where 1 indicates a high probability of using an “effective discipline strategy” and 7 indicates a high probability of “ineffective discipline”. Based on the original factor analysis (Arnold et al., 1993), parenting style can be divided into three separate response styles: “overreactivity” (harsh, angry discipline style, consistent with an authoritarian parenting style); “laxness” (permissive style of parenting); and “verbosity” (parents rely on verbal persuasion even when ineffective).

Psychometric properties

Arnold et al. (1993) reported internal consistency at $\alpha=.84$, with laxness and overreactivity at $\alpha=.83$ and $.82$, respectively. Over a 2-week period, test–retest reliability was relatively high for the total, laxness, and overreactivity scores, at $r=.84$, $.83$, and $.82$, respectively (Arnold et al., 1993). Scores have been significantly related to clinical observations of parent–child interactions (Arnold et al., 1993).

Critical review

The Parenting Scale was originally developed for parents of younger children (18–36 months), and subsequent validation studies with older children (2- to 12-year-old children) were found not to support the verbosity factor (Collett, Gimpel, Greenson, & Gunderson, 2001). The language used in this scale—particularly the terms “parenting dysfunction”

and “ineffective discipline”—may be viewed by parents as judgmental.

Parenting Stress Index–Short Form (PSI-SF)

The Parenting Stress Index–Short Form (PSI-SF) (Abidin, 1990) is a 36-item measure of parenting stress that includes three subscales: 1) parental distress; 2) parent–child dysfunctional interaction; and 3) difficult child. The parental distress subscale aims to measure a parent’s perception of competence in child-rearing, including conflict with his or her partner, social support, and other stressors. The parent–child dysfunctional interaction subscale measures interactions between the parent and child. The difficult child subscale looks at the parent’s views of the child’s behaviour and temperament. Each subscale consists of 12 items rated from 1 (strongly disagree) to 5 (strongly agree). Higher scores reflect greater levels of stress.

Psychometrics

Abidin (1990) reported Cronbach’s alpha as follows: total stress =.91; parental distress =.87; parent–child dysfunctional interaction =.80; difficult child =.85. Test–retest data for the entire normative sample of 800 parents showed: total stress, $r=.84$; parental distress, $r=.85$; parent–child dysfunctional interaction, $r=.68$; difficult child, $r=.78$.

Critical review

This measure may be useful for measuring changes in parental stress (for both the perpetrator and their [ex-]partner) over the course of an MBCP, as well as co-parenting stress and communication barriers between the perpetrator and their (ex-)partner. This questionnaire may be particularly relevant to MBCPs that target fathering skills. The PSI-SF has been used with mothers who are victims of DFV. The PSI-SF may be vulnerable to self-report limitations, particularly in clinical samples where mothers are at risk of depression (Lee, Hayat, Spratling, Sevcik, & Clark, 2018).

Parental Acceptance–Rejection Questionnaire (PARQ)

The Parental Acceptance–Rejection Questionnaire (PARQ) (Rohner, 1990; Rohner & Ali, 2016; Rohner & Khaleque, 2005)

is a 60-item, self-report questionnaire designed to measure adults' retrospective accounts of parental (maternal and paternal) acceptance or rejection in childhood, or children's current experiences. The measure consists of four scales: 1) warmth and affection; 2) hostility and aggression; 3) indifference and neglect; and 4) undifferentiated rejection. There are four versions: PARQ–Mother, PARQ–Father, Early childhood PARQ, and Childhood PARQ. The instruments are nearly identical, except for minor changes to accommodate varying pronouns and tense (“My mother/father yells/yelled at me when she/he is/was angry”).

Psychometric properties

The overall internal consistency of the PARQ across three versions (Mother, Father and Childhood versions) has been reported as $\alpha=.89$ (Khaleque & Rohner, 2002).

Critical review

The PARQ has been used with perpetrator populations. However, the outcomes of the PARQ—which measure retrospective accounts of parental acceptance and rejection in childhood—do not align with the program logic and overall goals of MBCPs in Australia. The Childhood PARQ may be useful, but only in situations where deriving this data from children would be safe and ethical.

Measures of children's safety and wellbeing and family functioning

Research strongly suggests that children and adolescents who experience DFV are at greater risk of developing mental health problems, including anxiety, depression and developmental problems (Cicchetti & Doyle, 2016; Holt, Buckley, & Whelan, 2008). Outcome measures involving children in family violence situations have tended to focus on psychological wellbeing and narrow behavioural outcomes. Such outcomes often include children's symptoms and diagnoses (e.g. trauma symptoms and externalising problems) and, cognisant of some perpetrators' deliberate tactics to harm their co-parent's parenting capacity, evaluations of parenting in the non-abusing parent (e.g. skills and efficacy). Importantly, however, there is some concern that reports of a mother's parenting often ignore the importance of context, as well as

specific tactics the perpetrator may be using to deliberately sabotage a women's parenting capacity (Vlais, 2014; Vlais et al., 2017).

In addition, measures of “good outcomes” in interventions for children need to extend beyond psychological wellbeing and include measures of resilience, empowerment, and qualitative measures to gauge improvements in interpersonal relationships (Howarth et al., 2016). For this reason, in our review of outcome measures within this domain, we have included broad measures of children's wellbeing that encompass constructs such as life achievements, peer relationships and health, in addition to traditional measures of psychological distress.

Strengths and Difficulties Questionnaire (SDQ)

The Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997) is a brief emotional and behavioural screening questionnaire for children and young people. The tool can capture the perspectives of children and young people, their parents and teachers. There are currently three versions of the SDQ: a short form, a longer form with an impact supplement (which assesses the impact of difficulties on the child's life) and a follow-up form. The 25 items in the SDQ comprise five subscales of five items each: 1) emotional symptoms; 2) conduct problems; 3) hyperactivity/inattention; 4) peer relationship problems; and 5) prosocial behaviour (Goodman, 1997).

Psychometric properties

Cronbach's alpha for the parents' and teachers' version has a weighted average between 0.79–0.82. However, Cronbach's alpha was weaker for some of the subscales (weighted average parents' and teachers' version range 0.49–0.69 and 0.69–0.83) (Kersten et al., 2015).

Critical review

The SDQ has been used to assess mental health and wellbeing in children who have experienced DFV (Tomin & Cummins, 2011). The scale has been criticised as it only demonstrated clinical utility when there was agreement between teacher and parent reports.

Personal Wellbeing Index–School Children (PWI–SC)

The Personal Wellbeing Index–School Children (PWI–SC) (Cummins & Lau, 2005) is a 7-item instrument for measuring mental health and personal wellbeing in school-aged children and adolescents. The index assesses satisfaction with the following life domains: standard of living, health, life achievements, personal relationships, personal safety, community connectedness, and future security. These seven domains are theoretically embedded with reference to the global question, “How satisfied are you with your life as a whole?”

Psychometric properties

The PWI–SC has a reported internal consistency of $\alpha=.82$ (Tomyn & Cummins, 2011).

Critical review

The PWI–SC has been used in groups of vulnerable children, such as children in residential care. There is some criticism of gender differences in responses, with boys reporting greater satisfaction than girls (González-Carrasco, Casas, Malo, Viñas, & Dinisman, 2017). Subjective wellbeing has also been shown to decline with increasing age (Tomyn & Cummins, 2011).

Children’s Inventory of Anger (ChIA)

The Children’s Inventory of Anger (ChIA) (Nelson & Finch, 2000) is a 39-item, child-informed, self-report assessment that explores various elements of anger, including intensity, relationship focus and expression. Children respond to items outlining various home and school situations on a 4-point Likert scale, which reflects the level of anger they feel. In addition to a total anger score, ChIA yields subscales for frustration, physical aggression, peer relationships and authority relations.

Psychometric properties

The ChIA has demonstrated strong internal consistency (.95), adequate test–retest reliability (.66) and split–half reliability (.83 to .95). This measure has also established content and

discriminant validity via psychometric testing in clinical populations (Nelson & Finch, 2000).

Critical review

The ChIA includes pictorial prompts and simple language, which is useful for assessments with younger children and children who have reading difficulties or who exhibit behavioural problems. The ChIA has been used extensively with vulnerable groups, is user-friendly, and is applicable in both school and clinical settings. However, the scale is lengthy and the association between anger and other clinically significant disorders remains poorly understood (Flanagan & Allen, 2005; Nelson & Finch, 2000).

Conclusion

In completing this state of knowledge review, we had two broad aims. The first aim was to provide an updated summary of Australian standards of practice relating to MBCPs for DFV. The second aim was to identify validated outcome measures in three key outcome domains: 1) perpetrators’ violence and controlling behaviour; 2) adult victims/survivors’ and children’s safety, wellbeing and functioning; and 3) provision of a brief critical review of each measure. The outcomes of the two parts of this state of knowledge review were in turn used to inform the structure and content of the Evaluation guide.

Our review of the standards identified that since the Day et al. (2019) report, updates of standards have been undertaken in Queensland, Victoria and New South Wales. Commonalities emerging across state-based and national standards highlight as central aims of MBCPs the increased freedom and safety of women and children who are the victims/survivors of DFV. Consequently, there is also emerging consensus that in addition to measuring long-term reductions in violent behaviour, evaluations measuring the success of MBCPs must also measure outcomes related to the safety and wellbeing of victims/survivors, including children.

In our review of validated measures, we were able to identify a number of measures with established psychometric properties assessing various aspects of violence risk, victim/survivor

safety and victim/survivor wellbeing for both women and children that may be useful for evaluation of outcomes of Australian MBCPs. However, given the number of measures reviewed, there was a dearth of validated, comprehensive measures that could be used to determine effectiveness of MBCPs. We excluded measures that have not undergone psychometric testing, and we are aware that there are some comprehensive measures of MBCP outcomes being used that have not been validated. The results of this review highlight the need for adequate support to be provided to allow such comprehensive measures to be validated. Validation of such measures has the potential to improve quality and consistency of MBCP evaluation by providing tools for better evaluation to take place (see Appendix B).

The findings of this review were used to inform the further development of the Evaluation guide. The review of the standards highlights areas of consensus regarding the essential outcomes of MBCPs needing to be assessed in any MBCP evaluation. The review of validated measures enabled us to provide examples of established measures that can be used to assess these outcomes.

Methodology

The State of knowledge review informed development of the Evaluation guide. In addition, there were two main activities undertaken to develop the Evaluation guide:

1. interviews with evaluators of Australian MBCPs and a group consultation with the specialist No to Violence members' forum
2. collating and summarising key evaluation information to include in the Evaluation guide.

Real-world examples

Interviews with evaluators

There is already significant work taking place across Australia in the implementation and evaluation of MBCPs. To supplement the evaluation concepts outlined in the Evaluation guide, we included excerpts from interviews with evaluators involved in current or recent evaluations of MBCPs across Australia. We aimed to represent a range of examples of programs from various Australian states and territories, as well as programs aimed at a range of target groups. We identified several programs to include in the Evaluation guide and created a “case study matrix” to identify real-world examples from programs representing all states and territories, those based on “standard” and “innovative” practice, and programs designed for specific groups of participants (e.g. fathers, Aboriginal and Torres Strait Islander men, people from refugee backgrounds, new migrants, and people from LGBTIQ+ communities). We also aimed to include a range of evaluation types and to note whether the particular funding body had engaged an independent evaluator or whether the evaluation was conducted “in house”. We identified possible case studies through a desktop review of existing programs and through the industry knowledge of our research team and expert review panel.

Where an independent evaluator had been engaged by the funding body (often state government departments/units) to evaluate one of our chosen MBCPs, we sought permission from these bodies to contact the evaluators. A pro-forma letter was sent to the appropriate contact by the chief investigator. Once we received permission to contact evaluators, we contacted them directly. Where the evaluation was being

conducted internally, we made initial contact with the person managing the project. In our contact with these evaluators, we requested any available evaluation documentation and to conduct a 1-hour interview to provide further information not available through the documentation.

Broadly speaking, the information we gained from each of the real-world examples, for potential use in the Evaluation guide, related to the following topics:

- use of intervention theories of change, program logic models and evaluation frameworks to guide the evaluation
- how criteria for “success” were determined
- choice of tools or indicators to measure outcomes (“success”) of the program
- which methods of data collection were chosen and why
- barriers and enablers to carrying out the evaluation
- relationships with stakeholders and how stakeholders were involved in the evaluation
- ethical considerations in designing the evaluation and ethical issues arising during the evaluation.

A data collection guide for the interview is included as Appendix A. This data collection guide provides the specific questions we aimed to answer for each interview, though not all of the listed information was relevant or available for every interview.

Group consultation with the specialist No to Violence members forum

We took advantage of an opportunity offered via our expert review panel member from No to Violence to attend a specialist No to Violence members' forum. Thirteen members attended this forum, and were largely MBCP facilitators and program managers. We provided a general overview of the Evaluation guide and asked for feedback. Members also discussed a range of issues related to conducting evaluation of their programs. We gained written consent to record this session, and we include excerpts from this consultation in the guide to illustrate relevant issues.

Consolidating information for the Evaluation guide

The Evaluation guide is divided into the following main sections: introduction, scoping the evaluation, and ethical issues associated with conducting an evaluation of an MBCP. The section entitled “Scoping the evaluation” is further divided into six sub-sections:

1. understanding the program’s theoretical framework
2. articulating the program goal
3. developing a program logic
4. developing and prioritising evaluation questions
5. answering evaluation questions
6. deciding who should conduct the evaluation.

To develop the content for the Evaluation guide, we drew on information from a range of sources. These included seminal texts on health and social program evaluation, academic papers on evaluation of MBCPs, general program evaluation information that is available in the public domain from reputable sources (e.g. the Centers for Disease Control and Prevention’s Program Performance and Evaluation Office webpages), published MBCP evaluation documentation, and the considerable evaluation expertise of members of the expert review panel and project team. This was punctuated by information from the interviews (as described in the previous section).

To ensure the content of the Evaluation guide is useful and appropriate for use by people working in MBCPs, we also conducted a short workshop with members of the expert review panel to provide input into what should be included in the guide.

Expert review panel

An expert review panel was formed for the duration of the project to ensure the practical guide was reviewed and qualitative input was provided by a variety of stakeholders. The purposes of the expert review panel were twofold:

- to ensure that the guide is comprehensive in incorporating

evaluation principles and that examples are inclusive of a range of viewpoints

- to ensure the usefulness of the guide for those who are likely to use it.

Our expert review panel comprised six representatives from organisations with specific expertise in DFV, MBCPs and/or evaluation. The representatives and their organisations are as follows:

- Dr Georgina Sutherland, Respect Victoria, Centre for Health Equity and Centre for Mental Health, The University of Melbourne
- Dr Helen Jordan, Centre for Health Policy, University of Melbourne
- Dr Cathy Vaughan, Centre for Health Equity, University of Melbourne
- Professor Emerita Patricia Easteal, University of Canberra, Legal Lightbulbs
- Mr Jacob Pegg, replaced by Ms Ilana Jaffe, Family Safety Victoria
- Mr Michael Brandenburg, No to Violence.

The panel met in person three times over the life of the project to provide input into the development of the Evaluation guide and feedback on the three major deliverables of the project. They also provided other advice and feedback as requested.

Ethical considerations

Ethical approval for this project was gained from the Melbourne School of Population and Global Health (MSPGH) Human Ethics Advisory Group (HEAG) at the University of Melbourne as a minimal risk project (Ethics ID: 1853504.1). This is considered a minimal risk project as data-gathering methods were largely desktop reviews, and interviews were undertaken with personnel involved in evaluating or implementing MBCPs, rather than with perpetrators of DFV or victims/survivors. We also made an amendment to the

original ethics application to conduct and record the group consultation with the specialist No to Violence members' forum, for which we also gained approval before we proceeded.

Project limitations

To provide a feasible scope for completion of the work outlined in this report, we conducted the research within limited and specified parameters. This makes it possible that relevant information falling outside of this scope was excluded from the research. The review of standards was conducted within a specified timeframe (to June 2019) and it is possible that there have been changes made to these standards since this time. Given the rapidly changing nature of the sector, this is unavoidable. Similarly, the review of outcomes measures examined literature from January 2008 to May 2019, and only included validated measures. Therefore, there may be relevant measures that were not identified through this search strategy. The Evaluation guide was developed with input from a range of experts, including those working in DFV and complex program evaluation. However, we recognise that the evaluation of MBCPs is also developing, and it is likely that there may be alternative approaches that could have been outlined in the guide. However, such parameters and direction are necessary to make such work feasible.

THE EVALUATION GUIDE:

A guide for evaluating behaviour change programs for men who use domestic and family violence

Introduction

Purpose of this Evaluation guide

The purpose of this Evaluation guide is to provide easy-to-understand information on scoping an evaluation for men's behaviour change programs (MBCPs) focusing on men who use domestic and family violence (DFV). This Evaluation guide aims to improve the technical knowledge for personnel involved in the implementation of MBCPs who might be involved in commissioning an external evaluation or conducting an in-house evaluation. Upon working through this Evaluation guide, readers should:

- understand the purposes of conducting quality evaluations of MBCPs
- be able to develop a program logic and articulate appropriate evaluation questions
- be aware of a range of methodologies available to answer the evaluation questions
- gain an awareness of the important considerations in designing an evaluation of MBCPs
- be aware of the ethical issues that need to be considered when commissioning or conducting an evaluation of an MBCP.

The information contained in the guide will contain excerpts from interviews (presented in some of the boxes) with evaluators of MBCPs across Australia and a group consultation conducted in Melbourne, Victoria, involving program managers and facilitators of MBCPs.

Australian context

There are numerous definitions of MBCPs in use in Australia. For the purpose of this guide, we have limited our discussion to group work programs involving men who use violence and controlling behaviour against women. However, the evaluation principles outlined in this guide may be applicable to other perpetrator interventions, and indeed all DFV behaviour change programs. We also recognise that not all victims/survivors of DFV are women, nor all perpetrators men; however, for the purposes of this study, we use the term DFV to refer to that perpetrated against women by men.

This guide has also been informed by a state of knowledge review that assessed the current state of practice standards for MBCPs in place in Australia. This review also

examined the academic and grey literature on outcome measures used in MBCPs in the domains of:

1. long-term changes in perpetrators' violent and controlling behaviour
 2. adult victim/survivor safety, wellbeing and freedom
 3. children's safety, wellbeing and family functioning.
- (Adapted from Respect, 2017)

We also recognise that MBCPs are part of a complex environment and service system, and that an MBCP can only be as effective as the broader system within which it operates. The broader system elements may influence the success of the program and should be identified and considered in the design of any MBCP and its evaluation.

Structure of this guide

Following this introduction, the guide is divided into sections on scoping the evaluation and ethical issues associated with conducting an evaluation of an MBCP. The Scoping the evaluation section is further divided into six main sub-sections:

1. understanding the program's theoretical framework
2. articulating the program goal
3. developing a program logic
4. developing and prioritising evaluation questions
5. answering evaluation questions
6. deciding who should conduct the evaluation.



What is evaluation?

Formal program evaluation is “the systematic collection of information about the activities, characteristics, and outcomes of programs to make judgments about the program, improve program effectiveness, and/or inform decisions about future programming” (Patton, 1997, p. 23).

Why evaluate?

Evaluation of MBCPs for perpetrators of DFV allows us to ask important questions that can help identify points in the implementation of the program where improvements can be made, as well as identify if the program is achieving its aims. Based on these evaluation findings, important decisions are often made, such as whether to continue, expand, modify or discontinue the program. In the context of an innovative or new type of program, evaluation is essential to help funders decide whether the innovation should be scaled up.

The other important purpose of evaluation is to contribute to the evidence base on MBCPs; that is, to contribute to the general knowledge in the sector such that others can use the knowledge gained to inform their own program. This is especially important in the area of MBCPs, where the evidence base is limited. There is always

a need to balance the responsibility of contributing to the evidence base with accountability for understanding the impact of an MBCP program.

Before you begin: Evaluation readiness

This guide takes you through the key steps of scoping an evaluation before you commence your own evaluation activities or commission someone to do this on your behalf. In this way, you are making sure that your MBCP is “evaluation-ready”. Determining whether your MBCP is evaluation-ready helps you to decide whether, at that certain point in time, an evaluation is “justified, feasible, and likely to provide useful information” (Kaufman-Levy & Poulin, 2003, p. 4). Evaluation readiness is also often called “evaluability” (Hawe, Degeling, & Hall, 1990). Completing an evaluation before it is evaluation-ready means that the outcomes measured may not accurately reflect the true effects of the program, but rather that the MBCP is not yet being fully implemented.

Day et al. (2019), in their recent Australia’s National Research Organisation for Women’s Safety (ANROWS) research report on improving the quality of MBCPs, argued that one of the major factors hampering the development of a knowledge base in this field is the lack of evaluation readiness of programs being formally evaluated. The authors pointed, in particular, to the lack of guiding program logics, lack of clarity of theoretical frameworks underpinning programs, and program integrity in relation to evaluated programs. Program integrity refers to the degree to which the program is being implemented as intended (also sometimes called “fidelity”).

Completing the steps in this Evaluation guide in order to scope the evaluation will assist you in deciding whether the program is evaluation-ready. However, there are several other questions you can consider when attempting to gauge whether a program has sufficient program integrity to be evaluation-ready. These include the following:

- Are program staff, including group work facilitators, other practitioners and partner contact workers, sufficiently trained in MBCP work in general, as well as in the particular theoretical orientation or approach adopted by the program?
- Are internal and external supervisors aware of, and able to talk about, the theoretical approaches and behaviour change models adopted by the program, and can they identify when practice is drifting in unintended directions?
- Can live or recorded practice be observed to enable reflection on the moment-by-moment ways in which practitioners attempt to implement the program’s theoretical approaches and behaviour change models?
- Does program management, and the agency’s ethos, support the theoretical orientation taken in the program?
- Are the program’s theoretical underpinnings, conceptual foundations and assumptions documented in a way that program practitioners can understand and follow, and in a way that assists them to translate these underpinnings and foundations into practice?
- Does the program guide or operational manual have sufficient detail to guide facilitation and practice, yet also retain the flexibility to enable practitioners to be responsive rather than prescribing rigidly controlled practice? (Vlais et al., 2017)

Monitoring program integrity involves more than determining if interventions with men perpetrating DFV are consistent with the program’s theoretical underpinnings. It should also involve gauging whether the program’s risk identification, risk assessment and risk management procedures are operating according to agency policy and jurisdiction-wide MBCP minimum standards, and whether the program is operating as part of an integrated response system.

It is also important to determine whether partner and family safety contact with victims/survivors is being implemented as planned and according to minimum standards. A program that acts with fidelity according to its model of behaviour change but does not put into practice intended policies and procedures regarding risk management and victim/survivor support is not evaluation-ready.

Many of the above considerations are process evaluation questions and might not be known until the evaluation is underway. When there is significant doubt about whether a program is acting with fidelity according to its theoretical underpinnings and conceptual approach, it can be beneficial to split the evaluation in two halves—first, a process evaluation to determine how program implementation might need to be strengthened and, after process evaluation adjustments are made, a second evaluation stage focusing on measuring outcomes.

More information on evaluation readiness and related tools is available from many sources, including:

- *Evaluability assessment: Examining the readiness of a program for evaluation.* <http://www.jrsa.org/pubs/juv-justice/evaluability-assessment.pdf> (Kaufman-Levy & Poulin, 2003).
- Better evaluation: Evaluability assessment. https://www.betterevaluation.org/en/themes/evaluability_assessment (Davies, 2015).



Scoping the evaluation

Understanding the program's theoretical framework

Most MBCPs are grounded in a theory of how they will work to achieve desired outcomes, and borrow from a range of theoretical viewpoints about *why* domestic and family violence occurs and *how* it might be stopped (Mackay et al., 2015; Paymar & Barnes, 2007). Understanding the framework that underpins your MBCP is important because it is this framework that will inform the program logic. The theoretical approach of the MBCP was likely adopted in the early stages of planning and program development and will have informed the content, structure and design of the program.

Theoretical frameworks relate not only to the mechanisms by which a program aims to facilitate changes in men's behaviour. They also relate to other mechanisms through which the program attempts to work towards desired outcomes, such as frameworks informing partner contact and support for victims/survivors, and those informing the program's approach towards coordinated and collaborative practice with other agencies. Multiple frameworks can therefore inform different components of the program.

Articulating the program goal

A clear understanding of the MBCP's overarching goal ensures that program staff, evaluators and other stakeholders have a shared understanding of what the program is working towards. The goal is a broad statement of what the MBCP aims to achieve in the long term: your program's "mission".

There is an emerging consensus in the perpetrator intervention literature that all perpetrator interventions should aim to achieve safety and wellbeing for women, children and others who experience men's use of DFV. While program goal statements need to be relatively brief, it is important for the program and its evaluators to be clear about the meaning of terms used in the goal statement. For example, how is *safety and wellbeing* defined by the program? Does the term incorporate emotional and other forms of safety, in addition to physical safety? Does *wellbeing* include the strength of relationships between adult and child victims/survivors in the family that might have been harmed by the perpetrator's actions? The conceptualisation of key terms should be articulated in program and evaluation documentation following the program goal statement.

Presently, the most recognised understanding of "success" in MBCPs is typified by that outlined in the Project Mirabal study (Kelly & Westmarland, 2015). This United Kingdom-based project measured program success as "the extent to which perpetrator programs reduce violence and increase safety for women and children, and the routes by which they contribute to coordinated community responses to domestic violence" (Kelly & Westmarland, 2015, p. 7).

Project Mirabal (Kelly & Westmarland, 2015) included interviews with female partners and ex-partners, male participants, practitioners and funders to identify what success

in their MBCP meant to them. The results revealed that women were focused on six outcomes of success, which moved beyond stopping the violence and included the following:

1. respectful communication
2. expanded space for action for women, which restores their voice and freedom to make choices, while improving their wellbeing
3. safety and freedom from violence
4. safe, positive and shared parenting
5. enhanced perpetrator awareness about the impact of violence on others
6. safer lives for their children. (Kelly & Westmarland, 2015)

It is likely that your program goal will include some or all of these outcomes. However, programs may have other goals that are worth articulating. For example, the aim of the Caring Dads group intervention program is

to engage fathers who have used violence, to help them develop skills in child-centred fathering and take responsibility for the impacts of their violence upon their children and their children's mother. (Diemer et al., 2020, p. 12)



More information on articulating program goals can be found at https://www.betterevaluation.org/en/rainbow_framework/frame/decide_purpose (Better Evaluation, 2016a).

Quality of life indicators

Recent Australian research has focused on proposing a set of women's quality of life (QOL) indicators as outcomes for evaluating MBCPs (McLaren, Fischer, & Zannettino, 2020). One hundred women, 71 of whom had partners who had participated in a DFV perpetrator intervention program, were asked what quality of life meant to them. The most frequently endorsed QOL themes were autonomy, informal supports (family and friends), emotional health, safety (physical and psychological), children and pets, and mental health.

The authors argued that measuring women's QOL indicators before and after their (ex-)partner's participation in an MBCP provides a way of determining whether women's lives have improved, without needing to focus research interviews directly on the women's experiences of DFV (McLaren et al., 2020). They speculated that this might have potential in increasing women's participation rates in MBCP outcome evaluation studies, as the women would not be required (unless they choose) to talk directly about the violence or about the man's participation in the program. Given the substantial impact of DFV across a range of facets of women's lives, employing QOL indicators offers a promising avenue to focus outcome measures on the fundamental goal of MBCPs.

Developing a program logic: Describing the program and its mechanisms

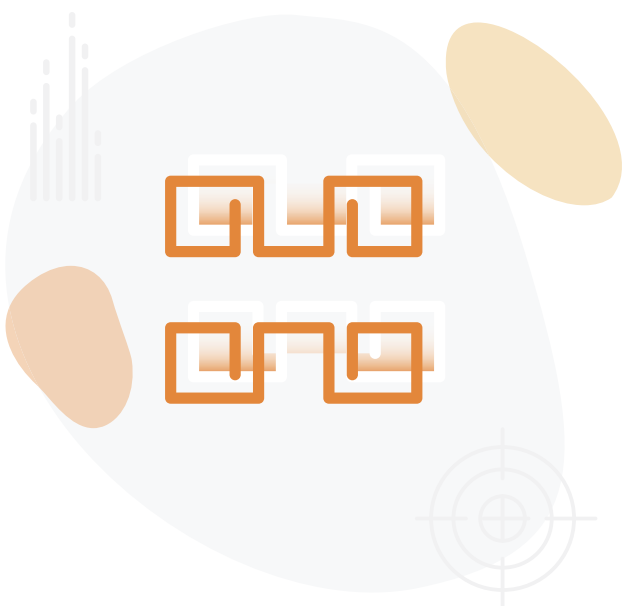
A program logic is a systematic way to present and share your understanding of the relationships between the resources used in the MBCP program, the activities you plan, and the changes or results you hope to achieve (Kellogg, 2004), including your articulated program goal. A program logic, if sufficiently detailed, articulates the “mechanisms” of how each of the MBCP activities leads to men reducing and ultimately ending their violent and controlling behaviour and to victims/survivors experiencing greater safety and wellbeing.

Why do you need a program logic?

Articulating the theoretical mechanism is important for many reasons:

- It can help you and all practitioners involved in running the program understand how the activities of the program are expected to lead to the desired outcomes.
- It is a useful device to use in collaboration with stakeholders to create a shared understanding of the desired outcomes of an MBCP program and the activities designed to achieve these (McKenzie, White, Minty, & Clancy, 2016).
- It can identify possible barriers to the quality implementation of the activities and the achievement of the desired outcomes (i.e. it helps to identify points at which the program implementation might “fall down” or where preceding outcomes may not lead to subsequent ones).
- It helps to inform the evaluation, particularly by helping in the selection of evaluation questions.

Logic models are not necessarily static, and can be revised or updated with continued program learning through formal and informal evaluation.



EXAMPLE**Developing the program logic with evaluation users**

For one evaluator we interviewed, the development and consideration of a program logic model had a significant impact on the evaluation users and helped to crystallise the evaluation questions:

Participants found the program logic workshops very useful, as it enabled them to reflect on the fullness of the intent of the programs, to consider differences between sites, and for the facilitators to consolidate their thoughts regarding their expectations and hopes for the program. As well as being useful for the program teams, it also helped the evaluation team to develop questions for future interviews with the facilitators.

Another evaluator suggested that adequate involvement of practitioners in the development of the program logic was essential to improve the external evaluators' understanding of the program:

Evaluators need to work collaboratively with the practitioners to develop the evaluation [program] logic; this helps practitioners to be on side with the evaluation, because the evaluation then reflects how the practitioners perceive the program and what's important for them to find out through the evaluation. Evaluators can come in with their own narratives or stories regarding the priorities for the evaluation, shaped by a range of influences (including expectations from the commissioner of the evaluation), and if there isn't sufficient involvement of practitioners in the development of the program logic, the essence of what needs to be evaluated can become lost.

The importance of collaboration was raised by another evaluator, who underlined the essential role of evaluators "doing up-front work" and developing strong working relationships with the MBCP practitioners. Co-designing the program logic might be one way to build this relationship:

Time and relationship-building effort is required for the program practitioners to be on board with the evaluation. Them being on board is crucial for things like recruiting program participants to be interviewed. If they aren't on board, evaluators can encounter a whole lot of problems across the evaluation. It's worth investing the collaboration-building time.

Developing the program logic with evaluation users

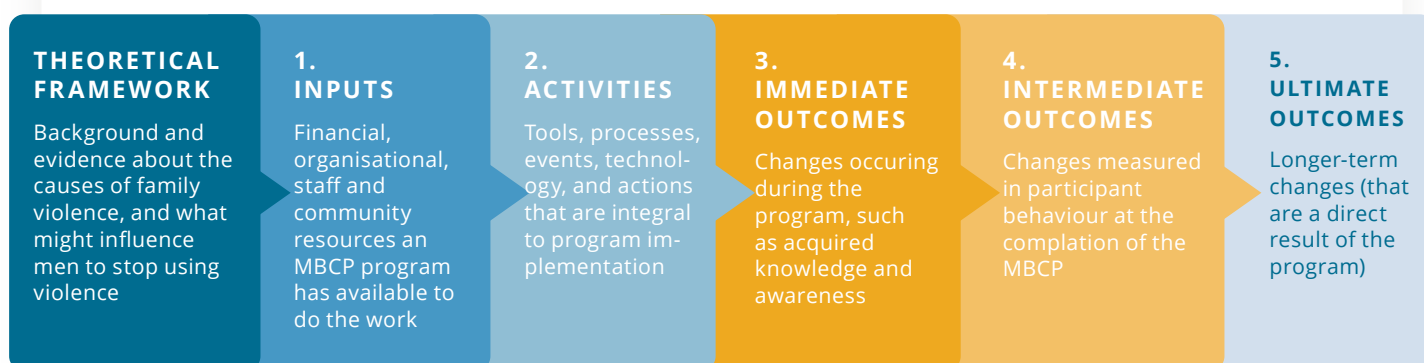
Where possible, it can be useful to develop the program logic with the evaluation users. The involvement of users in the development of the logic could range from providing feedback on a first draft written by the evaluator, to collaborative workshops where the logic is "co-designed". While program funders, especially those commissioning the evaluation, have an important stake in collaborating in program logic design, the program providers themselves will be in the strongest position to co-design the logic with the evaluator. Ideally, the program provider would have already constructed a program logic as part of the program's initial development process, but this is not always the case. The examples below show some of the advantages of this collaborative approach to the development of the program logic for the evaluators and for the evaluation users.

Key components of the program logic

The key components of a program logic are illustrated in Figure 1. It is worth noting here that there is much variation in the language used in evaluation, with different sources calling the same concepts by different names. For example, what we have

called “outcomes” are sometimes called “impacts”, and “inputs” are sometimes called “resources”. For simplicity and consistency, we use the terms outlined in Figure 1 throughout this guide. “Outputs” might be included in your program logic (see Box 1).

FIGURE 1:
Components of the program logic model



BOX 1:
A note on “outputs”

Outputs relate to the “products” of the activities of the MBCP. For example, an output of an MBCP risk review meeting (the activity) could be risk management plans for higher risk men (the output). Outputs are often confused with the indicators of the activities—for example, “number of MBCP group work sessions” as an output measure of the activity “MBCP group work sessions delivered”. Therefore, they often duplicate the activity, sitting alongside it in the program logic. The program logic should aim to demonstrate the hypothetical mechanisms of the program or how the program is supposed to work. The number of MBCP group work sessions is instead one indicator of how this activity was implemented. Outputs can be included in program logics, but take care that they are not just replicating the activity as many logic diagrams do.

Adapted from Knowlton and Phillips (2012)

The framework represented in Figure 2 is used to present a simple program logic for a hypothetical MBCP program. While this program logic illustrates the inputs, general program activities and immediate, intermediate and ultimate outcomes, the linkages between each are minimal. The program logic presented in Figure 3, however, is an example where a greater number of linkages can be used to illustrate the theoretical mechanism underpinning the program in more detail. Detail such as this can inform program design and identify potential weak points on which to focus additional resources, or where unintended consequences of the program might arise. If we look at Figure 3 under “activities”, at the first box: if the perpetrators are not deemed “ready” for the group-based component of the program as determined by the intake assessment criteria for suitability, but are nevertheless immediately enrolled in the group work component of the program without one-to-one sessions to improve readiness, then the program is unlikely to achieve some of its intended outcomes. In fact, we might instead see negative unintended outcomes as a result, whereby a perpetrator who was not ready at time of intake might undermine the facilitator in the group work or use some of the information provided to expand his controlling behaviours.

Focusing on multiple change mechanisms in the program logic

It is important that your program logic captures all of the main mechanisms and components through which your program works towards its ultimate goal. MBCPs work towards desired ultimate outcomes through multiple mechanisms. Attempting to change men's violence-supporting attitudes and behaviours is, of course, one of them. However, other mechanisms include the provision of partner and family safety contact, and risk management processes that identify and respond to perpetrator-driven risk. For MBCPs to operate safely, ethically and according to jurisdiction-based minimum standards, mechanisms to provide victims/survivors with partner and family safety contact, to identify and respond to perpetrator-driven risk, and to collaborate with other agencies as part of an integrated system response, are as important as mechanisms that work towards changing perpetrator attitudes and behaviours.

By focusing only on the mechanisms through which the program works towards men's attitudinal and behavioural change, the impacts of other mechanisms that work towards the ultimate program goal remain invisible. The evaluation then captures only part of the story.

FIGURE 2:

A hypothetical simple logic diagram for an MBCP adapted from a program logic by No to Violence

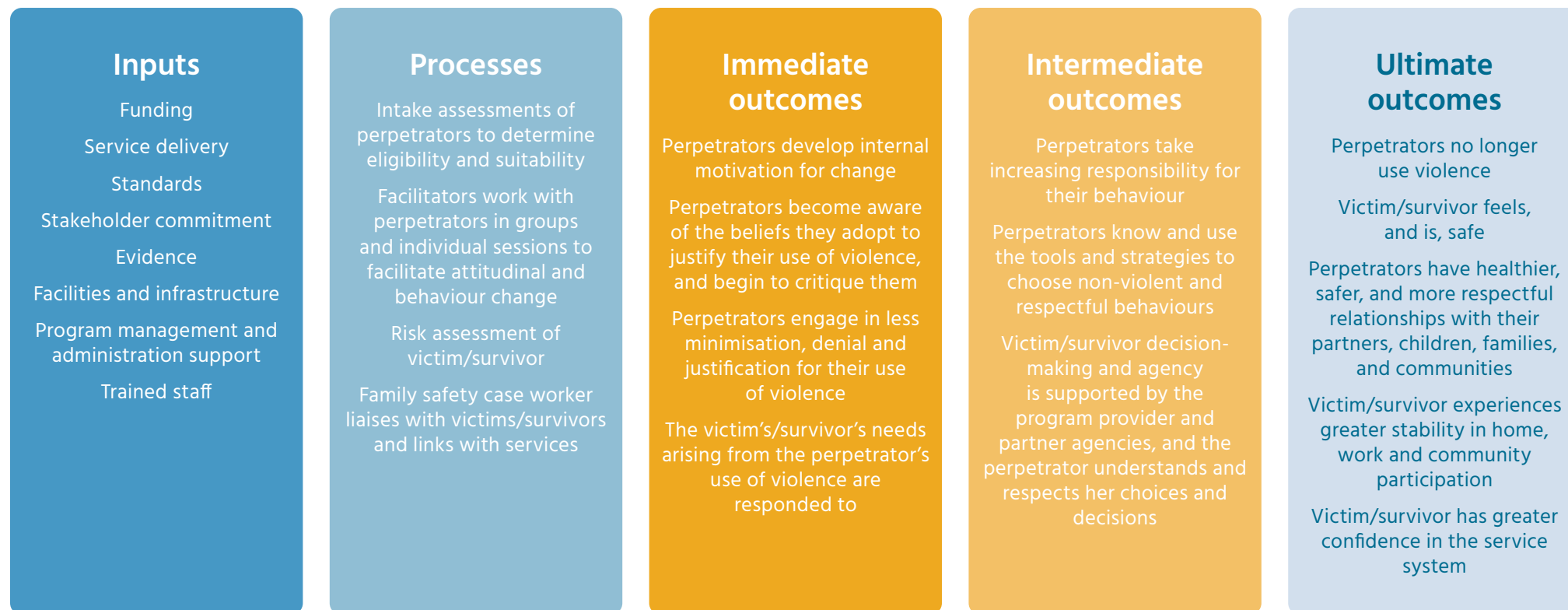
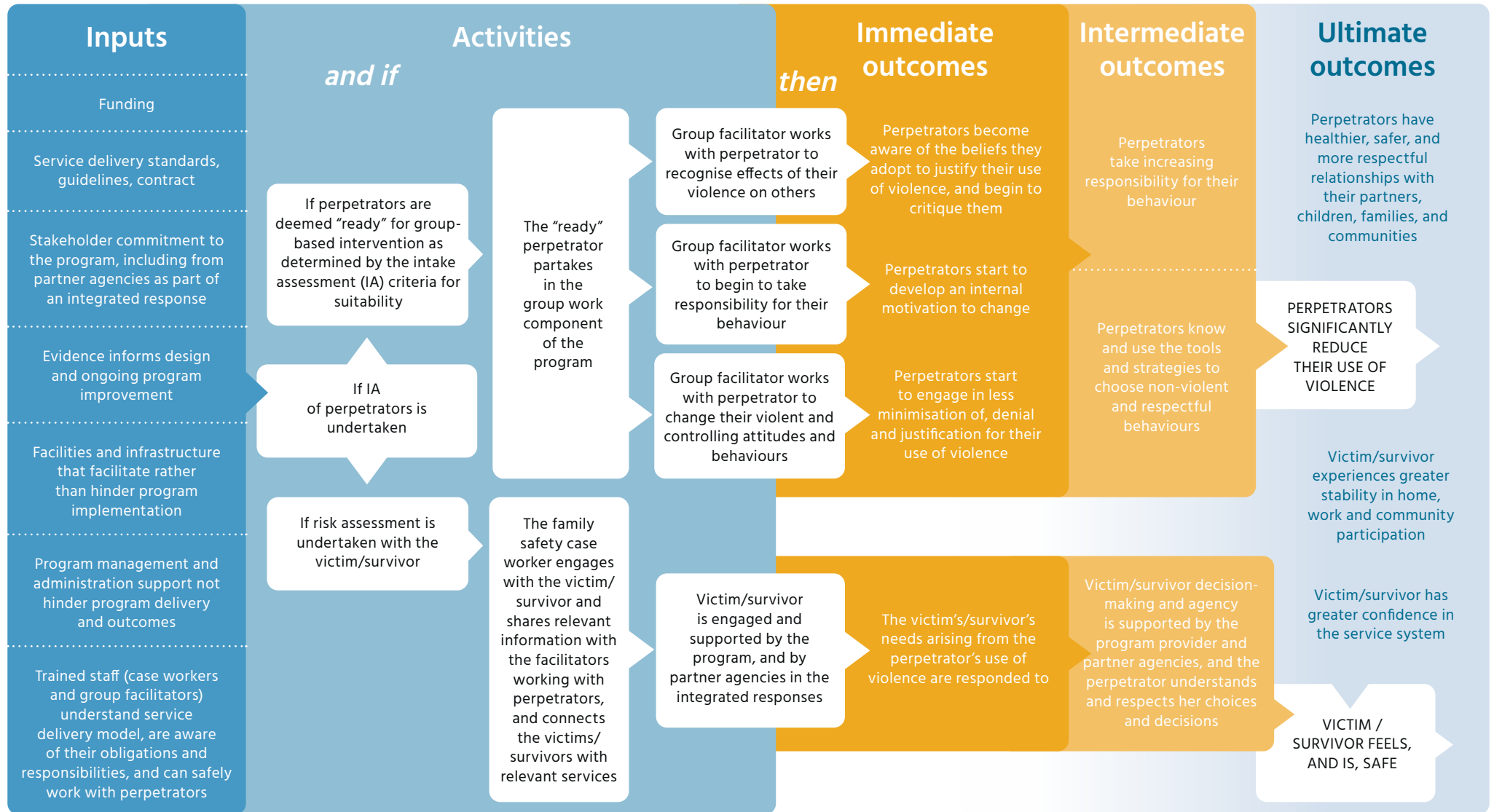


FIGURE 3:

A hypothetical “if ... then” program logic for an MBCP adapted from a program logic by No to Violence



Developing and prioritising evaluation questions

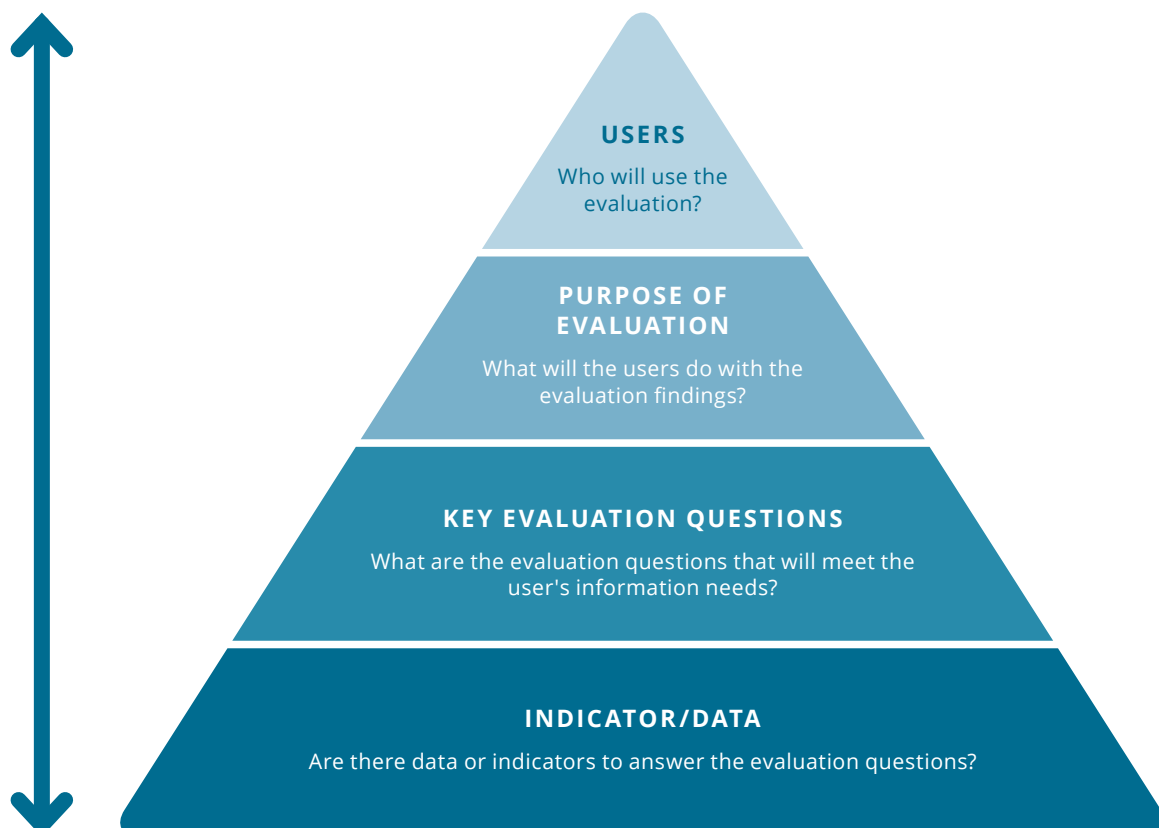
Evaluation questions can relate to any point in the program logic: the inputs, for example, “Is there support for the program in this community?”; the activities, for example, “Are victims/survivors engaged and supported by the program?”; and the outcomes, for example “Does the MBCP improve the safety of women and children?” (ultimate outcome). There are many possible evaluation questions, so how do you decide which to choose? In this section, we consider the steps to creating useful key evaluation questions. Figure 4 illustrates that the evaluation users’ information needs should inform the purpose of the evaluation, which should in turn determine the evaluation questions. From there, the evaluation questions should inform the indicators or data collected. The arrows illustrate the linkages between the users’ needs, the purpose of the evaluation, the key questions and data.



More information on developing evaluation questions can be found at https://www.betterevaluation.org/en/rainbow_framework/frame/specify_key_evaluation_questions (Better Evaluation, 2016b).

FIGURE 4:

Steps in developing evaluation questions



Working with evaluation “users” (stakeholders)

It is wise to create evaluation questions in consultation with key stakeholders who are likely to be “users” of the evaluation (Figure 4). Evaluation users are people or groups who will or could use evaluation findings to make decisions about the program. Different users will have varying views on the most important type of information to gather from the evaluation because they will use that information for different purposes (Figure 4). Users might be people who have inputs into the program (e.g. government funders), whose actions are affected by the program (e.g. victims/survivors), and/or those who interact directly with the program (e.g. participants, program managers, facilitators and referrers). When working with stakeholders to identify evaluation questions, it is important to consider “who will use the evaluation and for what purpose”. Therefore, it is a useful step to list your evaluation users and consider consulting with each of them regarding how they want to use the results of the evaluation.

Engaging users in developing the evaluation questions will increase the usefulness, relevance and credibility of the evaluation. Thinking strategically, users who are engaged in the evaluation process are also more likely to support it and act on any outcomes or recommendations that arise from the evaluation.



More information on identifying the evaluation users and their uses for the evaluation can be found here: <https://idl-bnc-idrc.dspacedirect.org/bitstream/handle/10625/47278/133624.pdf?sequence=1&isAllowed=y> (International Development Research Centre, 2012).

Determining which evaluation questions to include

There will be many possible evaluation questions, especially when there are multiple evaluation users, each with their own priorities. As one of the participants in our consultation stated:

There is such a diversity in the cultures of stakeholders, both in terms of practice and also in terms of what they want from evaluation ... there is often clashing with stakeholders, who might want very different things.

Once you understand how evaluation findings will be used you may need to prioritise some evaluation questions, depending on a number of factors such as:

- “non-negotiable” questions required by the funder
- resources available (e.g. budget)
- timing of the evaluation (how long you have to do it)
- decisions to be informed by the evaluation (determined by consulting with users)
- which stakeholder groups to be consulted.

For instance, it might be a condition of the program funding agreement to evaluate and report on particular aspects of the MBCP. For some stakeholder groups (e.g. funders and program managers), it would be useful to determine decisions expected to be informed by the evaluation (e.g. funding decisions and program resourcing) and when they will be made to ensure that evaluation findings are relevant and timely. There also might be an opportunity for presenting the findings (e.g. a national conference, community forum, or organisational strategic review).

Negotiation is sometimes required to ensure that evaluation questions are realistic for budget, timeframe and data collection. Some stakeholders might be most interested in questions related to ultimate outcomes, which in the MBCP field often require relatively large evaluation budgets and an evaluation timeframe upwards of 2 years. Some stakeholders might place less value on process evaluation and whether an MBCP is being implemented with integrity.

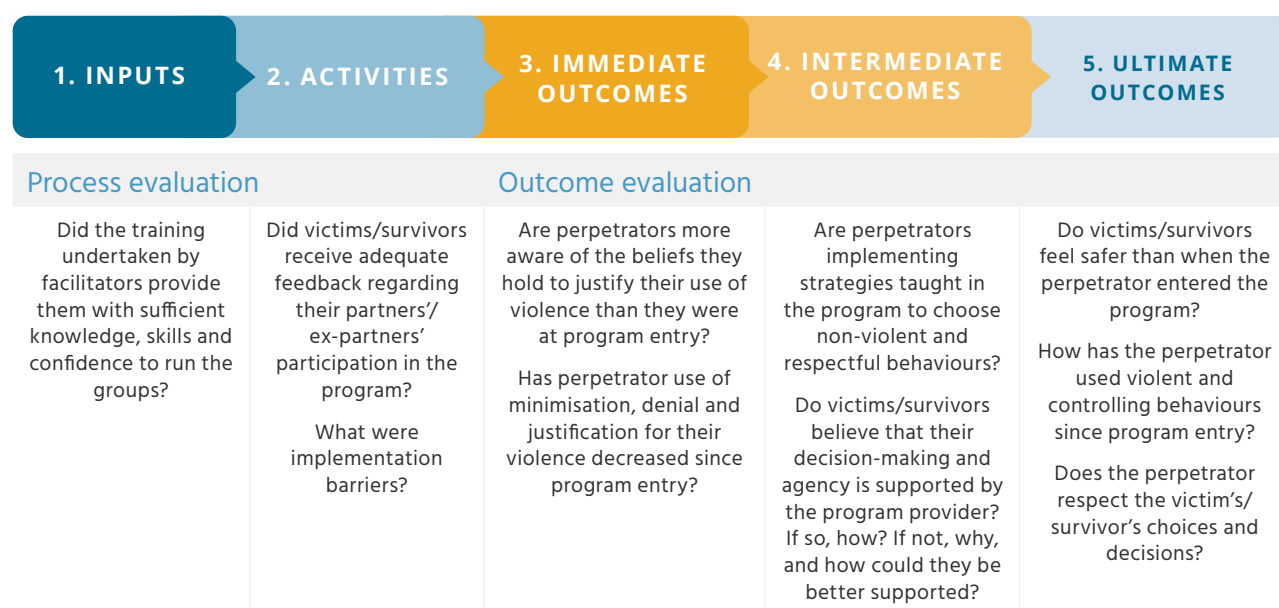
The program logic will also assist in deciding the order of answering evaluation questions. Evaluation questions on inputs and processes, for example, may need to be answered before questions about outcomes. This also relates back to evaluation readiness, where it is important to know if your program is being implemented properly before you explore program outcomes.

Developing evaluation questions from your program logic

Generally speaking, there are two main types of evaluation questions: process evaluation (how the program was delivered) and outcome evaluation (MBCP achievements). Figure 5 gives an example of how to use the program logic illustrated in Figure 3 to map evaluation questions.

FIGURE 5:

Examples of evaluation questions that can be linked to the program logic in Figure 3





Process evaluation

Process evaluation involves describing and assessing program inputs and activities and linking these to outcomes. Process evaluation questions are primarily concerned with the implementation, or process, of delivering MBCP activities. Process evaluation aims to assess how well the program is working, whether it is being implemented as designed (sometimes called “fidelity”), and whether it is accessible and acceptable to the target population. Process evaluation provides data and information about which aspects of the MBCP might need to be addressed or improved.

MBCP evaluators we interviewed placed a strong emphasis on process evaluation and included questions about participant recruitment (referral processes, participant characteristics, program operation), retention (in the overall service system), facilitator training and issues with program integrity. Importantly, process evaluation is key to determining evaluation readiness and whether the program has been implemented in a way that enables the MBCP to be ready for an outcome evaluation.

Process evaluation can reveal the internal workings of the program, but also how the system in which the program sits affects its implementation—for example, how men are being referred into the program, whether they are sufficiently prepared by the referrer and come into the program with realistic expectations, the overall impact of these factors on program retention, and how the program provider and other agencies work together to share information and manage risk. In this way, process evaluation helps determine whether the program has been implemented as planned, sits appropriately within the broader system, and is ready for an impact and outcome evaluation.

Process evaluation questions should integrate components of the program logic (i.e. inputs and activities). Table 2 shows examples of this.

TABLE 2:

Examples of process evaluation questions arising from the program logic in Figure 3

Process evaluation questions	Indicative questions
Process evaluation questions relating to inputs	<ul style="list-style-type: none"> • Is the program in line with current evidence on what works in MBCPs? • What is the cost per participant of running the program? Is it being sufficiently funded? • How can program managers better assist facilitators to implement the program?
Process evaluation questions relating to activities	<ul style="list-style-type: none"> • Does the current intake process adequately identify perpetrators who are suitable for the MBCP? • Is there sufficient communication between the victim/survivor case worker and the facilitator to inform the facilitator’s work with perpetrators in the program? • Are participants satisfied with MBCP activities? • Does the victim/survivor whose partner/ex-partner is involved in the MBCP feel their needs have been met? • Is the program encouraging perpetrators to take responsibility for their behaviour? If not, why not? • Which aspects of the facilitated session delivery can be improved and how? • What are program implementation barriers?

BOX 2:**Process evaluation example****An important question: “What are the barriers to implementation of the program?”**

One MBCP evaluator reported examples of facilitators’ answers when asked about barriers to implementation. Some barriers were systems-generated, such as not receiving enough referrals to the program or a lack of clarity about program processes such as how to report back to referrers. Barriers were assessed at numerous time points to explore how facilitators addressed barriers throughout MBCP implementation.

A systems-level approach to process evaluation

Evaluators explained the importance of a systems-level approach to process evaluation:

New MBCPs can struggle to get up and going because they rely on establishing and strengthening new referral pathways with other agencies and sectors. When a new program arises, potential referrers might misconstrue what the program is about, have unrealistic expectations concerning what it can achieve, it might be difficult to recruit suitable trained practitioners who practice in a way that’s consistent with the program’s theoretical approach, et cetera. A process evaluation enables these and other issues to be explored, rather than wasting evaluation dollars at such early stages on evaluating impact and outcome—this can be a waste, as if the program hasn’t yet got off the ground in the way that it’s intended, then an impact evaluation is premature.

The extent to which a program works is dependent heavily on the system in which the program is embedded. The evaluation’s methodology was finely tuned to focus on this system as this MBCP represented a new program in Australia. What was most pressing in this evaluation was how this new program could be introduced successfully into a pre-existing system. How do other services relate to it? The focused process evaluation methodology, including interviews with referrers, and the systems focus, enabled them to get at the roots of any barriers, and what would be needed to address these barriers.

An overarching consideration:**Working with program providers throughout the evaluation**

One evaluator raised the ethical issue in evaluation design of reporting back to program providers throughout the evaluation, saying that it was unethical when conducting a long-term evaluation to wait until the end to report on findings. This issue is particularly relevant to MBCP evaluations when interim findings indicate that program activities seem to be having a positive or negative impact on victim/survivor safety, suggesting that certain activities should be either scaled up or changed.

Having an ongoing relationship with providers is also important for evaluation quality. A close relationship can identify difficulties early and ensure they are resolved such that they do not threaten the integrity of the evaluation. For example, it would be important to identify if service providers do not understand what data needs to be collected or how to record it and then provide clarification and assistance to ensure data integrity.

If changes to the program are made during the evaluation on the basis of interim or ongoing evaluation results, these need to be documented in the evaluation report and considered in an evaluation of program outcomes.



Outcome evaluation

Outcome evaluation focuses on the immediate, short- and long-term effects of an MBCP on its target populations (i.e. perpetrators, women and children). Generally, outcome evaluations of MBCP programs aim to provide evidence of change in attitudes, motivations, awareness, behaviours and experiences (e.g. victims/survivors' safety and wellbeing) over time. Outcome evaluation questions should reflect the outcomes depicted in your logic model (i.e. immediate, intermediate, and ultimate outcomes). Some examples are included in Table 3.

TABLE 3:

Examples of outcomes evaluation questions devised from the program logic in Figure 3

Outcome evaluation questions	Indicative questions
Immediate outcomes evaluation questions	<ul style="list-style-type: none"> • Are perpetrators challenging beliefs justifying their use of violence upon completion of the program? • Are perpetrators engaging in less denial and justification of their use of violence than when they entered the program? • Are victim/survivor needs arising from the perpetrator's use of violence being responded to?
Intermediate outcomes evaluation questions	<ul style="list-style-type: none"> • Are perpetrators demonstrating increased use of non-violent and respectful behaviours toward their partner from when they entered the program? • Are perpetrators respecting the choices and decisions of their partner/ex-partner more than when they entered the program? • Do victims/survivors believe that their decision-making and agency is being supported by the program provider and partner agencies?
Ultimate outcomes evaluation questions	<ul style="list-style-type: none"> • Are perpetrators who have completed the program engaging in fewer violent, coercive and controlling behaviours and is this change sustained over time? • Are victims/survivors experiencing greater stability in their life (home, work, community participation) than when the perpetrators entered the program? • Are perpetrators relating to children in more respectful and child-centred ways than when they entered the program?



Evaluating MBCP ultimate outcomes

The overarching goal of MBCP evaluation is to examine if the outcomes of increased safety and wellbeing for women and children are achieved, and if these outcomes are sustained over time (Gondolf, 2012). It is important to note that the outcome of “increased safety and wellbeing for women and children” may be conceptualised in numerous ways. For example, is it about preventing injury from physical violence? Is it about victims/survivors’ perceived feelings of safety? Is it about freedom to live their lives and release them from their partner’s coercive control? Or is it about improvements in quality of life indicators that matter most to women and children?

It is important that ultimate outcomes measures focusing on victim/survivor safety do so in ways that are not solely linked to changes in the perpetrator’s behaviour. Victim/survivor safety can be enhanced by an MBCP even in situations where a perpetrator does not change his behaviour—for example, if the program in the short term is able to help contain the risk a perpetrator poses to their partner and family, and over the longer term support a victim’s/survivor’s wellbeing.

The focus on the safety of women and their children as the core priority of all MBCPs also encompasses wider issues of risk and ethical considerations in evaluation, which will be addressed in the following sections.

Other considerations regarding measuring outcomes

Measuring outcomes for victims/survivors

Program evaluations of MBCPs should incorporate a dedicated evaluation stream focusing on the partner and family safety contact component of the program, including data obtained from victims/survivors about their experiences of partner and family safety contact and the impact it has (or hasn’t) had. It is important that outcomes measures focusing on victim/survivor safety do so in ways that are not solely linked to changes in the perpetrator’s behaviour. Victim/survivor safety can be enhanced by an MBCP even in situations where a perpetrator does not change his behaviour—if, for example, the program in the short term is able to help contain the risk a perpetrator poses to their partner and family, and over the longer term support a victim’s/survivor’s wellbeing.

MBCP evaluations can include a focus on questions about the ability of the program to identify and respond appropriately to new or escalated perpetrator-driven risk. In some instances where a program is unable to change a perpetrator’s attitudes or behaviours, managing the risk he poses to adult and child victims/survivors can be an important outcome. This includes the extent to which the program is able to collaborate effectively with other agencies towards a multi-agency risk management response.

Systems-level outcomes

Evaluating behaviour change outcomes can include the ways in which the program is intended to exert an influence on the integrated DFV response system of which

it is a part. Some program logics in the MBCP field therefore divide immediate and intermediate outcomes into two sections—one that represents systems-level outcomes, and the other at the program level.

Immediate systems-level outcomes are important to consider in the course of process evaluations. Examples include:

- agencies refer appropriately into the program
- referring agencies actively collaborate with the MBCP provider during the man's participation in the program
- agencies appropriately and proactively share and exchange information with the MBCP provider to help to identify, assess and manage perpetrator-driven risk.

Intermediate systems-level outcomes can include:

- the MBCP provider and partner agencies involved in an integrated response understand their own, and each other's, roles and responsibilities in engaging effectively with perpetrators and in responding to perpetrator-driven risk
- partner agencies and other stakeholders develop a nuanced understanding of MBCPs and adopt realistic expectations about their effectiveness
- the MBCP assists child protection and intensive family support services to hold perpetrators responsible for their impacts on children's welfare and family functioning and to ally with the non-offending parent
- the MBCP contributes to the ability of the integrated DFV response to manage high-risk, high-harm perpetrators.

Of course, there are many things that can influence intermediate (and ultimate) systems-level outcomes, beyond the influence of the MBCP. It is important to state these outcomes in terms of what the MBCP may impact.

Measuring ultimate outcomes

Many issues specific to MBCP evaluation complicate the likelihood of determining the achievement of "ultimate outcomes" (Vlais & Green, 2018), such as maintaining contact with victims/survivors over the long term when they are no longer partnered with the perpetrator, and maintaining contact with perpetrators once they have completed the program. This difficulty was illustrated by an evaluator we interviewed:

In most programs, the follow-up numbers of men that will respond to a questionnaire after a program is completed is below 10 percent ... We follow up with all men that agree to participate in the evaluation (usually just over 50%), but less than 10 percent of that cohort respond after 6 months.

As a result, many MBCP evaluations focus more on the *immediate* and *intermediate outcomes* of the program rather, than on *ultimate outcomes*. While this has long been viewed as problematic emerging research suggests validity in using robust measures to assess shorter-term change (Semiatin, Murphy, & Elliott, 2013; Silvergleid & Mankowski, 2006). These immediate and intermediate outcomes could include changes in perpetrators' attitudes and behaviour, assessments of perpetrator risk and accounts of victim/survivor autonomy and wellbeing. Some researchers argue that

measuring more immediate outcomes shows progress towards ultimate outcomes of a program logic (Silvergleid & Mankowski, 2006; Vlasis & Green, 2018). Thus, while it is important to continue to attempt to assess ultimate outcomes, immediate and intermediate outcomes can be used as indicators of ultimate outcomes.

Determining realistic outcomes of MBCPs

Recent increases in funding provision for MBCPs have meant that program providers are under greater pressure to produce adequate evidence of program effectiveness (Vlasis & Green, 2018). At the same time, there is growing acknowledgement in the sector that for most perpetrators, perpetrator engagement in a single MBCP is unlikely to lead to long-term and sustained behaviour change (Dutton & Corvo, 2006), and that MBCPs need to be viewed within a wider system of supports and interagency measures that aim to reduce violence against women and keep a perpetrator “in view”. Recent shifts in policy and practice highlight the importance of multi-agency information sharing during and following MBCP completion. Service knowledge of perpetrator behaviour, motivation to change, and potential risk informs decisions of other agencies and ensures that perpetrators are kept “in view” through integrated systems (No to Violence, 2006; Respect, 2017).

An important evaluation question to consider: Unintended outcomes

It is possible that unintended outcomes arise from an MBCP, which may be positive (additional value-added benefits) or negative (harmful). For example, one unintended outcome of psychoeducation on types of abusive and controlling behaviours may be that MBCP participants actually “learn” about and take on new problematic behaviours. Other types of unintended negative consequences of MBCP activity unfortunately appear reasonably common: some perpetrators attempt to use their participation in a program to manipulate service system and community responses for their own benefit (Opitz, 2014; Vlasis & Campbell, 2019), or distort what is covered by the program to extend control over their partner (Wistow, Kelly, & Westmarland, 2017).

Evaluation stakeholders, such as program facilitators, who are sometimes able to observe outcomes directly might be in a good position to pick up unintended outcomes like this. It is important, therefore, to ensure that these stakeholders are given the opportunity to discuss outcomes beyond those the evaluation is designed to measure.

Measuring unintended outcomes might be done in a direct way, such as by asking victims/survivors directly if there have been any additional negative or positive outcomes or consequences of the program after asking other evaluation questions. Another way of identifying positive and negative unintended outcomes is to use the “most significant change” technique (Davies & Dart, 2005) discussed later in the guide. Examples are shown in Table 4.

TABLE 4:

Examples of questions and indicators for assessing unintended consequences of MBCPs

Example evaluation questions	Example indicators
<ul style="list-style-type: none"> • What are the ways in which men's participation in the MBCP could inadvertently cause harm to family members? • Is identification of these risks and impacts a customary focus in family safety support work? • How are these risks and impacts minimised? 	<ul style="list-style-type: none"> • Documented occurrences of men using program participation as a tactic of control in the family safety support contact recording tool or risk management register/matrix • Documented occurrences of other harms connected with men's participation determined by auditing a representative sample of case files or completion reports to referrers • Practice guidance and prompts in the program manual for facilitators to monitor, identify and respond to harm connected with MBCP participation

Transferability

Transferability is "the extent to which the measured effectiveness of an intervention could be achieved in another setting" (Rychetnik, Frommer, Hawe, & Shiell, 2002, p. 119), including in a different place, organisation or target group. Whether evaluation findings are transferable depends largely on MBCP and context similarities. In our interviews with MBCP evaluators, the difficulty of generalising evaluation results was raised. Evaluators stressed how substantially different MBCPs are in terms of approach, program length, practitioner skill level, organisational support for the program, degree and depth of partner contact, and integration with the broader system, making MBCP evaluation findings unlikely to be transferable.

To assess whether a program will be effective in a different setting, it can be useful to ask the questions outlined in Box 3. Transferability is important to consider when deciding whether to implement a program that has been tested in another setting and when reporting evaluation findings for others to make decisions about transferability of your MBCP.

BOX 3:

Assessing transferability of MBCPs

To determine to what degree the results of one evaluation might be transferable to another, it is useful to ask the following questions:

- What is the level of DFV in the community in which you will implement the MBCP? How does it compare to the communities exposed to the primary MBCP?
- Are characteristics of men to be involved in the program similar to those in the program you want to transfer? For example, are they of similar cultural backgrounds and socioeconomic status? Is it possible that differences in characteristics will affect MBCP implementation or effectiveness? What adaptations might be made?
- Are the resources available to implement the MBCP similar to those for the program you want to transfer? For example, is there a similar organisational environment in terms of the partner agencies available to deliver an integrated response? Is there the same or a similar level of organisational support?

Adapted from Wang, Moss, and Hiller (2006, p. 79)

Replicability considers whether an MBCP—and which of its parts—should be implemented elsewhere. This may have important funding implications.



Additional information about transferability can be found here: <https://academic.oup.com/heapro/article/21/1/76/646412> (Wang et al., 2006).

A useful framework for thinking further about replicability can be found here: <https://cdn.catch-22.org.uk/wp-content/uploads/2015/11/Realising-Ambition-Programme-Insight-Evidence-is-Confidence.pdf> (Catch-22, n.d.).

Answering the evaluation questions

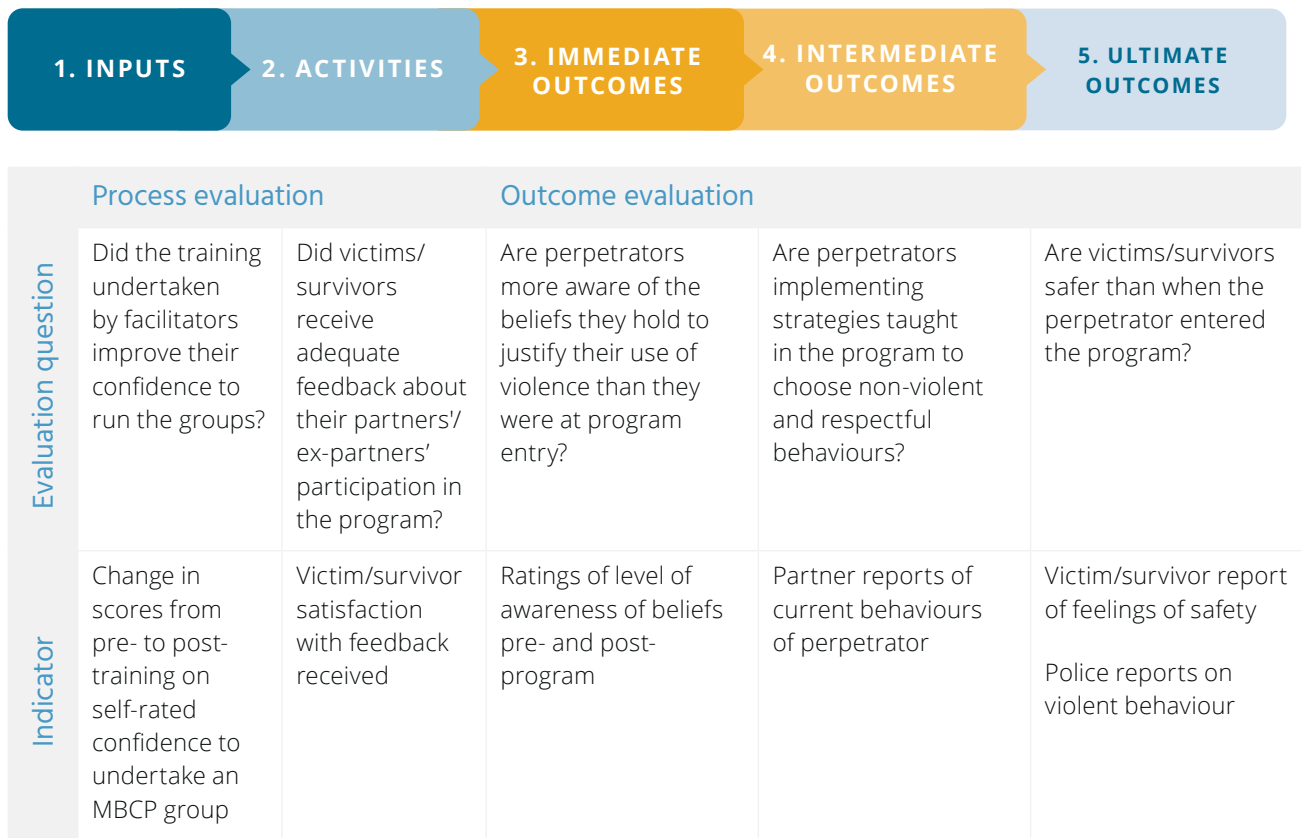
Choosing key indicators to answer your evaluation questions

Once a program logic and evaluation questions are decided, it is time to consider *how* these evaluation questions will be answered. To do this, you need to consider key indicators for each of the evaluation questions. An indicator is “a specific, observable, and measurable [marker of] accomplishment or change that shows the progress made toward achieving a specific output or outcome in your logic model” (Salabarría-Peña, Apt, & Walsh, 2007, p. 175). Generally speaking, indicators can be quantitative or qualitative. Most frequently, evaluations of complex programs, like MBCPs, use both types of indicators.

Quantitative indicators are reported numerically and can measure the scale of changes produced through the program (e.g. how much has changed and how many have changed). Qualitative indicators tend to describe “meaning and subjective experience” (Mark, Henry, & Julnes, 2000, p. 161), allowing you to answer questions unrelated to quantity or scale. For example, the evaluation questions “What are barriers to program delivery?” or “How can this aspect of the program be improved?” may result in the following indicators: “facilitator perceptions of how the program can be improved” or “facilitator views on barriers to perpetrators taking responsibility for their behaviour”. The indicators you select in your evaluation should contribute to answering your evaluation questions. Figure 6 shows some possible indicators that could be used to answer some of the process and outcome evaluation questions we derived from the program logic earlier in the guide.

FIGURE 6:

Indicators to answer evaluation questions arising from the program logic in Figure 3



Choosing quality indicators: SMART indicators

There will likely be many possible indicators that can be chosen to answer your evaluation questions. One way to assess the quality of proposed indicators is to assess whether it is a SMART indicator: specific, measurable, achievable, relevant and time-bound. Table 5 shows the example of using a facilitator's self-report of confidence to conduct the MBCP sessions as an indicator, which might be used before and after training.

TABLE 5:
How to assess indicators against the SMART criteria

SMART	Criteria	Meets criteria?
Specific	Does the indicator relate directly to the desired outcome?	✓
Measurable	Could we repeat the measurement? Do we have the resources required to undertake the measurement?	✓ Yes, though self-reported confidence is subjective and may vary over time
Achievable	Is the expected change achievable as a result of the intervention?	✓ If set reasonable expectation about how much confidence will increase
Relevant	Does the indicator reflect the expectations of stakeholders?	✓
Time-bound	Could the expected change happen in the measurement period?	✓ Could also include follow-up assessments to identify additional training needs after facilitation begins, as confidence may change once the groups commence



This information is adapted from “What makes a good performance indicator”, found here: <https://coastadapt.com.au/how-to-pages/identifying-indicators-monitoring-and-evaluation> (CoastAdapt, 2017).

Note that you can have more than one indicator for each evaluation question. In fact, it is often advisable to use more than one indicator to answer the same evaluation question. This is called data triangulation.

Data triangulation

Data triangulation in evaluation means combining different indicators to answer the same evaluation question. Triangulation can involve using different methods to get a better view of the answer to a particular question (also sometimes called “mixed methods”). For example, to measure “reduction in violent behaviour”, an evaluator might access police records and also ask victims/survivors about a perpetrator’s

frequency of violent behaviours. Data triangulation is also using the same data collection method with different points of view. For MBCPs this will likely involve getting the perspective of the perpetrator, victim/survivor, facilitators, and/or other case workers.

Problematic indicators requiring triangulation

Historically, there have been commonly used indicators in MBCP evaluations that are now considered less valuable, particularly when used alone. It may be worth considering why these are no longer considered sole indicators of MBCP outcomes and why they need to be triangulated with alternative indicators.

Program completion

Program completion or retention has often been used as an indicator of “success” in MBCPs, particularly by funders and commissioners of research. However, while retention remains a key concern in MBCPs generally, merely completing a program cannot account for change in perpetrator behaviour (Westmarland et al., 2010).

Perpetrator self-report

Frequently, comparisons are made between perpetrators’ self-reported data “pre-program” and “post-program” as an indicator of the program’s outcomes (Day & Casey, 2010). The use of self-report alone to measure change is problematic and unreliable for MBCPs. There are often vast differences in self-reports by perpetrators and ratings by victims/survivors, which raise concerns about the accuracy and usefulness of perpetrator self-reported data. As one of the evaluators we interviewed noted:

One of the really big challenges is that men will consistently self-report a huge amount of [positive] change, but their partners will not. So, what feels like a large change for them isn’t necessarily seen the same way by those affected by their violence. Given these issues, all measurements of change and safety need to be made through family [victim/survivor] contact.

The victim/survivor should always be given the opportunity to provide feedback regarding her own experiences of safety and wellbeing rather than relying on perpetrator self-reports of perceived change.

Recidivism data

Recidivism data is not considered a good indicator of victims/survivors’ and children’s safety as it fails to capture perpetrators’ use of a range of tactics to control women and their children. These tactics include emotional and financial controls and sabotaging women’s relationships with their support networks, all of which MBCPs ultimately aim to reduce (Vlais & Green, 2018; Walby et al., 2017). Recidivism is also not an accurate measure of attitudinal change (Kelly & Westmarland, 2015), as recidivism data might measure non-MBCP-related behaviour change or miss non-criminal forms of DFV. Yet many published evaluations of MBCPs have relied on recidivism data as the main outcome measures for program effectiveness, including data on re-offending (both DFV-specific and general criminological re-offending) and the time period to next re-offence. This reliance on recidivism data reflects the predominance

of corrections programs in the published MBCP literature (particularly in the United States), where victim-centred data has not been collected. While recidivism measures can in some circumstances have a role as part of a broader suite of measures, it is not recommended that they be sole measures of MBCP success.

Choosing data collection methods to answer the evaluation questions

Choosing data collection methods and overall study design is a key part of planning an evaluation, as it is essential to the collection of high-quality data, and to better enable the evaluator to infer a link between the activities of the MBCP and its outcomes. The choice of data collection method and evaluation design will be influenced by the evaluation questions and indicators selected, the intended use of evaluation findings, and users' confidence in the findings, as well as other contributing factors such as resources available for evaluation.

Some examples of quantitative and qualitative data collection methods and their advantages and disadvantages are shown in Table 6. Figure 7 provides examples of possible data collection methods for the indicators we outlined in Figure 6.



TABLE 6:

Advantages and disadvantages of various data collection methods

Method	Advantages	Disadvantages
Quantitative methods		
Questionnaire/ survey questions asking closed- ended questions (e.g. "how many times in the last week" ...); or ratings (e.g. rate from "very bad" to "very good"); or using psychometric tests	<ul style="list-style-type: none"> • Can be anonymous • Inexpensive and time-efficient • Can use validated measures that already exist • Can yield large sample sizes and substantial data to build evidence base 	<ul style="list-style-type: none"> • Provide information on "how much" but not "how" or "why" • Issues with partial completion • Can be an impersonal way to collect sensitive personal information • May need assistance of statistician or expert to perform the analysis • Some psychometric measures (i.e. validated questionnaires) need to be administered by a qualified person and there can be a cost for access • Need to ensure safety of victim/survivor respondents (e.g. might be risky to send via email or post, making distribution and collection more difficult)
Interviews with closed-ended questions (e.g. asking questions beginning "how many times ...")	<ul style="list-style-type: none"> • Face-to-face engagement with participants can facilitate rapport-building and result in more accurate data • Can provide explanation and clarification of questions to improve data accuracy and richness • May be more suitable than written methods for gathering sensitive information, as in an MBCP evaluation 	<ul style="list-style-type: none"> • Expensive • Time-intensive • Limited number of interviews you can conduct and therefore how much data can be collected
Observations using ratings, scores, checklists	<ul style="list-style-type: none"> • Can record program operation in real time and make necessary changes • Can reduce response biases of self-reports through observation by an independent third party • Direct observation can result in more accurate data and provide useful context 	<ul style="list-style-type: none"> • Checklists, ratings and scores may not capture important contextual information • Being observed can influence behaviours of program facilitators and participants • Expensive • If using multiple observers, will require training and guidance to yield reliable data (i.e. to achieve inter-rater reliability or consistency of ratings across observers)
Collation of administrative data (e.g. attendance records, recidivism data)	<ul style="list-style-type: none"> • Uses standard data and therefore does not place additional burden on program facilitators and administrators or evaluators • May have been collected over a long period of time, so could be used to test effects of new programs or changes to programs by looking at the data before and after 	<ul style="list-style-type: none"> • Gaining ethical approval to use administrative data can be difficult due to privacy concerns • Can be unclear how accurate such records are, as they are not collected by evaluators directly • Administrative data collected might not be useful for evaluation purposes, as it is recorded for other reasons

Method	Advantages	Disadvantages
Qualitative methods		
Questionnaires, surveys asking for comment, opinion or description (e.g. “please describe ...”)	<ul style="list-style-type: none"> • Can provide context to quantitative data or survey questions (e.g. “Please explain why you gave the rating above. What could we do to improve your rating?”) • Can be anonymous • Inexpensive • Can yield large sample which can then be analysed to identify common themes across responses 	<ul style="list-style-type: none"> • Can be impersonal, particularly if you are requesting sensitive information • Need to ensure safety of victim/survivor respondents (e.g. might be risky to send via email or post, making distribution and collection more difficult) • Wide variations in responses can be difficult to interpret • Can only provide minimal detail
Documentation review	<ul style="list-style-type: none"> • Comprehensive information (e.g. meeting minutes and facilitator manual) • Data can be retrieved from existing resources • Does not interrupt the program activities 	<ul style="list-style-type: none"> • Information may be incomplete or out-of-date • Restricted in terms of reviewing information that already exists (e.g. facilitator manuals might not be detailed enough to understand the activities of the program)
Individual interviews that ask open-ended questions	<ul style="list-style-type: none"> • Can gather in-depth information about “how” and “why” • Promotes engagement with the participant, which may yield better quality data than other methods and may be more suitable than written methods for gathering sensitive information • Can provide opportunity for victim/survivor to have her story heard, perhaps for the first time 	<ul style="list-style-type: none"> • Time-consuming to schedule, conduct and analyse • Difficult to analyse and draw causal conclusions from the data • Costly • May be prone to interviewer bias • Victim/survivor may be unable to disclose current risks, etc.
Case studies	<ul style="list-style-type: none"> • Can collect and comprehensively examine all aspects of participant experiences • Engaging data source through which to illustrate participant experiences to evaluation users • Allow for cross-comparison of cases 	<ul style="list-style-type: none"> • Time-consuming to collect and produce • Case studies might not be generalisable or transferable, especially since MBCPs widely differ
Observations that result in descriptions/ qualitative feedback	<ul style="list-style-type: none"> • Can record program operation in real time and make necessary changes • Can reduce response biases of self-reports through observation by an independent third party • Direct observation can result in more accurate data and provide useful context 	<ul style="list-style-type: none"> • Can be difficult to interpret and categorise observations • Being observed can influence behaviour of program facilitators and participants • Expensive
Focus groups	<ul style="list-style-type: none"> • Enable common issues and themes to be discussed • Efficient in terms of time and expense compared with individual interviews • Can provide a range of responses on key issues 	<ul style="list-style-type: none"> • Difficult and time-consuming to analyse responses • Requires a trained facilitator • Difficult to schedule due to large number of participants • Not suitable for the discussion of sensitive information • Confidentiality issues—may not be appropriate to ask about details of experiences of DFV in a group setting

FIGURE 7:

Possible data collection methods used for indicators outlined in Figure 6



	Process evaluation		Outcome evaluation		
Evaluation question	Did the training undertaken by facilitators improve their confidence to run the groups?	Did victims/survivors receive adequate feedback regarding their partners'/ ex-partners' participation in the program?	Are perpetrators more aware of the beliefs they hold to justify their use of violence than they were at program entry?	Are perpetrators implementing strategies taught in the program to choose non-violent and respectful behaviours?	Are victims/survivors safer than when the perpetrator entered the program?
Indicator	Change in scores from pre- to post-training on self-rated confidence to undertake an MBCP group	Victim/survivor satisfaction with feedback received	Ratings of level of awareness of beliefs pre- and post-program	Partner reports of current behaviours of perpetrator	Victim/survivor report of feelings of safety Police reports on violent behaviour
Possible data collection methods	Questionnaire completed before and after the training that includes a rating scale of confidence to run an MBCP group (quantitative)	Interview with victim/survivor in which she is asked to rate satisfaction from "not at all satisfied" to "completely satisfied" (quantitative), and to answer a question asking what additional feedback she would have found useful (qualitative)	Intake worker rating of perpetrator awareness of beliefs (quantitative) Facilitator rating of awareness of beliefs post-MBCP (quantitative)	Victim/survivor interview (qualitative)	Interviewer-administered questionnaire of victim-/survivor-rated perceptions of safety (quantitative) Document review of police reports (qualitative)

Validated scales

One of the data collection methods listed above is the use of psychometric tests or validated scales. A difficulty in evaluation is identifying appropriate measures to assess the impacts of an MBCP (Day et al., 2019). In our consultation with MBCP evaluators, there was consensus that both understanding of, and access to, appropriate outcome measures represented a significant roadblock to evaluation. Other issues such as time and funding constraints meant that validated outcomes measures to assess the impact of an MBCP were rarely used.

In Appendix B of the full research report, we have included a table that outlines a number of validated measures across the three key outcome domains of:

1. long-term changes in perpetrators' violent and controlling behaviour
2. adult victims/survivors' safety, wellbeing and freedom
3. children's safety, wellbeing and family functioning. (Adapted from Respect, 2017)

While there are few measures suitable to use alone for the purposes of MBCP evaluation, the table in Appendix B provides guidelines on which instruments, or combinations of instruments, may be valuable for use in an MBCP evaluation. Information is also provided on access, costs and/or administration requirements associated with each measure.

Most Significant Change

One lesser-known method of qualitative data collection worth noting is the Most Significant Change (MSC) technique. MSC can be especially useful for illustrating to different groups of stakeholders what "success" looks like from other perspectives, for uncovering organisational values and for identifying "unintended outcomes" that were not anticipated in the program logic (Davies & Dart, 2005). MSC is a useful addition to other techniques of measuring intermediate outcomes and feeding them back to stakeholders throughout the program.

Using the MSC technique involves asking those "in the field", such as facilitators of MBCPs, a question like "What was the most significant change that took place for the men in your program over the last month?" as well as why they think that change was "most significant". This question might be asked at regular intervals. The responses are then allocated to pre-determined "domains" or themes that the program was designed to assess (e.g. changes in understanding, parenting, or communication). Davies and Dart (2005) refer to a process of collating and summarising these responses through consecutive stages with different groups of people within an organisational line management structure.



This process requires substantial time and investment but may be suitable to consider for your evaluation. A guide to the MSC technique has been written by the creators of the technique and can be found here: <https://www.mande.co.uk/wp-content/uploads/2005/MSCGuide.pdf> (Davies & Dart, 2005).

An overarching consideration: Reducing the burden of data collection

A major challenge in MBCP evaluations is the additional burden placed on program participants, facilitators and others to provide additional data beyond that which the program provider routinely collects. Considerable data and risk-related information is already collected from program participants (perpetrators and victims/survivors) in the course of initial and ongoing risk assessment and case reviews. If not sensitively and carefully planned, introducing new data collection requirements on top of existing assessment processes can be onerous both for program participants and practitioners. One approach for an internally conducted evaluation is to embed data collection in a program's initial and ongoing assessment processes.

One example of how this can be done comes from Project Impact, arising from the Work with Perpetrators European Network (WWP-EN, 2019). Project Impact developed tools that served the dual purposes of initial assessment templates and program evaluation outcome measures. WWP-EN developed separate tools to use with perpetrators and their (ex-)partners that are administered at different time points (pre-program, mid-program, post-program and follow-up) and capture use/experience of coercive controlling violence, victim/survivor safety, hopes for the relationship, and wellbeing of children. Provided that the particular items and question wording were not altered, these tools could be merged into the program's customary initial assessment, risk assessment and program review templates. They could then be administered by program practitioners, so that the same items contribute towards both program clinical and program evaluation goals.

Selecting an appropriate and achievable study design

Types of study designs

A plan for measuring outcomes will need a description of the study design. There are many possible study designs, though relatively few have been used in formal MBCP evaluations due to the real-life constraints of conducting evaluation in this context. The type of design you choose will depend to some extent on how much you want to be able to determine "causality".

The extent to which you can attribute your findings to the MBCP and not to other external factors may be referred to as "causality". In outcome evaluation, it is important to discern whether your program activities led to the measured outcomes, rather than these outcomes being the result of other external factors. This is especially true as MBCPs take place in the context of other perpetrator interventions and integrated response systems. However, in process evaluation, you might just want to understand stakeholders' experiences of the program, rather than being concerned about changes produced. In this case, causality is not a primary concern.

Designs that are best at determining causality often involve randomly assigning participants to either an intervention or control group (e.g. randomised controlled trials). Randomly assigning participants to either group means that both groups are assumed to be "equivalent" (e.g. in terms of sociodemographic variables and relevant history). One group receives the MBCP (i.e. intervention group) and one does not

(i.e. control group). This allows the evaluator to conclude that any differences in outcomes between the two groups are due to the intervention (Tharp et al., 2011). However, for ethical and practical reasons, this design is rarely used in MBCP evaluation (see next section: Experimental rigour vs. ethical considerations).

Evaluation designs can vary in many ways, including whether participants are randomised to a control group, whether there is a control group, and at how many time points outcomes are measured. These design components can also be combined in different ways to create numerous variations of study design, each with a varying ability to determine causality. For example, a non-equivalent control group design is one involving a “natural” control group where participants are not randomly assigned to intervention or control, but outcomes are compared (see Box 3 for an example). Pre-test to post-test design compares outcomes before and after the intervention to determine if change has occurred. Sometimes a control group is used and compares pre–post changes between groups. Longitudinal designs measure outcomes before and after the intervention but also at one or more time points following completion (see Box 4.) One sub-type of longitudinal design is time series design, where outcomes are measured at multiple time points before and after the intervention.

There are other types of study design that do not focus on causality, for example cross-sectional designs in which a measurement is taken at a single point in time (e.g. a measure of attitudes toward women completed by perpetrators in the MBCP to determine the most common attitudes in MBCP participants). Cross-sectional designs might also assess the correlation between two concepts (e.g. scores on an attitudes toward women scale and [ex-]partners’ ratings of violence) at a single point in time. This type of correlational design is limited in its ability to determine that one construct “caused” the other because it doesn’t rule out other influences and cannot determine the direction of the relationship (i.e. is violence the result of poor attitudes or do poor attitudes develop as a way to justify violence?). However, these designs may also be useful if they assist in answering your evaluation questions, depending on what they are.

Examples of various study designs, and their relative advantages and disadvantages, are described in more detail in Table 7.



TABLE 7:

Examples of types of study design

Study design	Description/example	Strengths	Weaknesses
Randomised controlled trials (RCTs)	RCTs involve the random allocation of participants into intervention and control groups (i.e. some participants will partake in the MBCP and others will be allocated into a no-intervention “control” group). The two groups are compared on outcomes	<ul style="list-style-type: none"> • Considered the “gold standard” test of effectiveness • More likely to produce evidence that outcomes are due to the MBCP rather than other factors, such as maturation, demand characteristics, or other possible mechanisms of change which participants might be exposed to (Tharp et al., 2011) 	<ul style="list-style-type: none"> • Most resource-intensive • May not be ethically appropriate for use in DFV contexts, particularly given that random assignment may put women victims/survivors at risk of further harm and psychological distress (Arai et al., 2019) • May be impacted by a range of intersecting support systems (e.g. support for victims/survivors from other integrated response agencies) which may reduce the likelihood of “controlled” experimental contexts for MBCP evaluation (Howarth et al., 2019, p. 60)
Non-equivalent control groups design	Trials using a control group and intervention group where participants are not randomly assigned to either group (e.g. “control” participants in a different geographic location where MBCP will commence soon, but outcomes are measured before they attend the MBCP)	<ul style="list-style-type: none"> • An alternative when you are unable to randomly allocate participants • Enables comparison across groups and/or across time points 	<ul style="list-style-type: none"> • Differences between comparison groups may confound the results (e.g. could their different geographic location make them different to the intervention group in a way that will affect the outcomes?). See Box 3 for Project Mirabal (Kelly & Westmarland, 2015) • May be ethically inappropriate given that the “control group” will be men who use violence, or men who are on a “waiting list” for the MBCP and should participate in an MBCP as soon as possible • There are further ethical issues in allocating a “waiting list” of men for MBCPs as there may be immediate and long-term risks to women
Single group pre-test–post-test design	Also called “single-group” or “within subjects” design; involves collecting data in one group, who are compared before and after the MBCP	<ul style="list-style-type: none"> • Simple design • May be used when comparison groups are not available or ethically inappropriate 	<ul style="list-style-type: none"> • Limited ability to infer causality (i.e. that changes were the result of the MBCP)

Study design	Description/ example	Strengths	Weaknesses
Longitudinal design (see Box 4 for a broad example of longitudinal design)	Measures of outcomes are repeated over a long period of time (e.g. pre-MBCP, post-MBCP, 12 months follow-up and 24 months follow-up)	<ul style="list-style-type: none"> • Allows for measurement of long-term maintenance of change in attitudes and behaviours (i.e. sustained outcomes) following the MBCP • Could identify when positive outcomes are lost and therefore when repeat intervention might be needed 	<ul style="list-style-type: none"> • Access to perpetrators and (ex-) partners is very difficult following completion of the MBCP. This requires an ability to get in contact and the willingness of evaluation participants to stay involved • Time- and resource-intensive to track participants over time (see Box 5 for Brown, Flynn, Fernandez Arias, and Clavijo, 2016)
Case control	Case control (post-intervention only) retrospectively compares data between intervention and non-intervention groups (e.g. compare recidivism data for perpetrators who did and did not participate in an MBCP)	<ul style="list-style-type: none"> • May be used when baseline data is not available and for descriptive study • Might be able to use a large retrospective dataset to make comparisons between MBCP and non-MBCP groups 	<ul style="list-style-type: none"> • Limits to determining causality • Difficult to determine retrospectively what external factors might have caused differences in outcomes between those allocated and not allocated to MBCPs (e.g. were some perpetrators historically not considered “suitable” for MBCPs and are the reasons and the data supporting this still available?) • Consistent data would need to be available to allow collation of the dataset over time
Cross-sectional design	Measures a construct at a point in time in order to describe it (e.g. prevalence of particular types of controlling behaviour among men who enter MBCPs), or measures two or more constructs and correlates them (e.g. does holding particular attitudes relate to particular types of controlling behaviours?)	<ul style="list-style-type: none"> • Can be quick and easy as measures a single point in time • Can be useful describing how common a construct is (e.g. a particular attitude or type of controlling behaviour, victims'/ survivors' level of wellbeing at the beginning of perpetrators' entry into the MBCP) • Can be useful to generate further hypotheses for testing • Can collect information on multiple constructs at once (e.g. attitudes and behaviours) 	<ul style="list-style-type: none"> • Cannot determine causality nor directions of relationships between constructs

BOX 4:**Non-equivalent control group design: Domestic violence perpetrator programmes—Steps towards change (Kelly & Westmarland, 2015)**

Project Mirabal had a unique approach to providing a “non-intervention” comparison group (non-equivalent/non-randomly assigned control group) for comparison of outcomes between those attending and not attending a perpetrator program. The project attempted to compare outcomes from female (ex-) partners of men attending the perpetrator intervention with those whose partners were not attending because no program existed in the area. The women in the intervention and comparison groups were matched on “basic demographics, length of relationship and baseline levels of violence and abuse” (Kelly & Westmarland, 2015, p. 7).

The project illustrates one of the major challenges of a non-equivalent comparison group design: the two groups differed substantially in important ways. The comparison group were more likely to have children not in contact with their father, and they were far more likely to still be partnered with the perpetrator than the intervention group. These alone could explain differences in the presentation of outcomes and might reflect that the women were “at different points in the process of dealing with domestic violence” (Kelly & Westmarland, 2015, p. 8). Differences in life circumstances might also affect men’s motivations for change. Consequently, Kelly and Westmarland (2015, p. 8; see also Kelly et al., 2013) concluded that

whilst we do have comparison group data [...] the fact that they are not an equivalent comparison group rendered the comparative data difficult to interpret in a way where we could be sure of our explanations [...] If we had much higher numbers of men going through [domestic violence perpetrator programs] and higher numbers of research participants, it would have been possible to control for these differences. However studies of this nature do not tend to recruit the numbers that would have been required, and developing appropriate comparison or control groups unfortunately remains methodologically problematic.



BOX 5:**Longitudinal evaluation: European model for MBCP evaluations**

Lilley-Walker, Hester, and Turner (2018), in a review of the methodologies and measures used in European (and United Kingdom) evaluations of MBCPs, proposed a multi-point, longitudinal evaluation model with common elements of quality evaluations in the field. Based on their review of European MBCP evaluations, the authors concluded that studies need to

specify who exactly is participating, completing, and dropping out, at what point, and their motivations for doing so. Thus, careful attention must be paid to the types of information being collected—and also then reported—at different time points in order to better understand what and how behaviour and attitudes might change throughout the course of the programme. (Lilley-Walker et al., 2018, p. 880)

They recommend that the structure of a quality MBCP evaluation be summarised as follows:

T0 (pre-program)

- Size and type of sample at intake
- Referral routes and program pathways
- Excluded referrals and referral drop-outs

T3 (program end)

- Size and type of sample at end of program
- Measures
- Program completers

T1 (start of program)

- Size and type of sample at start of program
- Initial measures
- Excluded participants and drop-outs

T4 (follow-up)

- Measures
- Completers vs. non-completers

T2 (during program)

- Process/role and quality of facilitation and other program components
- Measures
- Drop-outs



BOX 6:**Challenges of longitudinal evaluation:**

A study of the impact on men and their partners in the short term and long term of attending MBCPs (Brown et al., 2016)

This well-known evaluation of MBCPs from 12 sites across Australia used a longitudinal design. This involved attempting to follow up with men participating in the program and their partners for 3 years, commencing at the start of the program to 2 years after the end, using surveys and interviews.

The evaluation report for this study illustrates one of the difficulties of longitudinal design: keeping in contact with participants over a long period and maintaining their participation in the evaluation. In the first wave of data collection on entry to the MBCP, there were 270 men in the study. The second wave at the end of the program attracted 110 men, but only after significant efforts to contact many of them. The third wave, 1 year later, included just 45 men, and their inclusion also required significant effort:

This number was gained after mail outs and emails, followed by phone calls to men [...] offers to pay for the completed surveys, further phone calls and further mail outs and emails. (Brown et al., 2016, p. 33)

Attempts to contact partners met with so little success that no quantitative data was able to be obtained.

Attempting to conduct any longitudinal evaluation should consider the significant effort and resources needed to ensure ongoing participation over a long period of time. Consideration should be given to what strategies can be used and therefore the resources required to encourage ongoing engagement in the evaluation. Furthermore, it is imperative that quality measures of more immediate outcomes are used, as discussed in the section “Measuring ultimate outcomes”.

Experimental rigour vs. ethical considerations

Most MBCP evaluations do not employ rigorous experimental design in the form of a randomised controlled trial (RCT).¹ While RCTs are considered the “gold standard” in evaluation, there is debate regarding the ethics and appropriateness of RCTs in MBCP contexts (Bender, 2017; Logan, Walker, Shannon, & Cole, 2008). Random assignment may put victims/survivors at risk of further harm and psychological distress because those in the control group do not immediately receive the intervention (Dutton et al., 2003). In addition, MBCPs take place in a broader system that influences perpetrator outcomes, significantly reducing the likelihood of “controlled” experimental contexts for MBCPs (Howarth et al., 2019, p. 60) as the MBCP is never delivered in isolation from other systemic interventions.

External factors affecting measured outcomes

Where it is not possible to use an evaluation design that can determine causality, and when causality is important, it can be useful to record factors external to the MBCP that might affect the measured outcomes. Such factors may include:

- Overall group differences: for example, one participant group may include a high number of high-risk, high-harm perpetrators, skewing the results because positive outcomes might be more difficult to achieve.

¹ RCTs are regarded as the most rigorous evaluation method because the random allocation of participants into different “treatment” groups (where some participants will partake in an intervention and others will be allocated into a no-intervention “control” group) provides evidence that the outcomes are due to the program itself, and not other possible factors.



- Individual differences: for example, participants of an MBCP group may be led by demand characteristics, such as a motivation to illustrate they have reduced their violent and controlling behaviour in order to see their children, but the change may not be maintained in the long-term.
- Outside factors: MBCPs likely represent one aspect in a range of multi-agency initiatives that aim to reduce violence against women. Men attending the program (and/or victims/survivors) may be receiving additional interventions such as supervision from a corrections officer. Similarly, victims/survivors may receive multiple services (e.g. safe housing, access to financial resources and legal advice) at the same time that their partner is completing the MBCP. These external services may influence the overall change recorded in pre–post MBCP measures.

Measurement of these “external” factors that might affect program outcomes can assist in better understanding the evaluation findings. For example, if the evaluation shows that participants with particular characteristics (e.g. men under 30 years of age) were most likely to drop out before program completion, then it might indicate, firstly, that a different approach is needed to maintain engagement of younger men, but also that the program outcomes might not be generalisable to younger men, since their intermediate outcomes (measured at program completion) were not measured.

Collective impact

Throughout most of the guide, we have referred specifically to measuring outcomes related to a single MBCP. However, it is worth noting a growing movement toward measuring “collective impact” (Kania & Kramer, 2011). Collective impact refers to “the commitment of a group of important actors from different sectors to a common agenda for solving a specific social problem” (Kania & Kramer, 2011, p. 36). Initiatives designed to have a collective impact require several key elements, not least of which are for participating organisations to have a common social agenda, a shared measurement framework and a shared plan of action (Kania & Kramer, 2011).

Collective impact initiatives for reducing the impacts of DFV within communities are increasing in Australia.² These initiatives require evaluation, and these evaluations are necessarily long-term and complex, as outcomes must be measured across the broad community they aim to target.

While the evidence base for collective impact as a strategy and the most accepted means of evaluating collective impact are still emerging, there are resources available if collective impact evaluation is something you want to know more about. A good place to start may be the Australian Institute of Family Studies’ webpage on evidence and evaluation for collective impact.³

2 See e.g. Cairns Collective Impact on Domestic and Family Violence <https://www.dvcairns.org/collective-impact-on-dv/>; Cardinia Shire Together We Can <http://togetherwecan.org.au/>; Logan Together <https://logantogogether.org.au/resources/>

3 See <https://aifs.gov.au/cfca/publications/collective-impact-evidence-and-implications-practice/evidence-and-evaluation>

Deciding who should conduct your evaluation

Evaluations vary widely in size and scope, and whether the evaluation is conducted by staff within the program or by an external evaluator will depend on a variety of factors like time, budget constraints, evaluation questions and study design, as well as ethical considerations. There are advantages and disadvantages to conducting both internal and external evaluations, and some of these are described in Table 9.

There is also the option to combine the internal and external evaluation components such that external evaluators are sub-contracted to conduct part of the evaluation. In one of the consultations for this guide, a program coordinator described employing an evaluator to observe and provide feedback on the facilitation of some groups. This is an example of when having an external evaluator would be most useful for providing an independent and objective view. Similarly, an external evaluator may be useful when particular expertise is required—for example, when a strong understanding of the cultural sensitivities of working with Aboriginal and Torres Strait Islander peoples is needed.

More information on internal and external evaluators can be found in the document, *A fundamental choice: Internal or external evaluation?* (Conley-Tyler, 2005). You can also complete the checklist to help make a decision about whether to do your evaluation internally or externally.⁴

An overarching consideration: Partnerships between program providers and evaluators

In some situations, working relationships between a program provider and an evaluator might be temporary or one-off. This can particularly be the case when the funder commissions an evaluation project for a new or existing program; the program provider might have no pre-existing relationship with the evaluator, and the working relationship ends once the evaluation is completed.

In other situations, opportunities can arise for program providers and evaluators to develop a longer-term partnership. While relevant applied research centres cannot fund program evaluations, industry and community partnerships and opportunities to conduct research that makes a difference can be highly valued.

These partnerships can help to strengthen a program provider's confidence about participating in evaluation and research activities. Although large agencies might have their own (still generally small) research unit and internal research ethics committee, in general, MBCP providers lack the knowledge and skills required to drive evaluation and research activity. Correspondingly, most applied research centres in the social or human sciences lack the subject matter expertise and understanding of some of the complexities and contentious issues involved in evaluating DFV services and programs. Partnership development—including researchers sitting in on MBCP group sessions and being involved in program planning and review activities—can help lay the foundation for quality evaluation activity once funding is in place to commence an evaluation process.

⁴ See https://www.aes.asn.au/images/stories/files/Publications/Vol4No1-2/fundamental_choice.pdf

TABLE 8:

Advantages and disadvantages of internal and external evaluators

Internal	External
+ive: Free or low-cost	-ive: Expensive
+ive: Evaluation can be implemented more quickly as facilitators and program managers are familiar with program and evaluation needs	-ive: Can require significant lead time before it begins. The evaluators need to be available, and then to spend some time understanding the program
+ive: Immediately useful as evaluators and users may be the same or in close proximity	-ive: The evaluator is not around in the longer term to ensure the results are utilised
+ive: Can capitalise on existing relationships to leverage participation of stakeholders in the evaluation (e.g. perpetrators, case workers)	-ive: It can be difficult for men who use violence and victims/survivors of DFV to trust in someone they don't know and for them to gain the willingness to talk to them for the purpose of evaluation
+ive: Facilitators as evaluators are already trained in avoiding collusion with perpetrators and are aware of the sensitivities of interviewing victims/survivors	-ive: Requires expertise in interviewing victims/survivors in ways that do not re-traumatise, and program participants in ways that avoid collusion with their violence-supporting narratives (see the section "Ethical considerations" for a more detailed discussion)
-ive: Participants in the evaluation (e.g. perpetrators) may be less likely to report negative thoughts or opinions of the program to internal evaluators in case they jeopardise their relationship with program staff	+ive: The evaluator's lack of involvement with the MBCP means stakeholders might be more likely to report things they don't like or that they don't think are working about the program. This can make the evaluation findings more useful
-ive: Participants can feel pressured or coerced into participating in an internal evaluation because of their existing relationships with those involved	+ive: Participants can feel less pressured or coerced into participating in the evaluation because they have no pre-existing relationship with the evaluator
-ive: Program staff might not have strong evaluation expertise	+ive: Evaluation consultants and organisations typically have a great deal of expertise and experience in a range of evaluation designs and methods. This means they might design a more comprehensive and useful evaluation, and that they might be able to carry out more complex evaluations than can be done by internal evaluators. These evaluation experts can also help to build the capacity of internal staff in the evaluation of their programs
-ive: Internal evaluators need to find extra time to do the evaluation. This might impact on the quality of the evaluation particularly if other demands are placed upon them	+ive: External evaluators are contracted to conduct the evaluation, and there would be less demand on the time of program staff to conduct evaluation activities
-ive: Users of the evaluation findings (e.g. program funders) may perceive internal evaluations as less trustworthy than external evaluations because they are seen as more likely to report favourable results (especially if there are funding decisions involved)	+ive: External evaluators are likely to be less invested in "positive findings" from the evaluation and are therefore likely to be more objective. They are also perceived as more objective by users, who therefore might see them to be more "credible" and thus have more faith in the findings

Note: "+ive" refers to positive and "-ive" refers to negative

Ethical considerations

Principles of ethical research

In Australia, all research (including evaluation activities) conducted with human participants must abide by the National Health and Medical Research Council's (NHMRC) *National Statement on Ethical Conduct in Human Research* (the Statement) (NHMRC, 2018). If you are planning to conduct an evaluation yourself, it is best that you read the Statement carefully, as it details how to adhere to a number of ethical research principles focused on participant protection.⁵

Broadly, the principles of ethical research involving humans are *research merit and integrity, justice, beneficence, and respect*.

In practice this means that, regardless of whether you are conducting your own evaluation or commissioning external evaluators, you must ensure that the evaluation design and processes address the following:

- how participants will be fairly and safely recruited and provided with enough information to give informed consent to participate. This will be especially true when involving women and children who are victims/survivors of DFV. Their participation must not place them at any additional risk
- how informed consent will be obtained
- what data will be generated or collected and how; how this data will be used and analysed, stored, disposed of and shared; risks associated with data collection; how the data collection adheres to the general ethical principles
- how the findings will be communicated to participants and to whom else they will be communicated and how, adhering to all the ethical principles
- how the findings will be communicated to a wider audience to ensure they contribute to broader knowledge in practice or the broader good
- how long the data will be retained after the project, how it will be stored to maintain confidentiality, and whether it will be available to others for future use.

Ethical guidelines specifically for evaluation

As well as these ethical guidelines for all research activities, the Australasian Evaluation Society (AES) has *Guidelines for the Ethical Conduct of Evaluations* (Australasian Evaluation Society, 2013). These guidelines encompass commissioning and preparing for an evaluation, conducting the evaluation, and reporting the results of the evaluation.⁶

Particular ethical issues to consider in evaluations of MBCPs

This section outlines some of the practical and ethical challenges evident in MBCP evaluation that may require special consideration.

⁵ See https://www.nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research-2007-updated-2018#toc_95

⁶ See https://www.aes.asn.au/images/stories/files/membership/AES_Guidelines_web_v2.pdf

Involving perpetrators in evaluation

An overarching ethical principle of engaging both victims/survivors and perpetrators in MBCP evaluation is that they are positioned as active agents and can make informed decisions about their participation (Downes, Kelly, & Westmarland, 2014).

It is important to acknowledge that perpetrator engagement with an MBCP may intersect, and be motivated by, outside pressures and expectations. This is particularly relevant if a perpetrator engages in an MBCP while child contact and custody disputes are ongoing or if they are affected by civil/criminal proceedings. Their participation in the MBCP may also be linked to their motivation to maintain or resume a relationship with an (ex-)partner (Downes et al., 2014).

Perpetrators may also be reluctant to be involved in research and/or evaluation and have reservations about speaking openly about their current circumstances and past actions. Those engaging with perpetrators need to recognise potential risks but also acknowledge their involvement as voluntary participants and as men and fathers capable of positive change (Downes et al., 2014).

There are also particular circumstances where engagement with perpetrators has potential to harm. For example, perpetrators may use the evaluation process to rationalise and justify their use of violence against their (ex-)partner (Hearn, Andersson, & Cowburn, 2007). The possibility of collusion with the narratives and thinking that perpetrators use to minimise, deny and rationalise their use of violence should be acknowledged, and interviewers need to be adequately skilled or supervised to counteract such instances in the evaluation process.

There are numerous other potential risks in the interview process that need to be observed and mitigated in MBCP evaluation, including:

- risk of inadvertently reinforcing the perpetrator's position as a victim, or what is known in the MBCP field as his "victim stance" (e.g. "I'm glad this evaluator understands my situation and how unfair she has been to me, because the group facilitators sure don't!")
- risks associated with the provision of a private one-to-one interview setting. Perpetrator engagement and disclosure in the interview process may mean they avoid disclosing information in the more transparent, accountability-based, group setting
- the possibility of a perpetrator disclosing information in the course of the evaluation suggesting potential risk of ongoing or future violence. This includes the associated risk he poses to victims/survivors that the program practitioners might not be aware of
- risks associated with the perpetrator's reflection of the evaluation process. For example, the interview may incite problematic thinking that the facilitators aren't aware of (e.g. "This interview has reinforced for me how hard I'm trying to change my behaviour, and how ungrateful she is for my efforts").

The significant majority of perpetrators involved in an evaluation or research process will not pose a physical safety risk to interviewers. However, it is important for program providers to alert the evaluator when any such risk might arise. Risk mitigation strategies can then be put into place (e.g. a program practitioner being

present at the interview). It is also good practice for evaluators to follow general precautions, such as never interviewing a perpetrator in a building alone and having ready access to a duress alarm.

Evaluators and the program provider need to have protocols in place so that if the evaluator believes that the interview with the perpetrator has either in itself escalated risk, or that new information relating to risk was revealed, they will be directed to confer with the program provider to determine appropriate action, which could include contacting the victim/survivor about safety planning.

It is important for program staff of the MBCP being evaluated to consider what these risks could be and how to mitigate them. One way to mitigate such risk is through appropriate training and supervision of evaluators.

The importance of training and expertise in MBCPs

It is well recognised that MBCP delivery requires specialist skills and experience, including a strong understanding of the intersections of gender, behavioural factors, safety and cultural aspects which impact and support violence against women and children. It is imperative that evaluators have a good knowledge of the complexities in MBCPs and are well equipped to identify any issues as the evaluation progresses. This is particularly important if evaluators are engaging in qualitative interviews and individual consultations with men who use violence.

One of the evaluators interviewed for the study noted that because data from victims/survivors are often very important, it's crucial that the evaluators conducting these interviews have sensitive and sufficiently advanced interviewing skills and understand DFV. When an evaluator conducts an interview with a victim/survivor, it may be the first lengthy face-to-face discussion that the victim/survivor has had with a "professional". She might have a lot to tell, and it's critical that the evaluator does not try to limit her voice or the narration of her story if the victim/survivor is finding that the telling of her story is helpful. Similarly, interviewers need to be able to conduct interviews in ways that don't re-traumatise victims/survivors.

When you commence the process of ethical approval, you will also likely be asked to consider whether members of the evaluation team have the necessary knowledge and skills to conduct the evaluation.

A number of peak bodies operating at the state and territory level implement specialist training in DFV. For example, there are training and short courses that offer a basic understanding of the dynamics of working with perpetrators of violence (e.g. they address concepts such as responding appropriately to perpetrator invitations to collude; identification of risk escalation; perpetrator accountability; responsibility and victim/survivor safety; basic micro-skills in perpetrator engagement), and more complex specialist training for experienced practitioners.

The importance of training may also depend on whether the MBCP evaluation is being conducted by internal staff or external evaluators, and the level of expertise in the evaluation team. Internal evaluations are likely to engage staff or expert practitioners who have knowledge of the MBCP field. Internal staff are also expected to have DFV-

specific training, supportive organisational policies and procedures, and available supervision from experienced DFV practitioners. However, given the complexities involved in data collection in MBCPs, external evaluators require specific training when engaging with perpetrators for evaluation purposes. For example, workshops addressing incidents of possible collusion will assist external evaluators to effectively identify and respond to male perpetrators by learning how to identify and resist such invitations.

Management and collection of data from victims/survivors

There are specific considerations when collecting and managing data from victims/survivors related to recruitment, formal written consent, and data collection that require significant planning and an ongoing awareness of risks. Victim/survivor safety needs to be prioritised in the process. At the same time, the range of data collection methods should also attend to victim/survivor agency and advocacy. This includes ethical practice in data collection to ensure the evaluation does not encourage any of the following issues.

Placing victims/survivors at risk of further violence from their (ex-)partners

Victims/survivors need to be informed about aspects of the evaluation in order to voluntarily participate. However, evaluators also need to ensure that their methods of communication are not putting victims/survivors at undue risk of further violence and abuse. For example, if a victim/survivor remains in the household or in contact with her abusive partner, it may not be safe to send information about the evaluation to her home address (Downes et al., 2014). Equally, if a victim/survivor is requested to participate in a telephone interview, there may be particular considerations regarding obtaining signed consent to participate (i.e. whether she feels safe to conduct the interview over the phone if the perpetrator is in the household) (Downes et al., 2014).

In the consultation, one MBCP evaluator made the following statement regarding coercion:

One of the issues for me with evaluation over the years has been about getting the opinion from the family about how the participant is doing, and in particular men taking accountability and women's safety. So, I think it is uniquely tricky in MBCPs. You also don't know the level of coercion that goes on. So how can you be sure that she is not just saying that he has changed or improved? So [are] the data authentic or is the victim/survivor being pressured to say things? And if she does offer an authentic account, is there going to be retaliation?

Given the multitude of issues that could potentially put a victim/survivor at risk, evaluators need to work with women's specialist support workers to ensure that they are appropriately briefed about expectations of engagement in the evaluation and that necessary safeguards are in place to mitigate risk of further abuse and violence. In the first instance, victims/survivors should be approached by support workers to see if they consent to their contact information being shared with the evaluators.

Importantly, victims/survivors should be given an opportunity to make decisions about when and how they would like to be contacted and receive information about the evaluation. Specialist workers engage in various forms of communication with victims/survivors, including text updates (to inform women of men's participation in the MBCP), phone support, and one-to-one contact with the women's support worker (Downes et al., 2014). Specialist women's workers are also crucial to evaluation activities with victims/survivors. They are able to appropriately advise evaluators how best to make contact, ensure voluntary consent to be involved, and minimise risks to safety.

Potentially re-traumatising women who do not have adequate support

Evaluators need to consider the risks for victims/survivors who engage in the evaluation. Victims/survivors may need to be assessed in relation to their *current* situation and vulnerability to take part, with recognition that their capacity and agency may vary greatly over the course of the evaluation. As a result, women need to be approached as individuals, and the potential for harm needs to be viewed as being located on a continuum that may change over time.

In the context of a qualitative interview, retelling and revisiting traumatic events and circumstances may have unanticipated consequences for victims/survivors. Adequate debriefing following an interview is important to ensure they have necessary supports in place. However, given the possibility that distress triggered in the interview may also be delayed for some victims/survivors, there is considerable evidence to suggest that victims/survivors should only be involved in the evaluation if they are currently receiving assistance from a support worker.

Risk of disempowering victims/survivors

Victims/survivors need to be informed that their participation in the evaluation is voluntary, and they should not be unduly pressured to take part. Evaluators need to work closely with women's support workers to ensure that victims/survivors adequately understand their role in the evaluation and can make an autonomous choice to participate.

One evaluator we interviewed for this guide emphasised that often the interviews are an opportunity for women to tell their story. Many victims/survivors positioned themselves as active agents in the evaluation and were open to being interviewed. The evaluation needs to consider the extent to which the interview process empowers victims/survivors. Given the context, where women often have little control over their lives due to perpetrators' use of violence and controlling behaviour, providing them with an opportunity to discuss their experience and to offer feedback about how the program might need to change is an opportunity for them to exert some agency.

Careful management and collection of data from women must be undertaken to ensure that evaluation is not at risk of disempowering victims/survivors, intruding on their lives and choices, misrepresenting their experiences or assuming they are incapable of making independent decisions (McDermott & Garofalo, 2004).

Equally importantly, evaluators are encouraged to view victims/survivors as active and autonomous agents. Adopting ethical practice that attends to women’s potential empowerment in the evaluation process, often through the “telling of their story”, is important. The design of data collection instruments—such as the interview schedule—needs to be carefully considered to ensure victims/survivors are also provided opportunities to reflect and experience positive impacts.

Ethical approval and working with vulnerable groups

Any evaluation involving “vulnerable” groups will require ethics approval from a human research ethics committee (HREC). These groups include perpetrators, victims/survivors, and their children. If you are working with an external, experienced evaluator, such as a university, this organisation will obtain its own ethical approval. However, if you are conducting the evaluation yourself, you will need to do so. Some community service organisations have their own ethics approval processes, and there are some organisations that accept external ethics applications. The NHMRC keeps a list of registered ethics committees by state.⁷

There are some types of evaluation that will not require ethics approval because the activity is considered to have “negligible risk” associated with it. These evaluations generally include “quality assurance/audit projects that do not involve access to or collection of private, sensitive or health data” (University of Melbourne, n.d.). This might include activities such as collecting participant feedback on group sessions. However, if you are uncertain whether ethics approval is needed, it is best to speak to a member of a human research ethics committee to clarify. “Secondary” use of data (i.e. analysing existing data) may still require ethics approval even though it involves no direct human contact. Whether ethics approval is required will depend on factors such as whether any individual can be identified within the data (e.g. participants in programs), and whether the data are being used for a purpose outside that to which the person providing the data has consented (Tripathy, 2013). Again, it is best to clarify these issues with an ethics committee to ensure that any evaluation activities are conducted ethically.



⁷ The list of registered ethics committees can be found at <https://www.nhmrc.gov.au/research-policy/ethics/human-research-ethics-committees>

Recommendations

Recommendations related to this guide

- To ensure this guidance is useful and meaningful to the work of MBCP providers we recommend a series of steps that positions the outcome of this project as a first iteration of an ongoing piece of work. This may include, for example:
 - knowledge exchange and learning activities, such as a national webinar that introduces the guide and engages in dialogue about its content with a range of potential end-users, including those with experience in design and evaluation of MBCPs (internal and external program providers), as well as evaluators with less experience with MBCP evaluation
 - active promotion of the Evaluation guide by harnessing the networks of the peak bodies, No to Violence and Stopping Family Violence, and the Services and Practitioners for the Elimination of Abuse Queensland (SPEAQ) network. These organisations are well positioned to promote the use of the guide among their networks of MBCP providers
 - an evaluation of the uptake of the guide in practice using a pilot study design with a select group of MBCP providers.
- In terms of future work, we recommend broadening the current guide to include how to scope an evaluation for a range of perpetrator interventions and behavioural change programs. While this project limited its scope to MBCPs, the considerations and guidance required to support an evaluation for the range of innovative and emerging perpetrator interventions is likely to be similar to that required to support a standard MBCP evaluation.
- Despite attempts to include real-world examples of evaluations of programs that have been specifically designed for diverse communities (e.g. programs for perpetrators from LGBTIQ+ communities and CALD groups), we were unable to do so. We recommend, therefore, that future iterations include broader consultation to include information relevant to evaluation of these programs.
- We recommend development of a specific guide on conducting evaluation of interventions that involve children and young people who use violence.

- While the broad framework of this guide has applicability for monitoring and evaluation activities across the whole of the perpetrator intervention system, our intended purpose was to guide approaches to evaluation at the programmatic level. We recommend, however, that this guide be used as a platform to develop a dedicated approach to support evaluation activity conducted at a systems level. Recent and upcoming publications on perpetrator intervention systems provide further guidance (see e.g. Chung et al., in press; Vlasis, Campbell, & Green, 2019).

Broader recommendations for policy and practice

- This guide does not attempt to inform or support providers conducting evaluations of MBCPs or other perpetrator interventions specifically provided by and for Aboriginal and Torres Strait Islander men and communities. Rather, we recommend that evaluation of any MBCP program by Aboriginal and Torres Strait Islander community-controlled organisations be led by those organisations.
- Process evaluations are critical to determine if an MBCP is being implemented as intended and is “evaluation-ready” when an opportunity arises to assess the program’s impact and outcomes. We recommend that state and territory governments fund MBCP provider peak bodies to support program providers to conduct periodic process evaluations, even if these providers do not have the capacity to conduct or commission impact and outcome evaluations.
- Although the scope of this work did not extend to exploring in detail what “success” looks like in terms of outcomes from MBCPs, we note that this remains a highly contested space. There is significant disagreement and variability in terms of how outcomes are conceptualised and measured in MBCP evaluations, limiting establishment of a comprehensive evidence base on “what works”. We therefore recommend that the Commonwealth Government commission work to develop a national outcomes framework for MBCPs and perpetrator intervention programs.

Conclusion

The primary aim of this project was to develop an easy-to-understand guide to scoping an evaluation for personnel involved in the implementation of MBCPs in Australia. The purpose of this guide is to improve the technical evaluation expertise of these personnel to better equip them to scope an evaluation in order to conduct their own in-house evaluation activities or to commission a quality external evaluation. The processes undertaken to develop this guide highlight a number of key considerations regarding the state of MBCP evaluation in Australia and, to some degree, internationally. These key points have consequences for the ongoing evaluation of perpetrator interventions.

The review of state and territory minimum standards for implementation of MBCPs highlighted that the perpetrator intervention area within most states is in flux, with most standards currently under review or with new standards recently implemented or awaiting formal approvals. This means that the minimum requirements for delivery of some MBCPs will soon change. These changes will have consequences for evaluation, including potential changes to minimum standards. While such standards differ between states, they are uniform in prioritising the freedom and safety of victims/survivors, including children. This focus suggests the need for victim/survivor freedom and safety to be a key outcome addressed in all MBCP evaluations.

We reviewed the usability and psychometric properties of established outcome measures commonly used in MBCP evaluations in Australia and internationally. This review uncovered a dearth of validated outcomes measures suitable for use in current evaluations of MBCPs in Australia. Problems with the identified measures include a lack of validation with target users (e.g. women who are victims/survivors of DFV), impracticality relating to the large number of items, failure to measure key constructs within outcome domains (e.g. failure to measure controlling behaviours in addition to physical violence), and outmoded terms and language (e.g. “battering”). Such measures are needed to improve the quality of MBCP evaluation and to improve replicability of outcomes measurement in order to provide a broader evidence base. Development of more suitable measures would be a valuable addition to MBCP evaluation.

In order to develop the Evaluation guide, we consulted with a review panel, a group of experts currently implementing MBCPs in Australia, and evaluators of current Australian MBCPs. The complexities of undertaking MBCPs were highlighted throughout this process, particularly ethical considerations. While we have tried to provide some guidance on many of these in our guide, the constraints of this project mean that primarily these issues have been “flagged” for further consideration by evaluators. Each of these complexities, such as processes for safely involving victims/survivors in MBCP evaluation and the ethical ramifications of involving perpetrators in evaluation, require more consideration. Further work on such issues would also be useful for improving the quality of MBCP evaluations.

Limitations notwithstanding, our Evaluation guide does provide concise and accessible information for those personnel scoping the evaluation of an Australian MBCP. In addition to outlining general principles of quality evaluation, it also provides real-life examples of MBCPs conducted nationally and internationally to highlight key points and make the information specifically relevant to MBCP evaluation. We have recommended that further work be done to pilot this guide for different user groups and that ongoing revisions be undertaken to ensure that this guide provides the most usable and useful information.

References

- Abidin, R. R. (1990). *Parenting stress index–Short form*. Charlottesville, VA: Pediatric Psychology Press.
- Aldarondo, E., & Malhotra, K. (2014). Domestic violence: What every multicultural clinician should know. In M. Miville & A. Ferguson (Eds.), *Handbook of race-ethnicity and gender in psychology* (pp. 379–403). New York: Springer.
- Andrews, D., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior*, 17(1), 19–52. <https://doi.org/10.1177/0093854890017001004>
- Arai, L., Heawood, A., Feder, G., Howarth, E., MacMillan, H., Moore, T. H., . . . Gregory, A. (2019). Hope, agency, and the lived experience of violence: A qualitative systematic review of children’s perspectives on domestic violence and abuse. *Trauma, Violence, & Abuse*, 20(3), 1–12. <https://doi.org/10.1177/1524838019849582>
- Arnold, D. S., O’Leary, S. G., Wolff, L. S., & Acker, M. M. (1993). The parenting scale: A measure of dysfunctional parenting in discipline situations. *Psychological Assessment*, 5(2), 137–144. <https://doi.org/10.1037/1040-3590.5.2.137>
- Australia. Department of Social Services. (2016). *National outcome standards for perpetrator interventions: Baseline report 2015–2016*. Canberra: DSS.
- Australia. Department of Social Services. (2019). *Fourth action plan 2019–2022: National Plan to Reduce Violence against Women and their Children 2010–2022*. Canberra: DSS.
- Australian Institute of Family Studies. (2017). *Collective impact: Evidence and implications for practice*. Retrieved 21 February, 2020, from <https://aifs.gov.au/cfca/publications/collective-impact-evidence-and-implications-practice/evidence-and-evaluation>
- Australasian Evaluation Society. (2013). *Guidelines for the ethical conduct of evaluations*. Retrieved from https://www.aes.asn.au/images/stories/files/membership/AES_Guidelines_web_v2.pdf
- Axford, N., Elliott, D. S., & Little, M. (2012). Blueprints for Europe: Promoting evidence-based programmes in children’s services. *Psychosocial Intervention*, 21(2), 205–214. <https://doi.org/10.5093/in2012a11>
- Babcock, J. C., Costa, D. M., Green, C. E., & Eckhardt, C. I. (2004). What situations induce intimate partner violence? A reliability and validity study of the proximal antecedents to violent episodes (PAVE) scale. *Journal of Family Psychology*, 18(3), 433–442. <https://doi.org/10.1037/0893-3200.18.3.433>
- Barnett, O. W., Martinez, T. E., & Keyson, M. (1996). The relationship between violence, social support, and self-blame in battered women. *Journal of Interpersonal Violence*, 11(2), 221–233. <https://doi.org/10.1177/088626096011002006>
- Bates, E. A., Graham-Kevan, N., Bolam, L. T., & Thornton, A. J. V. (2017). A review of domestic violence perpetrator programs in the United Kingdom. *Partner Abuse*, 8(1), 3–46. <https://doi.org/10.1891/1946-6560.8.1.3>
- Beck, C. J. A., Walsh, M. E., & Weston, R. (2009). Analysis of mediation agreements of families reporting specific types of intimate partner abuse. *Family Court Review*, 47(3), 401–415. <https://doi.org/10.1111/j.1744-1617.2009.01264.x>
- Begun, A., Brondino, M. J., Bolt, D., Weinstein, B., Strodthoff, T., & Shelley, G. (2009). The revised safe at home instrument for assessing readiness to change intimate partner violence. In C. M. Murphy & R. D. Maiuro (Eds.), *Motivational interviewing and stages of change in intimate partner violence* (pp. 199–223). New York: Springer Publishing Company.
- Begun, A. L., Murphy, C., Bolt, D., Weinstein, B., Strodthoff, T., Short, L., & Shelley, G. (2003). Characteristics of the safe at home instrument for assessing readiness to change intimate partner violence. *Research on Social Work Practice*, 13(1), 80–107. <https://doi.org/10.1891/0886-6708.23.4.508>
- Belfrage, H., Strand, S., Storey, J. E., Gibas, A. L., Kropp, P. R., & Hart, S. D. (2012). Assessment and management of risk for intimate partner violence by police officers using the spousal assault risk assessment guide. *Law and Human Behavior*, 36(1), 60–67. <https://doi.org/10.1037/h0093948>

- Bender, A. K. (2017). Ethics, methods, and measures in intimate partner violence research: The current state of the field. *Violence Against Women*, 23(11), 1382–1413. <https://doi.org/10.1177/107780121665897>
- Better Evaluation. (2016a). Decide purpose. Retrieved from https://www.betterevaluation.org/en/rainbow_framework/frame/decide_purpose
- Better Evaluation. (2016b). Specify the key evaluation questions (KEQs). Retrieved from https://www.betterevaluation.org/en/rainbow_framework/frame/specify_key_evaluation_questions
- Beydoun, H. A., Beydoun, M. A., Kaufman, J. S., Lo, B., & Zonderman, A. B. (2012). Intimate partner violence against adult women and its association with major depressive disorder, depressive symptoms and postpartum depression: A systematic review and meta-analysis. *Social Science & Medicine*, 75(6), 959–975. <https://doi.org/10.1016/j.socscimed.2012.04.025>
- Birkley, E. L., & Eckhardt, C. I. (2015). Anger, hostility, internalizing negative emotions, and intimate partner violence perpetration: A meta-analytic review. *Clinical Psychology Review*, 37(4), 40–56. <https://doi.org/10.1016/j.cpr.2015.01.002>
- Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *BMJ*, 314(7080), 572. <https://doi.org/10.1136/bmj.314.7080.572>
- Blatch, C., O'Sullivan, K., Delaney, J. J., van Doorn, G., & Sweller, T. (2016). Evaluation of an Australian domestic abuse program for offending males. *Journal of Aggression, Conflict and Peace Research*, 8(1), 4–20. <https://doi.org/10.1108/JACPR-10-2015-0194>
- Bloom, S. S. (2008). *Violence against women and girls: A compendium of monitoring and evaluation indicators*. Chapel Hill, NC: MEASURE Evaluation.
- Bonomi, A. E., Thompson, R., Anderson, M., Rivara, F. P., Holt, V. L., Carrell, D., & Martin, D. P. (2006). Ascertainment of intimate partner violence using two abuse measurement frameworks. *Injury Prevention*, 12(2), 121–124. <https://doi.org/10.1136/ip.2005.009563>
- Brown, T., Flynn, C., Fernandez Arias, P., & Clavijo, C. (2016). *A study of the impact on men and their partners in the short term and in the long term of attending men's behaviour change programs*. Retrieved from <https://violencefreefamilies.org.au/longitudinal-study-effectiveness-mens-behaviour-change-programs/>
- Bruce, N., Pope, D., & Stanistreet, D. (2018). *Quantitative methods for health research: A practical interactive guide to epidemiology and statistics*. West Sussex: John Wiley & Sons, Ltd.
- Buhrmester, D., Furman, W., Wittenberg, M. T., & Reis, H. T. (1988). Five domains of interpersonal competence in peer relationships. *Journal of Personality & Social Psychology*, 55(6), 991–1008. <https://doi.org/10.1037/0022-3514.55.6.991>
- Burgess, S. E., & Tavakoli, A. (2005). Psychometric assessment of the Burgess–Partner abuse scale for teens (B-PAST). *Revista Aquichan*, 5(1), 96–107.
- Buss, A., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology*, 63(3), 452–459. <https://doi.org/10.1037/0022-3514.63.3.452>
- Cairns Regional Domestic Violence Service. (n.d.). Cairns collective impact on domestic and family violence. Retrieved 21 February, 2020 from <https://www.dvcairns.org/collective-impact-on-dv>
- Campbell, J. C. (1986). Nursing assessment for risk of homicide with battered women. *Advances in Nursing Science*, 8(4), 36–51. <https://doi.org/10.1097/00012272-198607000-00006>
- Campbell, J. C. (2004). *Danger assessment* (Measurement instrument). Retrieved from <https://dojmt.gov/wp-content/uploads/2011/05/dangerassessment.pdf>
- Campbell, J. C., & Soeken, K. L. (1999). Women's responses to battering: A test of the model. *Research in Nursing & Health*, 22(1), 49–58. [https://doi.org/10.1002/\(SICI\)1098-240X\(199902\)22:1<49::AID-NUR6>3.0.CO;2-F](https://doi.org/10.1002/(SICI)1098-240X(199902)22:1<49::AID-NUR6>3.0.CO;2-F)

- Campbell, J. C., Webster, D., Koziol-McLain, J., Block, C., Campbell, D., Curry, M. A., ... Laughon, K. (2003). Risk factors for femicide in abusive relationships: Results from a multisite case control study. *American Journal of Public Health*, 93(7), 1089–1097. <https://doi.org/10.2105/AJPH.93.7.1089>
- Campis, L. K., Lyman, R. D., & Prentice-Dunn, S. (1986). The parental locus of control scale: Development and validation. *Journal of Clinical Child Psychology*, 15(3), 260–267. https://doi.org/10.1207/s15374424jccp1503_10
- Canales, O. G., Geldschläger, H., Nax, D., & Ponce, Á. (2015). European perpetrator programmes: A survey on day-to-day outcome measurement. *Studia Humanistyczne AGH*, 14(2), 33–52. <https://doi.org/10.7494/human.2015.14.2.33>
- Cardinia. (n.d.). Preventing family violence: Together we can. Retrieved 21 February, 2020 from <http://togetherwecan.org.au/>
- Cascardi, M., & O’Leary, K. D. (1992). Depressive symptomatology, self-esteem, and self-blame in battered women. *Journal of Family Violence*, 7(4), 249–259. <https://doi.org/10.1007/BF00994617>
- Catch-22. (n.d.). *Evidence is confidence: How to create a richer evidentiary tapestry*. Retrieved from <https://cdn.catch-22.org.uk/wp-content/uploads/2015/11/Realising-Ambition-Programme-Insight-Evidence-is-Confidence.pdf>
- Chung, D. (2014). Behavior change programs for intimate partner violence abusers: A means to promote the safety of women and children? In H. Johnson, B. S. Fisher, & V. Jaquier (Eds.), *Critical issues on violence against women: International perspectives and promising strategies* (1st ed.) (pp. 171–182). London: Routledge.
- Chung, D., Upton-Davis, K., Cordier, R., Campbell, E., Wong, T., Salter, M., ... Bisset, T. (in press). *Improved accountability: The role of perpetrator intervention systems* (Research report). Sydney: Australia’s National Research Organisation for Women’s Safety.
- Cicchetti, D., & Doyle, C. (2016). Child maltreatment, attachment and psychopathology: Mediating relations. *World Psychiatry*, 15(2), 89–90. <https://doi.org/10.1002/wps.20337>
- CoastAdapt. (2017). Identifying indicators for monitoring and evaluation. Retrieved from <https://coastadapt.com.au/how-to-pages/identifying-indicators-monitoring-and-evaluation>
- Cobb, A. R., Tedeschi, R. G., Calhoun, L. G., & Cann, A. (2006). Correlates of posttraumatic growth in survivors of intimate partner violence. *Journal of Traumatic Stress*, 19(6), 895–903. <https://doi.org/10.1002/jts.20171>
- Cohen, P. J., Glaser, B. A., Calhoun, G. B., Bradshaw, C. P., & Petrocelli, J. V. (2005). Examining readiness for change: A preliminary evaluation of the University of Rhode Island change assessment with incarcerated adolescents. *Measurement and Evaluation in Counseling and Development*, 38(1), 45–62. <https://doi.org/10.1080/07481756.2005.11909768>
- Collett, B. R., Gimpel, G. A., Greenson, J. N., & Gunderson, T. L. (2001). Assessment of discipline styles among parents of preschool through school-age children. *Journal of Psychopathology and Behavioral Assessment*, 23(3), 163–170. <https://doi.org/10.1023/A:1010965220517>
- Colley, D., Hall, R., Jenkins, A., & Anderson, G. (1997). *Competency standards for intervention workers—Working with men who perpetrate domestic abuse and violence for standards that promote best practice for intervention*. Adelaide: Department of Human Services, South Australia.
- Conley-Tyler, M. (2005). A fundamental choice: Internal or external evaluation? *Evaluation Journal of Australasia*, 4(1–2), 3–11. <https://doi.org/10.1177%2F1035719X05004001-202>
- Corvo, K., Dutton, D., & Chen, W.-Y. (2008). Toward evidence-based practice with domestic violence perpetrators. *Journal of Aggression, Maltreatment & Trauma*, 16(2), 111–130. <https://doi.org/10.1080/10926770801921246>
- Council of Australian Governments. (2015). *National outcome standards for perpetrator interventions*. Canberra: DSS.

- Council of Australian Governments. (2016). *COAG advisory panel on reducing violence against women and their children—Final report*. Canberra: Department of the Prime Minister and Cabinet.
- Cummins, R., & Lau, A. (2005). *Personal Wellbeing Index—School Children (PWI-SC) 3rd edition*. Retrieved from <http://www.acqol.com.au/uploads/pwi-sc/pwi-sc-english.pdf>
- Cutler, B. L. (2008). *Encyclopedia of psychology and law* (Vols. 1 & 2). Thousand Oaks: Sage Publications.
- Dainbrot, F. H., Papp, M. E., & Whitmore, C. (1984). The sex-role attitudes of three generations of women. *Personality and Social Psychiatry Bulletin*, 10(3), 469–473. <https://doi.org/10.1177/0146167284103015>
- Daugherty, C. G., & Dambrot, F. H. (1986). Reliability of the attitudes toward women scale. *Educational and Psychological Measurement*, 46(2), 449–453. <https://doi.org/10.1177/001316448604600221>
- Davies, R. (2015). Evaluability assessment. Retrieved 21 February, 2020 from http://betterevaluation.org/themes/evaluability_assessment
- Davies, R., & Dart, J. (2005). *The “most significant change”(MSC) technique: A guide to its use*. Retrieved from https://www.researchgate.net/publication/275409002_The_'Most_Significant_Change'_MSC_Technique_A_Guide_to_Its_Use
- Day, A., & Casey, S. (2010). Maintaining change programmes for offenders: Some suggestions for practice. *Psychology, Crime & Law*, 16(6), 449–458. <https://doi.org/10.1080/10683160902798086>
- Day, A., Vlasis, R., Chung, D., & Green, D. (2018). Standards of practice in domestic and family violence behaviour change programs in Australia and New Zealand. *Australian and New Zealand Journal of Family Therapy*, 39(4), 501–513. <https://doi.org/10.1002/anzf.1332>
- Day, A., Vlasis, R., Chung, D., & Green, D. J. (2019). *Evaluation readiness, program quality and outcomes in men's behaviour change programs* (Research report, 01/2019). Sydney: ANROWS.
- de Vogel, V. (2005). *Structured risk assessment of (sexual) violence in forensic clinical practice: The HCR-20 and SVR-20 in Dutch forensic psychiatric patients*. Amsterdam: Dutch University Press.
- DeKeseredy, W. S. (2011). *Violence against women: Myths, facts, controversies*. Toronto: University of Toronto Press.
- DeKeseredy, W. S., & Schwartz, M. D. (1998). *Measuring the extent of woman abuse in intimate heterosexual relationships: A critique of the conflict tactics scales*. Retrieved from https://vawnet.org/sites/default/files/materials/files/2016-09/AR_Ctscriit_0.pdf
- Devries, K. M., Mak, J. Y. T., García-Moreno, C., Petzold, M., Child, J. C., Falder, G., . . . Watts, C. H. (2013). The global prevalence of intimate partner violence against women. *Science*, 340(6140), 1527–1528. <https://doi.org/10.1126/science.1240937>
- Diemer, K., Humphreys, C., Fogden, L., Gallant, D., Spiteri-Staines, A., Bornemisza, A., & Vercoe, E. (2020). *Caring Dads program: Helping fathers value their children: Three site independent evaluation 2017–2019* (Final report). Retrieved from https://mcusercontent.com/735c980d83254d8db51ef91e3/files/294308b1-1756-43c2-88d4-2a2eac8e4e05/2020_Caring_Dads_Final_Report_3_3_2020.pdf?mc_cid=979b3fcc6e&mc_eid=1737a9d628
- Dillon, G., Hussain, R., Loxton, D., & Rahman, S. (2013). Mental and physical health and intimate partner violence against women: A review of the literature. *International Journal of Family Medicine*, 2013(313909), 1–15. <https://doi.org/10.1155/2013/313909>
- Dobash, R. P., Dobash, R. E., Cavanagh, K., & Lewis, R. (2000). Confronting violent men. In C. Itzin & J. Hanmer (Eds.), *Home truths about domestic violence: Feminist influences on policy and practice: A reader* (p. 289). New York: Routledge.
- Dobash, R. P., Dobash, R. E., Wilson, M., & Daly, M. (1992). The myth of sexual symmetry in marital violence. *Social Problems*, 39(1), 71–91. <https://doi.org/10.2307/3096914>
- Douglas, K. S., Hart, S. D., Webster, C. D., & Belfrage, H. (2013). *CR-20 (Version 3): Assessing risk of violence—User guide*. Burnaby, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.

- Douglas, K. S., & Reeves, K. A. (2010). The HCR-20 violence risk assessment scheme: Overview and review of the research. In R. K. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment* (pp. 147–186). New York: Routledge.
- Douglas, K. S., Webster, C. D., Hart, S. D., Eaves, D., & Ogloff, J. R. P. (Eds.). (2001). *HCR20: Violence risk management companion guide*. Burnaby, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.
- Downes, J., Kelly, L., & Westmarland, N. (2014). Ethics in violence and abuse research: A positive empowerment approach. *Sociological Research Online*, 19(1), 1–13. <https://doi.org/10.5153/sro.3140>
- Dutton, D. G. (1988). *The domestic assault of women: Psychological and criminal justice perspectives*. Boston: Allyn & Bacon.
- Dutton, D. G., & Corvo, K. (2006). Transforming a flawed policy: A call to revive psychology and science in domestic violence research and practice. *Aggression and Violent Behavior*, 11(5), 457–483. <https://doi.org/10.1016/j.avb.2006.01.007>
- Dutton, M. A., Holtzworth-Munroe, A., Jouriles, E., McDonald, R., Krishnan, S., McFarlane, J., & Sullivan, C. (2003). *Recruitment and retention in intimate partner violence research*. Rockville, MD: US Department of Justice.
- Dwyer, D. (1999). Measuring domestic violence. *Journal of Offender Rehabilitation*, 29(1–2), 23–33. https://doi.org/10.1300/J076v29n01_02
- Eckhardt, C. I., Babcock, J., & Homack, S. (2004). Partner assaultive men and the stages and processes of change. *Journal of Family Violence*, 19(2), 81–93. <https://doi.org/10.1023/B:JOFV.0000019839.98858.5c>
- Eckhardt, C., Holtzworth-Munroe, A., Norlander, B., Sibley, A., & Cahill, M. (2008). Readiness to change, partner violence subtypes, and treatment outcomes among men in treatment for partner assault. *Violence and Victims*, 23(4), 446–475. <https://doi.org/10.1891/0886-6708.23.4.446>
- Eckhardt, C. I., Murphy, C., Black, D., & Suhr, L. (2006). Intervention programs for perpetrators of intimate partner violence: Conclusions from a clinical research perspective. *Public Health Reports*, 121(4), 369–381. <https://doi.org/10.1177/003335490612100405>
- Eckhardt, C. I., Murphy, C. M., Whitaker, D. J., Sprunger, J., Dykstra, R., & Woodard, K. (2013). The effectiveness of intervention programs for perpetrators and victims of intimate partner violence. *Partner Abuse*, 4(2), 196–231. <https://doi.org/10.1891/1946-6560.4.2.196>
- Eckhardt, C. I., & Utschig, A. C. (2007). Assessing readiness to change among perpetrators of intimate partner violence: Analysis of two self-report measures. *Journal of Family Violence*, 22(5), 319–330. <https://doi.org/10.1007/s10896-007-9088-9>
- Eliason, M. J., Taylor, J. Y., & Arndt, S. (2005). Assessing intimate partner violence in incarcerated women. *Journal of Forensic Nursing*, 1(3), 106–110. <https://doi.org/10.1111/j.1939-3938.2005.tb00026.x>
- Family Safety Victoria. (2017). *Men's behaviour change minimum standards*. Retrieved from <https://www.ntv.org.au/wp-content/uploads/2018/03/FINAL-MBCP-Minimum-Standards-1.pdf>
- Family Safety Victoria. (2018). Family violence reform: MARAM. Retrieved from <https://www.vic.gov.au/familyviolence/family-safety-victoria/information-sharing-and-risk-management/maram.html>
- Fischer, J., & Corcoran, K. (1994). *Measures for clinical practice* (2nd ed.). New York: Free Press.
- Flanagan, R., & Allen, K. (2005). A review of the children's inventory of anger (ChIA): A needed measure. *Journal of Rational-Emotive and Cognitive-Behavior Therapy*, 23(3), 263–273. <https://doi.org/10.1007/s10942-005-0010-1>
- Fleming, J. S., & Courtney, B. E. (1984). The dimensionality of self-esteem. II. Hierarchical facet model for revised measurement scales. *Journal of Personality and Social Psychology*, 46(2), 404–421. <https://doi.org/10.1037/0022-3514.46.2.404>

- Flood, M. (2019). The problem: Men's violence against women. In M. Flood (Ed.), *Engaging men and boys in violence prevention* (p. 11–38). New York: Palgrave MacMillan.
- Ford-Gilboe, M., Wathen, C. N., Varcoe, C., MacMillan, H. L., Scott-Storey, K., Mantler, T., . . . Perrin, N. (2016). Development of a brief measure of intimate partner violence experiences: The Composite Abuse Scale (Revised)–Short Form (CASR-SF). *BMJ Open*, 6(12), e012824. <https://doi.org/10.1136/bmjopen-2016-012824>
- Furukawa, T. A., Kessler, R. C., Slade, T., & Andrews, G. (2003). The performance of the K6 and K10 screening scales for psychological distress in the Australian national survey of mental health and well-being. *Psychological Medicine*, 33(2), 357–362. <https://doi.org/10.1017/S0033291702006700>
- García-Ael, C., Recio, P., & Silvan-Ferrero, P. (2018). Psychometric properties of the Inventory of Beliefs about Intimate Partner Violence (IBIPV). *Social and Organisational Psychology*, 34(1), 135–145. <https://doi.org/10.6018/analesps.34.1.232901>
- Garner, J., & Hudson, W. (1992). Non-physical abuse of partner scale (NPAPS). In J. Fischer & K. Corcoran (Eds.), *Measures for clinical practice: A sourcebook* (pp. 152–153). New York: Free Press.
- Geldschläger, H., Ginés, O., Nax, D., & Ponce, Á. (2013). *Outcome measurement in European perpetrator programmes: A survey*. Berlin: Dissens Institute for Education and Research.
- Glick, P., & Fiske, S. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, 70(3), 491–512. <https://doi.org/10.1037/0022-3514.70.3.491>
- Glick, P., & Whitehead, J. (2010). Hostility toward men and the perceived stability of male dominance. *Social Psychology*, 41(3), 177–185. <https://doi.org/10.1027/1864-9335/a000025>
- Gölge, Z. B., Sanal, Y., Yavuz, S., & Arslanoglu-Çetin, E. (2016). Attitudes toward wife abuse of police officers and judiciary members in Turkey: Profession, gender, ambivalent sexism and sex roles. *Journal of Family Violence*, 31(6), 785–796. <https://doi.org/10.1007/s10896-016-9823-1>
- Gondolf, E. W. (2012). *The future of batterer programs: Reassessing evidence-based practice*. Boston: Northeastern University Press.
- González-Carrasco, M., Casas, F., Malo, S., Viñas, F., & Dinisman, T. (2017). Changes with age in subjective well-being through the adolescent years: Differences by gender. *Journal of Happiness studies*, 18(1), 63–88. <https://doi.org/10.1007/s10902-016-9717-1>
- Good, J. J., Woodzicka, J. A., & Wingfield, L. C. (2010). The effects of gender stereotypic and counter-stereotypic textbook images on science performance. *The Journal of Social Psychology*, 150(2), 132–147. <https://doi.org/10.1080/00224540903366552>
- Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38(5), 581–586. <https://doi.org/10.1111/j.1469-7610.1997.tb01545.x>
- Goodwin, J. (2010). *Research in psychology: Methods and design* (6th ed.). Hoboken, NJ: John Wiley & Sons.
- Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability: What they are and how to use them. *Educational and Psychological Measurement*, 66(6), 930–944. <https://doi.org/10.1177/0013164406288165>
- Graham, L. M., Sahay, K. M., Rizo, C. F., Messing, J. T., & Macy, R. J. (2019). The validity and reliability of available intimate partner homicide and reassault risk assessment tools: A systematic review. *Trauma, Violence, & Abuse*, 20(1), 1–23. <https://doi.org/10.1177/1524838018821952>
- Gremigni, P., & Stewart-Brown, S. (2011). Measuring mental well-being: Italian validation of the Warwick–Edinburgh Mental Well-being Scale (WEMWBS). *Giornale Italiano di Psicologia*, 38(2), 485–508. <https://doi.org/10.1421/35174>

- Groth-Marnat, G. (2009). *Handbook of psychological assessment* (5th ed.). New Jersey: John Wiley & Sons.
- Hamby, S., & Finkelhor, D. (2001). *Choosing and using child victimization questionnaires*. Washington, DC: US Government Printing Office.
- Hannem, S., Langan, D., & Stewart, C. (2015). "Every couple has their fights ...": Stigma and subjective narratives of verbal violence. *Deviant Behavior*, 36(5), 388–404. <https://doi.org/10.1080/01639625.2014.935688>
- Hardan-Khalil, K., & Mayo, A. M. (2015). Psychometric properties of the multidimensional scale of perceived social support. *Clinical Nurse Specialist*, 29(5), 258–261. <https://doi.org/10.1097/NUR.0000000000000148>
- Hawe, P., Degeling, D., & Hall, J. (1990). Evaluability assessment: Getting ready to test programme effects. In P. Hawe, D. Degeling, & J. Hall (Eds.), *Evaluating health promotion: A health worker's guide* (pp. 86–100). Artarmon, NSW: MacLennan & Petty.
- Hayes, E.-R., & Swim, J. K. (2013). African, Asian, Latina/o, and European Americans' responses to popular measures of sexist beliefs: Some cautionary notes. *Psychology of Women Quarterly*, 37(2), 155–166. <https://doi.org/10.1177/0361684313480044>
- Hearn, J., Andersson, K., & Cowburn, M. (2007). *Background paper on guidelines for researchers on doing research with perpetrators of sexual violence*. Pretoria: Sexual Violence Research Initiative.
- Hegarty, K., Bush, R., & Sheehan, M. (2005). The Composite Abuse Scale: Further development and assessment of reliability and validity of a multidimensional partner abuse measure in clinical settings. *Violence and Victims*, 20(5), 529–547. <https://doi.org/10.1891/0886-6708.2005.20.5.529>
- Hegarty, K., Sheehan, M., & Schonfeld, C. (1999). A multidimensional definition of partner abuse: Development and preliminary validation of the Composite Abuse Scale. *Journal of Family Violence*, 14(4), 399–415. <https://doi.org/10.1023/A:1022834215681>
- Henning, K., & Klesges, L. M. (2003). Prevalence and characteristics of psychological abuse reported by court-involved battered women. *Journal of Interpersonal Violence*, 18(8), 857–871. <https://doi.org/10.1177/0886260503253878>
- Holt, S., Buckley, H., & Whelan, S. (2008). The impact of exposure to domestic violence on children and young people: A review of the literature. *Child Abuse & Neglect*, 32(8), 797–810. <https://doi.org/10.1016/j.chiabu.2008.02.004>
- Howarth, E., Moore, T. H., Stanley, N., MacMillan, H. L., Feder, G., & Shaw, A. (2019). Towards an ecological understanding of readiness to engage with interventions for children exposed to domestic violence and abuse: Systematic review and qualitative synthesis of perspectives of children, parents and practitioners. *Health & Social Care in the Community*, 27(2), 271–292. <https://doi.org/10.1111/hsc.12587>
- Howarth, E., Moore, T., Welton, N., Lewis, N., Stanley, N., MacMillan, H., . . . Feder, G. (2016). *IMPROving Outcomes for children exposed to domestic Violence (IMPROVE): An evidence synthesis*. *Public Health Research*, 4(10), 1–341. <https://doi.org/10.3310/phr04100>
- Hudson W. (1992). *The Partner Abuse Scale*. Tempe, AZ: Walmyr Publishing.
- Hudson, W. (1997). *The WALMYR assessment scales scoring manual*. Tallahassee, FL: WALMYR Publishing Company.
- Hudson, W., & McIntosh, S. (1981). The assessment of spouse abuse: Two quantifiable dimensions. *Journal of Marriage and Family*, 43(4), 873–885. <https://doi.org/10.2307/351344>
- International Development Research Centre. (2012). *Identifying the intended user(s) and use(s) of an evaluation*. Retrieved from <https://www.betterevaluation.org/sites/default/files/idrc.pdf>
- Jenkins, A. (1990). Invitations to responsibility: The therapeutic engagement of men who are violent and abusive. *Australian and New Zealand Journal of Family Therapy*, 11(4), 201–201. <https://doi.org/10.1002/j.1467-8438.1990.tb00823.x>

- Jenkins, A. (2009). *Becoming ethical: A parallel, political journey with men who have abused*. Lyme Regis, UK: Russell House Publishing.
- Johansen, V. A., Wahl, A. K., Eilertsen, D. E., & Weisaeth, L. (2007). Prevalence and predictors of post-traumatic stress disorder (PTSD) in physically injured victims of non-domestic violence. *Social Psychiatry and Psychiatric Epidemiology*, 42(7), 583–593. <https://doi.org/10.1007/s00127-007-0205-0>
- Kania, J., & Kramer, M. (2011). Collective impact. *Stanford Social Innovation Review*, 9(1), 36–41. Retrieved from https://ssir.org/articles/entry/collective_impact
- Kasian, M., & Painter, S. L. (1992). Frequency and severity of psychological abuse in a dating population. *Journal of Interpersonal Violence*, 7(3), 350–364. <https://doi.org/10.1177/088626092007003005>
- Kaufman-Levy, D., & Poulin, M. (2003). *Evaluability assessment: Examining the readiness for program evaluation*. Retrieved from <http://www.jrsa.org/pubs/juv-justice/evaluability-assessment.pdf>
- Kellogg, W. K. (2004). *Using logic models to bring together planning, evaluation, and action: Logic model development guide*. Michigan: W.K. Kellogg Foundation.
- Kelly, J. B., & Johnson, M. P. (2008). Differences among types of intimate partner violence: Research update and implications for interventions. *Family Court Review*, 46(3), 476–499. <https://doi.org/10.1111/j.1744-1617.2008.00215.x>
- Kelly, L., & Westmarland, N. (2015). *Domestic violence perpetrator programmes: Steps towards change—Project Mirabal final report*. London and Durham: London Metropolitan University and Durham University.
- Kersten, P., Czuba, K., McPherson, K., Dudley, M., Elder, H., Tauroa, R., & Vandal, A. (2015). A systematic review of evidence for the psychometric properties of the strengths and difficulties questionnaire. *International Journal of Behavioral Development*, 40(1), 64–75. <https://doi.org/10.1177/0165025415570647>
- Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L., . . . Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine*, 32(6), 959–976. <https://doi.org/10.1017/s0033291702006074>
- Kessler, R. C., Barker, P. R., Colpe, L. J., Epstein, J. F., Gfroerer, J. C., Hiripi, E., . . . Zaslavsky, A. M. (2003). Screening for serious mental illness in the general population. *Archives of General Psychiatry*, 60(2), 184–189. <https://doi.org/10.1001/archpsyc.60.2.184>
- Khaleque, A., & Rohner, R. P. (2002). Perceived parental acceptance–rejection and psychological adjustment: A meta-analysis of cross-cultural and intracultural studies. *Journal of Marriage and Family*, 64(1), 54–64. <https://doi.org/10.1111/j.1741-3737.2002.00054.x>
- Kimmel, M. S. (2002). “Gender symmetry” in domestic violence: A substantive and methodological research review. *Violence Against Women*, 8(11), 1332–1363. <https://doi.org/10.1177/107780102237407>
- Knowlton, L. W., & Phillips, C. C. (2012). *The logic model guidebook: Better strategies for great results* (2nd ed.). London: Sage Publishing.
- Koehler, J. A., Lösel, F., Akoensi, T. D., & Humphreys, D. K. (2013). A systematic review and meta-analysis on the effects of young offender treatment programs in Europe. *Journal of Experimental Criminology*, 9(1), 19–43. <https://doi.org/10.1007/s11292-012-9159-7>
- Koeske, G. F., & Koeske, R. D. (1992). Parenting locus of control: Measurement, construct validation, and a proposed conceptual model. *Social Work and Research Abstracts*, 28(3), 37–46. <https://doi.org/10.1093/swra/28.3.37>
- Kropp, P., & Hart, S. (2000). The Spousal Assault Risk Assessment (SARA) guide: Reliability and validity in adult male offenders. *Law and Human Behavior*, 24(1), 101–118. <https://doi.org/10.1023/a:1005430904495>

- Kropp, P. R. (2008). Development of the spousal assault risk assessment guide (SARA) and the Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER). In A. C. Baldry & F. W. Winkel (Eds.), *Intimate partner violence prevention and intervention* (pp. 19–31). New York: Nova Science Publishers.
- Kropp, P. R., Hart, S. D., & Belfrage, H. (2005). *Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER): User manual*. Vancouver: ProActive ReSolutions, Inc.
- Kropp, P. R., Hart, S. D., Webster, C. D., & Eaves, D. (1994). *Manual for the Spousal Assault Risk Assessment (SARA) Guide*. Vancouver: The British Columbia Institute Against Family Violence.
- La Flair, L. N., Bradshaw, C. P., Mendelson, T., & Campbell, J. (2015). Intimate partner violence and risk of psychiatric symptoms: The moderating role of attachment. *Journal of Family Violence, 30*(5), 567–577. <https://doi.org/10.1007/s10896-015-9681-2>
- Laing, L., & Humphreys, C. (2013). *Social work and domestic violence: Developing critical and reflective practice*. London: Sage Publishing.
- Lee, J., Hayat, M. J., Spratling, R., Sevcik, R. A., & Clark, P. C. (2018). Relationship of mothers' mental and physical health to characteristics of mothers and their children with developmental disabilities. *Nursing Research, 67*(6), 456–464. <https://doi.org/10.1097/nnr.0000000000000309>
- Levesque, D. A., Gelles, R. J., & Velicer, W. F. (2000). Development and validation of a stages of change measure for men in batterer treatment. *Cognitive Therapy and Research, 24*(2), 175–199. <https://doi.org/10.1023/A:1005446025201>
- Lila, M., Gracia, E., & Garcia, F. (2013). Ambivalent sexism, empathy and law enforcement attitudes towards partner violence against women among male police officers. *Psychology, Crime & Law, 19*(10), 907–919. <https://doi.org/10.1080/1068316X.2012.719619>
- Lila, M., Oliver, A., Catalá-Miñana, A., & Conchell, R. (2014). Recidivism risk reduction assessment in batterer intervention programs: A key indicator for program efficacy evaluation. *Psychosocial Intervention, 23*(3), 217–223. <https://doi.org/10.1016/j.psi.2014.11.001>
- Lilley-Walker, S. J., Hester, M., & Turner, W. (2018). Evaluation of European domestic violence perpetrator programmes: Toward a model for designing and reporting evaluations related to perpetrator treatment interventions. *International Journal of Offender Therapy and Comparative Criminology, 62*(4), 868–884. <https://doi.org/10.1177/0306624X16673853>
- Litwin, M. S. (1995). *How to measure survey reliability and validity*. Thousand Oaks: Sage Publications.
- Logan, T. K., Walker, R., Shannon, L., & Cole, J. (2008). Combining ethical considerations with recruitment and follow-up strategies for partner violence victimization research. *Violence Against Women, 14*(11), 1226–1251. <https://doi.org/10.1177/1077801208323976>
- Logan Together. (2017). Logan Together: Resources. Retrieved 27 February, 2020 from: <https://logantogogether.org.au/resources>
- Losada, A., de los Angeles Villareal, M., Nuevo, R., Marquez-Gonzalez, M., Salazar, B. C., Romero-Moreno, R., . . . Fernandez-Fernandez, V. (2012). Cross-cultural confirmatory factor analysis of the CES-D in Spanish and Mexican dementia caregivers. *The Spanish Journal of Psychology, 15*(2), 783–792. https://doi.org/10.5209/rev_SJOP.2012.v15.n2.38890
- Lucente, S. W., Fals-Stewart, W., Richards, H. J., & Goscha, J. (2001). Factor structure and reliability of the Revised Conflict Tactics Scales for incarcerated female substance abusers. *Journal of Family Violence, 16*(4), 437–450. <https://doi.org/10.1023/A:1012281027999>
- Mackay, E., Gibson, A., Lam, H., & Beecham, D. (2015). *Perpetrator interventions in Australia: Part One—Literature review* (State of knowledge paper) (ANROWS Landscapes, 01/2015) Sydney: ANROWS.
- MacMillan, H. L., Wathen, C. N., Jamieson, E., Boyle, M. H., Shannon, H. S., Ford-Gilboe, M., . . . McMaster Violence Against Women Research Group. (2009). Screening for intimate partner violence in health care settings: A randomized trial. *JAMA, 302*(5), 493–501. <https://doi.org/10.1001/jama.2009.1089>

- Mark, M., Henry, G., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs*. San Francisco: Jossey-Bass Inc.
- Marshall, L. L. (1992). Development of the severity of violence against women scales. *Journal of Family Violence*, 7(2), 103–121. <https://doi.org/10.1007/BF00978700>
- Masser, B., Viki, G. T., & Power, C. (2006). Hostile sexism and rape proclivity amongst men. *Sex Roles*, 54(7), 565–574. <https://doi.org/10.1007/s11199-006-9022-2>
- McConnell, N., Barnard, M., Holdsworth, T., & Taylor, J. (2014). *Caring Dads: Safer Children: Interim evaluation report*. London: National Society for the Prevention of Cruelty to Children.
- McDermott, M. J., & Garofalo, J. (2004). When advocacy for domestic violence victims backfires: Types and sources of victim disempowerment. *Violence Against Women*, 10(11), 1245–1266. <https://doi.org/10.1177/1077801204268999>
- McFarlane, J., Parker, B., & Soeken, K. (1995). Abuse during pregnancy: Frequency, severity, perpetrator, and risk factors of homicide. *Public Health Nursing*, 12(5), 284–289. <https://doi.org/10.1111/j.1525-1446.1995.tb00150.x>
- McKenzie, J., White, T., Minty, R., & Clancy, L. (2016). *Program planning and evaluation: How to use evidence to support communities for children services*. Melbourne: Centre for Family Research and Evaluation.
- McLaren, H., Fischer, J., & Zannettino, L. (2020). *Defining quality of life indicators for measuring perpetrator intervention effectiveness* (Research report, 05/2020). Sydney: ANROWS
- Mechanic, M. B., Uhlmansiek, M. H., Weaver, T. L., & Resick, P. A. (2000). The impact of severe stalking experienced by acutely battered women: An examination of violence, psychological symptoms and strategic responding. *Violence and Victims*, 15(4), 443–458. <https://doi.org/10.1891/0886-6708.15.4.443>
- Migliore, C., Ziersch, E., & Marshall, J. (2014). *Intervention orders and the intervention response model: Evaluation report 3* (Statistical overview and outcome evaluation). Adelaide: Attorney General's Department.
- Mihalic, S. F., & Elliott, D. S. (2015). Evidence-based programs registry: Blueprints for healthy youth development. *Evaluation and Program Planning*, 48(5), 124–131. <https://doi.org/10.1016/j.evalprogplan.2014.08.004>
- Miles, J., & Gilbert, P. (2005). *A handbook of research methods for clinical and health psychology*. Oxford: Oxford University Press.
- Murphy, C. M., & Cascardi, M. (1993). Psychological aggression and abuse in marriage. In R. L. Hampton, T. P. Gullotta, G. R. Adams, E. H. Potter III & R. P. Weissberg (Eds.), *Family violence: Prevention and treatment* (pp. 86–112). Newbury Park, CA: Sage.
- Murray, C. E., & Graves, K. N. (2013). *Responding to family violence: A comprehensive, research-based guide for therapists*. New York: Routledge.
- National Health and Medical Research Council, Australian Research Council, & Universities Australia. (2018). *National statement on ethical conduct in human research 2007* (Rev. ed.). Canberra: NHMRC.
- Nelson, W. M., & Finch, A. J. (2000). *Children's Inventory of Anger: ChIA Manual*. Los Angeles: Western Psychological Services.
- New South Wales Department of Justice. (2012). *Minimum standards for men's domestic violence behaviour change programs*. Retrieved from [http://www.crimeprevention.nsw.gov.au/domesticviolence/Documents/Mini/dfv_behaviour_change_program_standards_april_2012%20\(1\).pdf](http://www.crimeprevention.nsw.gov.au/domesticviolence/Documents/Mini/dfv_behaviour_change_program_standards_april_2012%20(1).pdf)
- New South Wales. Department of Justice. (2017). *Practice standards for men's domestic violence behaviour change programs*. Retrieved from <http://www.crimeprevention.nsw.gov.au/domesticviolence/PublishingImages/NSW%20Department%20of%20Justice%20-%20Men%27s%20Behaviour%20Change%20Programs%20-%20Practice%20Standards.pdf>
- New South Wales. Department of Justice. (2018). *Compliance framework for men's behaviour change programs*. Retrieved from <http://www.crimeprevention.nsw.gov.au/domesticviolence/Documents/mbcp-compliance-framework.pdf>

- Nguyen, H. T., Kitner-Triolo, M., Evans, M. K., & Zonderman, A. B. (2004). Factorial invariance of the CES-D in low socioeconomic status African Americans compared with a nationally representative sample. *Psychiatry Research*, 126(2), 177–187. <https://doi.org/10.1016/j.psychres.2004.02.004>
- Nicholls, T. L., Pritchard, M. M., Reeves, K. A., & Hilterman, E. (2013). Risk assessment in intimate partner violence: A systematic review of contemporary approaches. *Partner Abuse*, 4(1), 76–168. <https://doi.org/10.1891/1946-6560.4.1.76>
- No to Violence. (2006). *Men's behaviour change group work: Minimum standards and quality practice*. Melbourne: NTV.
- No to Violence. (2018). *Implementation guide: Men's behaviour change minimum standards*. Retrieved from <https://www.ntv.org.au/wp-content/uploads/2018/09/Minimum-Standards-manual-August-2018-FINAL-140818-Screen-ready-FA1.pdf>
- No to Violence, & Red Tree Consulting. (2012). *Towards safe families: A practice guide for men's domestic violence behaviour change programs*. Retrieved from http://www.crimeprevention.nsw.gov.au/domesticviolence/Documents/Mini/agj_domestic_violence_practice_guide_final_consolidated_sec.pdf
- Opitz, C. (2014). Considerations for partner contact during men's behaviour change programs: Systemic responses and engagement. *Ending Men's Violence Against Women and Children: The No To Violence Journal*, Autumn, 114–142. Retrieved from <http://shantiworks.com.au/wp-content/uploads/2014/08/2014-ntv-journal-tc-article.pdf>
- Pape, K. T., & Arias, I. (2000). The role of perceptions and attributions in battered women's intentions to permanently end their violent relationships. *Cognitive Therapy and Research*, 24(2), 201–214. <https://doi.org/10.1023/A:1005498109271>
- Patton, M. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Thousand Oaks: Sage Publishing.
- Paymar, M., & Barnes, G. (2007). *Countering confusion about the Duluth model*. Duluth, MN: Domestic Abuse Intervention Programs.
- Pence, E., Paymar, M., Ritmeester, T., & Shepard, M. (1993). *Education groups for men who batter: The Duluth model*. New York: Springer Publishing Company.
- Pham, M. T., Rajić, A., Greig, J. D., Sargeant, J. M., Papadopoulos, A., & McEwen, S. A. (2014). A scoping review of scoping reviews: Advancing the approach and enhancing the consistency. *Research Synthesis Methods*, 5(4), 371–385. <https://doi.org/10.1002/jrsm.1123>
- Price, B. J., & Rosenbaum, A. (2009). Batterer intervention programs: A report from the field. *Violence and Victims*, 24(6), 757–770. <https://doi.org/10.1891/0886-6708.24.6.757>
- Price, P. C., Jhangiani, R. S., & Chiang, I.-C. A. (2015). Chapter 5: Psychological measurement. In P. C. Price, R. Jhangiani, & I.-C. A. Chiang (Eds.), *Research methods in psychology* (2nd ed.) (pp. 79–97). Canada: BCCampus and OpenEd.
- Prochaska, J. O., & DiClemente, C. C. (1983). Stages and processes of self-change of smoking: Toward an integrative model of change. *Journal of Consulting and Clinical Psychology*, 51(3), 390–395. <https://doi.org/10.1037//0022-006X.51.3.390>
- Queensland. Department of Child Safety, Youth and Women. (2016a). *Queensland says not now, not ever: Domestic and family violence prevention strategy 2016–2026*. Retrieved from <https://www.csyw.qld.gov.au/resources/campaign/end-violence/dfv-prevention-strategy.pdf>
- Queensland. Department of Child Safety Youth and Women. (2016b). *Professional practice standards: Working with men who perpetrate domestic and family violence*. Retrieved from <https://publications.qld.gov.au/storage/f/2014-10-21T05%3A16%3A49.846Z/professional-practice-standards>
- Queensland. Department of Child Safety Youth and Women. (2016c). *Second action plan of the domestic and family violence prevention strategy 2016–17 to 2018–19*. Retrieved from <https://www.communities.qld.gov.au/resources/gateway/campaigns/end-violence/dfv-second-action-plan.pdf>

- Queensland. Department of Child Safety, Youth and Women. (2019). *Approved providers and approved intervention programs*. Retrieved from <https://www.csyw.qld.gov.au/resources/dcsyw/violence-prevention/approved-providers-and-intervention-programs.pdf>
- Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1(3), 385–401. <https://doi.org/10.1177/014662167700100306>
- Respect. (2017). *Respect outcomes framework*. Retrieved from <http://respect.uk.net/wp-content/uploads/2017/02/Respect-Outcomes-Framework-15.11.17.pdf>
- Riemer, A., Chaudoir, S., & Earnshaw, V. (2014). What looks like sexism and why? The effect of comment type and perpetrator type on women's perceptions of sexism. *The Journal of General Psychology*, 141(3), 263–279. <https://doi.org/10.1080/00221309.2014.907769>
- Rodenburg, F. A., & Fantuzzo, J. W. (1993). The measure of wife abuse: Steps toward the development of a comprehensive assessment technique. *Journal of Family Violence*, 8(3), 203–228. <https://doi.org/10.1007/BF00988769>
- Rohner, R., & Ali, S. (2016). Parental Acceptance–Rejection Questionnaire (PARQ). In V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of personality and individual differences* (pp. 1–4). London: Springer International Publishing.
- Rohner, R., & Khaleque, A. (2005). *Handbook for the study of parental acceptance and rejection* (4th ed.). Storrs, CT: Rohner Research Publications.
- Rohner, R. P. (1990). *Handbook for the study of parental acceptance and rejection* (3rd ed.). Storrs, CT: Rohner Research Publications.
- Rollero, C., Glick, P., & Tartaglia, S. (2014). Psychometric properties of short versions of the ambivalent sexism inventory and ambivalence toward men inventory. *TPM*, 21(2), 1–11. <https://doi.org/10.4473/TPM21.2.3>
- Ronan, G. F., Dreer, L., Maurelli, K., Ronan, D., & Gerhart, J. (2013). *Practitioner's guide to empirically supported measures of anger, aggression, and violence*: Cham, Switzerland: Springer International Publishing.
- Ros, L., Latorre, J. M., Aguilar, M. J., Serrano, J. P., Navarro, B., & Ricarte, J. J. (2011). Factor structure and psychometric properties of the Center for Epidemiological Studies Depression Scale (CES-D) in older populations with and without cognitive impairment. *International Journal of Aging and Human Development*, 72(2), 83–110. <https://doi.org/10.2190/AG.72.2.a>
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton: University Press.
- Rychetnik, L., Frommer, M., Hawe, P., & Shiell, A. (2002). Criteria for evaluating evidence on public health interventions. *Journal of Epidemiology & Community Health*, 56(2), 119–127. <https://doi.org/10.1136/jech.56.2.119>
- Sackett, L. A., & Saunders, D. G. (1999). The impact of different forms of psychological abuse on battered women. *Violence and Victims*, 14(1), 105–117. <https://doi.org/10.1891/0886-6708.14.1.105>
- Salabarría-Peña, Y., Apt, B. S., & Walsh, C. M. (2007). *Practical use of program evaluation among sexually transmitted disease (STD) programs: Vol. 4—Gather credible evidence*. Atlanta: Centers for Disease Control and Prevention.
- Salkind, N. J. (2010). *Encyclopedia of research design*. Thousand Oaks: Sage Publications.
- Saunders, D. G., Lynch, A. B., Grayson, M., & Linz, D. (1987). The inventory of beliefs about wife beating: The construction and initial validation of a measure of beliefs and attitudes. *Violence and Victims*, 2(1), 39–57. Retrieved from <https://deepblue.lib.umich.edu/bitstream/handle/2027.42/50455/The%20Inventory%20of%20Beliefs%20about%20Wife%20BeatingSaunders%20EtAl%20SEC.pdf?sequence=1&isAllowed=y>
- Schwarzer, R., & Jerusalem, M. (1995). Generalized self-efficacy scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), *Measures in health psychology: A user's portfolio* (pp. 35–37). Windsor, UK: NFER-Nelson.

- Scott, K. L., & Wolfe, D. A. (2003). Readiness to change as a predictor of outcome in batterer treatment. *Journal of Consulting and Clinical Psychology, 71*(5), 879–889. <https://doi.org/10.1037/0022-006X.71.5.879>
- Semiatin, J. N., Murphy, C. M., & Elliott, J. D. (2013). Observed behavior during group treatment for partner-violent men: Acceptance of responsibility and promotion of change. *Psychology of Violence, 3*(2), 126–139. <https://doi.org/10.1037/a0029846>
- Shepard, M., & Campbell, J. (1992). The Abusive Behavior Inventory: A measure of psychological and physical abuse. *Journal of Interpersonal Violence, 7*(3), 291–305. <https://doi.org/10.1177/088626092007003001>
- Silvergleid, C. S., & Mankowski, E. S. (2006). How batterer intervention programs work: Participant and facilitator accounts of processes of change. *Journal of Interpersonal Violence, 21*(1), 139–159. <https://doi.org/10.1177/0886260505282103>
- Smith, P., Gendreau, P., & Swartz, K. (2009). Validating the principles of effective intervention: A systematic review of the contributions of meta-analysis in the field of corrections. *Victims and Offenders, 4*(2), 148–169. <https://doi.org/10.1080/15564880802612581>
- Smith, P., Smith, J., & Earp, J. (1999). Beyond the measurement trap: A reconstructed conceptualization and measurement of battering. *Psychology of Women Quarterly, 23*(1), 179–195. <https://doi.org/10.1111/j.1471-6402.1999.tb00350.x>
- Smith, P. H., Earp, J. A., & DeVellis, R. (1995). Measuring battering: Development of the Women's Experience with Battering (WEB) scale. *Women's Health: Research on Gender, Behavior, and Policy, 1*(4), 273–288.
- Smith, P. H., Thornton, G. E., DeVellis, R., Earp, J., & Coker, A. L. (2002). A population-based study of the prevalence and distinctiveness of battering, physical assault, and sexual assault in intimate relationships. *Violence Against Women, 8*(10), 1208–1232. <https://doi.org/10.1177/107780120200801004>
- Special Taskforce on Domestic and Family Violence in Queensland. (2015). *Not now, not ever: Putting an end to domestic and family violence in Queensland*. Retrieved from <https://www.csyw.qld.gov.au/campaign/end-domestic-family-violence/about/not-now-not-ever-report>
- Spence, J. T., & Hahn, E. D. (1997). The attitudes toward women scale and attitude change in college students. *Psychology of Women Quarterly, 21*(1), 17–34. <https://doi.org/10.1111/j.1471-6402.1997.tb00098.x>
- Spence, J. T., & Helmreich, R. L. (1972). The Attitudes toward Women Scale: An objective instrument to measure the attitude toward the rights and roles of women in contemporary society. *JSAS: Catalogue of Selected Documents in Psychology, 2*, 66–67.
- Spence, J. T., & Helmreich, R. L. (1978). *Masculinity & femininity: Their psychological dimensions, correlates, & antecedents*. Austin: University of Texas Press.
- Spence, J. T., Helmreich, R. L., & Stapp, J. (1973). A short version of the Attitudes toward Women Scale (AWS). *Bulletin of the Psychonomic Society, 2*(4), 219–220. <https://doi.org/10.3758/BF03329252>
- Sprague, S., McKay, P., Madden, K., Scott, T., Tikasz, D., Slobogean, G. P., & Bhandari, M. (2016). Outcome measures for evaluating intimate partner violence programs within clinical settings: A systematic review of the literature. *Trauma, Violence, & Abuse, 18*(5), 508–522. <https://doi.org/10.1177/1524838016641667>
- State of Victoria. (2016). *Royal Commission into Family Violence: Report and recommendations, Vol. III* (Parl paper no. 132, 2014–16). Retrieved from <http://www.rcfv.com.au/MediaLibraries/RCFamilyViolence/Reports/Final/RCFV-Vol-I.pdf>
- State of Victoria. (2018). *Expert Advisory Committee on Perpetrator Interventions: Final Report*. Retrieved from <https://apo.org.au/sites/default/files/resource-files/2019-11/apo-nid268781.pdf>
- Stewart, A. L. (2014). The men's project: A sexual assault prevention program targeting college men. *Psychology of Men & Masculinity, 15*(4), 481–485. <https://doi.org/10.1037/a0033947>

- Stith, S. M., Smith, D. B., Penn, C. E., Ward, D. B., & Tritt, D. (2004). Intimate partner physical abuse perpetration and victimization risk factors: A meta-analytic review. *Aggression and Violent Behavior, 10*(1), 65–98. <https://doi.org/10.1016/j.avb.2003.09.001>
- Stover, C. S., Meadows, A. L., & Kaufman, J. (2009). Interventions for intimate partner violence: Review and implications for evidence-based practice. *Professional Psychology: Research and Practice, 40*(3), 223. <https://doi.org/10.1037/a0012718>
- Straus, M. A. (1979). Measuring intrafamily conflict and violence: The Conflict Tactics (CT) scales. *Journal of Marriage and Family, 41*(1), 75–88. <https://doi.org/10.2307/351733>
- Straus, M. A., Hamby, S. L., Boney-McCoy, S. U. E., & Sugarman, D. B. (1996). The Revised Conflict Tactics Scales (CTS2): Development and preliminary psychometric data. *Journal of Family Issues, 17*(3), 283–316. <https://doi.org/10.1177/019251396017003001>
- Straus, M. A., Hamby, S., & Warren, W. (2003). *The conflict tactics scales handbook: Revised Conflict Tactics Scale (CTS2) and CTS: Parent–Child Version (CTSPC)*. Los Angeles: Western Psychological Services.
- Street, A. E., & Arias, I. (2001). Psychological abuse and posttraumatic stress disorder in battered women: Examining the roles of shame and guilt. *Violence and Victims, 16*(1), 65–78. <https://doi.org/10.1891/0886-6708.16.1.65>
- Stroshine, M. S., & Robinson, A. L. (2003). The decision to end abusive relationships: The role of offender characteristics. *Criminal Justice and Behavior, 30*(1), 97–117. <https://doi.org/10.1177/0093854802239165>
- Stylianou, A. M. (2018). Economic abuse experiences and depressive symptoms among victims of intimate partner violence. *Journal of Family Violence, 33*(6), 381–392. <https://doi.org/10.1007/s10896-018-9973-4>
- Talley, P., Heitkemper, M., Chicz-Demet, A., & Sandman, C. A. (2006). Male violence, stress, and neuroendocrine parameters in pregnancy: A pilot study. *Biological research for nursing, 7*(3), 222–233. <https://doi.org/10.1177/1099800405283182>
- Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., . . . Stewart-Brown, S. (2007). The Warwick–Edinburgh Mental Well-being Scale (WEMWBS): Development and UK validation. *Health and Quality of Life Outcomes, 5*(1), 63. <https://doi.org/10.1186/1477-7525-5-63>
- Tharp, A. T., DeGue, S., Lang, K., Valle, L. A., Massetti, G., Holt, M., & Matjasko, J. (2011). Commentary on Foubert, Godin, & Tatum (2010): The evolution of sexual violence prevention and the urgency for effectiveness. *Journal of Interpersonal Violence, 26*(16), 3383–3392. <https://doi.org/10.1177/0886260510393010>
- Thompson, M., Basile, K., Hertz, M., & Sitterle, D. (2006). *Measuring intimate partner violence, victimization and perpetration: A compendium of assessment tools*. Retrieved from <https://stacks.cdc.gov/view/cdc/11402>
- Tjaden, P., & Thoennes, N. (2000). Prevalence and consequences of male-to-female and female-to-male intimate partner violence as measured by the national violence against women survey. *Violence Against Women, 6*(2), 142–161. <https://doi.org/10.1177/10778010022181769>
- Toivonen, C., & Backhouse, C. (2018). *National risk assessment principles for domestic and family violence*. Sydney: ANROWS.
- Tolman, R. (2001). The validation of the psychological maltreatment of women inventory. In K. O’Leary & R. Maiuro (Eds.), *Psychological abuse in domestic violence relationships* (pp. 47–60). New York: Springer Publishing Company.
- Tomyn, A., & Cummins, R. (2011). The subjective wellbeing of high-school students: Validating the Personal Wellbeing Index–School Children. *Social Indicators Research, 101*(3), 405–418. <https://doi.org/10.1007/s11205-010-9668>
- Torres, A., Navarro, P., García-Esteve, L., Tarragona, M. J., Ascaso, C., Herreras, Z., . . . Martín-Santos, R. (2010). Detecting domestic violence: Spanish external validation of the index of spouse abuse. *Journal of Family Violence, 25*(3), 275–286. <https://doi.org/10.1007/s10896-009-9290-z>

- Tripathy, J. P. (2013). Secondary data analysis: Ethical issues and challenges. *Iranian Journal of Public Health*, 42(2), 1478–1479. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4441947/>
- Trochim, W. M., Donnelly, J. P., & Arora, K. (2016). *Research methods: The essential knowledge base*. Boston: Cengage Learning.
- Trousselard, M., Steiler, D., Dutheil, F., Claverie, D., Canini, F., Fenouillet, F., . . . Franck, N. (2016). Validation of the Warwick–Edinburgh Mental Well-Being Scale (WEMWBS) in French psychiatric and general populations. *Psychiatry Research*, 245(2016), 282–290. <https://doi.org/10.1016/j.psychres.2016.08.050>
- University of Melbourne (n.d.). Do I need ethics approval? Retrieved from <https://staff.unimelb.edu.au/research/ethics-integrity/human-ethics/faq/do-i-need-ethics-approval>
- Victoria. Department of Human Services. (2012). *Family violence risk assessment and risk management framework and practice guides 1–3*. Melbourne: Department of Human Services.
- Vlais, R. (2014). *Ten challenges and opportunities for domestic violence perpetrator program work*. Melbourne: No to Violence.
- Vlais, R., Campbell, E., & Green, D. (2019). *Foundations for family and domestic violence perpetrator interventions systems*. Melbourne: Centre for Innovative Justice, RMIT University.
- Vlais, R., & Green, D. (2018). *Developing an outcomes framework for men’s behaviour change programs*. Perth: Stopping Family Violence.
- Vlais, R., Ridley, S., Green, D., & Chung, D. (2017). *Family and domestic violence perpetrator programs: Issues paper of current and emerging trends, developments and expectations*. Perth: Stopping Family Violence.
- Wakeling, H. C., & Barnett, G. D. (2014). The relationship between psychometric test scores and reconviction in sexual offenders undertaking treatment. *Aggression and Violent Behavior*, 19(2), 138–145. <https://doi.org/10.1016/j.avb.2014.01.002>
- Walby, S., Towers, J., Balderston, S., Corradi, C., Francis, B. J., Heiskanen, M., . . . Strid, S. (2017). *The concept and measurement of violence against women and men*. Bristol: Policy Press.
- Wang, S., Moss, J. R., & Hiller, J. E. (2006). Applicability and transferability of interventions in evidence-based public health. *Health Promotion International*, 21(1), 76–83. <https://doi.org/10.1093/heapro/dai025>
- Wathen, C. N., Jamieson, E., MacMillan, H. L., & McMaster Violence Against Women Research Group. (2008). Who is identified by screening for intimate partner violence? *Women’s Health Issues*, 18(6), 423–432. <https://doi.org/10.1016/j.whi.2008.08.003>
- Western Australia. Department for Child Protection and Family Support. (2015). *Practice standards for perpetrator intervention: Engaging and responding to men who are perpetrators of family and domestic violence*. Perth: Department for Child Protection and Family Support.
- Westmarland, N., Kelly, L., & Chalder-Mills, J. (2010). *Domestic violence perpetrator programmes: What counts as success?* London: London Metropolitan University and Durham University.
- Wistow, R., Kelly, L., & Westmarland, N. (2017). “Time out”: A strategy for reducing men’s violence against women in relationships? *Violence Against Women*, 23(6), 730–748. <https://doi.org/10.1177/1077801216647944>
- Work with Perpetrators of Domestic Violence–European Network (WWP–EN). *Project IMPACT*. Retrieved 21 February, 2020 from <https://www.work-with-perpetrators.eu/projects/project-impact>
- World Health Organization, & Pan American Health Organization. (2012). *Understanding and addressing violence against women: Intimate partner violence* (Information sheet). Geneva: WHO.
- Zhang, B., Fokkema, M., Cuijpers, P., Li, J., Smits, N., & Beekman, A. (2011). Measurement invariance of the Center for Epidemiological Studies Depression Scale (CES-D) among Chinese and Dutch elderly. *BMC Medical Research Methodology*, 11(74), 1–10. <https://doi.org/10.1186/1471-2288-11-74>

Zimet, G. D., Dahlem, N.W., Zimet, S.G. & Farley, G.K. (1988). The Multidimensional Scale of Perceived Social Support. *Journal of Personality Assessment*, 52(1), 30–41. https://doi.org/10.1207/s15327752jpa5201_2

Zimet, G. D., Powell, S. S., Farley, G. K., Werkman, S., & Berkoff, K. A. (1990). Psychometric characteristics of the Multidimensional Scale of Perceived Social Support. *Journal of Personality Assessment*, 55(3–4), 610–617. <https://doi.org/10.1080/00223891.1990.9674095>

Data collection guide for real-world examples

This document outlines the information we aim to collect for each evaluation example. This information will be gained through a combination of reviewing available evaluation documentation and a 1-hour interview with the evaluator(s).

Program information

- Program name
- Program provider
- Target group
- Program structure (e.g. timing, duration of sessions and number of attendees)
- Program approach, theory or philosophy

Evaluation information

- Evaluators and organisation
- Evaluation funding source
- Evaluation duration and dates

Use of theory and frameworks

- Was the evaluation guided by theories of behaviour change, program logic models or evaluation frameworks?
 - What were these models or frameworks? (Can we get a copy?)
 - How were these important in helping to design the evaluation?
 - Did the intervention provider already have these, or did the evaluator assist the provider in developing them?
 - What was the development process? (If applicable)
 - How useful was this process of assisting development for the program provider, and for the evaluator? (If applicable)

Determining criteria for success (targeted outcomes)

- What were the criteria for success of the program?
- What process was used to determine what counted as success?
- What success criteria, if any, were judged too difficult to address in the evaluation?
- Were there any differences in expectations of the evaluation (e.g. in what success criteria were valued) and other stakeholders?
 - Intervention/evaluation funder (generally a state or territory government department)
 - Intervention provider
 - Other stakeholders and the evaluator

Choosing measures

- What were the measures or indicators used (including process evaluation measures)?
- Why and how were these measures or indicators chosen?
 - What factors influenced the search for measures?

- What criteria were prioritised over others in terms of choosing measures? For example, how were the psychometric properties considered in terms of conceptual fit with the program logic model?
- What were the difficulties encountered in determining measures for any component of the evaluation?
 - Were there compromises in terms of needing to choose measures that didn't align as closely as desired with the program logic model and evaluation framework? Why were such compromises required?

Choosing data collection methods

Quantitative components

- What quantitative methods of data collection were used?
- What informed the choice of these methods?
 - What compromises need to be made, if any?
 - Were some creative workarounds or tweaks used to nudge the design closer towards "experimental" design? (E.g. creativity in employing naturalistic experimental designs)
- Were there difficulties encountered in conducting quantitative data collection?

Qualitative components

- What qualitative methods were used?
- Why were these qualitative methods chosen?
- How important were the data gathered using these qualitative methods in helping to interpret overall findings?
- Were there difficulties encountered in conducting qualitative data collection?

General methods

- Did the "process of the process evaluation" turn out differently than planned?
 - Was the focus on process evaluation bigger than expected because the intervention didn't exactly go as planned, for e.g.?
- How did the evaluation methodology address the difference between program-level and systems-level impacts? (If relevant)
- Were any new components introduced to the evaluation mid-stream?

Barriers and enablers

- What were the overall difficulties and barriers in conducting the evaluation?
 - Did this take longer than expected?
- What were some of the helpful practices that maximised data obtainment?
 - Include obtaining data from victims/survivors if relevant
- Were there any other compromises required in the evaluation—parts or aspects of the evaluation that could not be implemented?
- What were the things that were most important about the evaluation design, methodology and process that contributed towards the success of the evaluation process?
- What were the things that looking back now, you'd do differently in the evaluation?
 - Put more emphasis or focus on
 - Substitute different methodologies or measures
- What did the evaluation experience say about the capacity required to conduct this type of evaluation?
 - The capacity required of the evaluator
 - The capacity of the intervention provider to contribute to the evaluation

Relationships with stakeholders

- What was the nature of the relationship between the evaluator, intervention provider, funder and any other stakeholders?
- What role did the program providers play in supporting the evaluation?
- Did the evaluation leave the program provider with a bit more knowledge of evaluation methods/enhance their internal expertise to contribute towards evaluation and research in the future?
- Were there tensions or barriers in the relationship? (Again, this might be too sensitive to report)

Ethical issues

- Did you gain ethical approval from any organisation to carry out the evaluation?
- Were there any particular ethical issues you were aware of before you began the evaluation?
- Did any additional ethical issues arise during the evaluation?
- How did you manage these ethical issues?

APPENDIX B:

Validated scales for use in MBCP evaluation

Measure	Overview	Number of items and subscales	Target group	To be used in conjunction with	Access/administration	Link to measure/permissions/original publication
Long-term changes in perpetrators' violent and controlling behaviour						
Revised Safe at Home Instrument (SAHI; Begun et al., 2009) ^a	The SAHI is designed to assess individuals' readiness to change their intimate partner violence behaviours	35 items Four stages of change: 1. precontemplation 2. contemplation 3. preparation/action 4. maintenance	Men who use violence	Partner reports from victims/survivors	Publicly available (must cite original publication: see link) No cost Self-administered Less than 20 minutes for administration	https://www.researchgate.net/publication/23253921_The_Revised_Safe_At_Home_Instrument_for_Assessing_Readiness_to_Change_Intimate_Partner_Violence
Ambivalent Sexism Inventory (ASI; Glick & Fiske, 1996) ^a	The ASI is a perpetrator self-report measure. It was developed to measure endorsement of sexism (attitudes only)	22 items Two subscales: 1. hostile sexism 2. benevolent sexism	Men who use violence	Partner reports from victims/survivors	Publicly available (must cite original publication: see link) No cost Self-administered Less than 15 minutes for administration	https://www.researchgate.net/publication/232548173_The_Ambivalent_Sexism_Inventory_Differentiating_Hostile_and_Benevolent_Sexism/link/0deec52092e61ddbbae000000/download
Proximal Antecedents to Violent Episode (PAVE; Babcock et al., 2004) ^a	The PAVE is a measure of perpetrator's anger and aggression designed to assess a perpetrator's self-reported likelihood to perpetrate IPV	30 items Three subscales: 1. violence to control 2. violence out of jealousy 3. violence following verbal abuse	Men who use violence	Partner reports from victims/survivors	No cost for research and educational use (must cite original publication: see link) Self-administered Less than 15 minutes for administration	https://sabi.unc.edu/pdf/PAVE_Babcock%20(1).pdf

Measure	Overview	Number of items and subscales	Target group	To be used in conjunction with	Access/administration	Link to measure/permissions/original publication
Modified Abusive Behavior Inventory (ABI; Shepard & Campbell, 1992) ^a	The ABI is a measure of perpetrator's physical and psychological abuse and victim's/survivor's experience of physical and other forms of violence	29 items Two subscales: 1. physical abuse (12 items) 2. psychological abuse (17 items) The forms are identical for victims/survivors and perpetrators except for the pronoun use (e.g. "kicked you" vs. "kicked her")	Men who use violence; victims/survivors of violence	Additional measures to capture victim/survivor safety, wellbeing and freedom from coercive control Children's safety, wellbeing and development Effects on family functioning	Need permission for use Self-administered Less than 15 minutes for administration	Copyright © 1992, Shepard and Campbell, Sage Publications, Thousand Oaks https://www.researchgate.net/publication/249723399_The_Abusive_Behavior_Inventory/link/0deec5303b7f22894200000/download
Partner Abuse Scale: Non-Physical (PASNP; Hudson, 1992) and Partner Abuse Scale: Physical (PASPH; Hudson, 1992)	The PASNP and PASPH measure the extent of physical and non-physical abuse perpetrated against an intimate partner, completed by the victim/survivor	25 items Two subscales: 1. physical abuse 2. non-physical abuse	Victims/survivors of violence	Additional measures to capture victim/survivor safety, wellbeing and freedom from coercive control Children's safety, wellbeing and development Effects on family functioning	Need permission for use Self-administered Less than 10 minutes for administration	Copyright © 1992, James W. Garner and Walter W. Hudson, WALMYR Publishing Company walmyr@walmyr.com https://stacks.cdc.gov/view/cdc/11402/cdc_11402 See a copy of the PASNP on pp. 68-69 See a copy of the PASPH on pp. 19-20
Index of Spouse Abuse (ISA; Hudson & McIntosh, 1981)	The ISA measures victims'/survivors' self-reported severity of both physical and non-physical partner abuse	30 items Two subscales: 1. physical abuse 2. non-physical abuse	Victims/survivors of violence	Additional measures to capture victim/survivor safety, wellbeing and freedom from coercive control Children's safety, wellbeing and development Effects on family functioning	Limited access Original article needs subscription for full access Self-administered Less than 15 minutes for administration	Hudson, W. W., & McIntosh, S. R. (1981). The assessment of spouse abuse: Two quantifiable dimensions. <i>Journal of Marriage and the Family</i> , 43(4), 873-885. https://doi.org/10.2307/351344

Measure	Overview	Number of items and subscales	Target group	To be used in conjunction with	Access/administration	Link to measure/permissions/original publication
Composite Abuse Scale (CAS; Hegarty et al., 1999)	The CAS is a measure for victims/survivors to report the frequency of experience in relation to 30 violent acts over a 12-month period	30 items Four subscales: 1. combined abuse 2. emotional abuse 3. physical abuse 4. harassment	Victims/survivors of violence	Additional measures to capture victim/survivor safety, wellbeing and freedom from coercive control Children's safety, wellbeing and development Effects on family functioning	Need author permission for use No fee Self-administered Less than 15 minutes for administration	Copyright © Hegarty 1999 Contact Professor Kelsey Hegarty at k.hegarty@unimelb.edu.au bmjopen-2016-012824supp.pdf
Severity of Violence Against Women Scale (SVAWS; Marshall, 1992)	The SVAWS measures both frequency and severity of violent behaviours experienced by women in intimate relationships	49 items Three subscales: 1. threats of violence 2. acts of violence 3. sexual aggression	Victims/survivors of violence	Additional measures to capture victim/survivor safety, wellbeing and freedom from coercive control Children's safety, wellbeing and development Effects on family functioning	Publicly available Self-administered Less than 20 minutes for administration	http://www.midss.org/content/severity-violence-against-women-scale-svaws
Psychological Maltreatment of Women Inventory (PMWI; Tolman, 1989)	The PMWI is a measure of psychological abuse experienced by victims/survivors in intimate relationships	58 items Two sub-scales: 1. dominance-isolation 2. verbal-emotional abuse	Victims/survivors of violence	Additional measures to capture victim/survivor safety, wellbeing and freedom from coercive control Children's safety, wellbeing and development Effects on family functioning	Need author permission for use No fee Self-administered Less than 30 minutes for administration	Terms of use: http://www-personal.umich.edu/~rtolman/index.html

Measure	Overview	Number of items and subscales	Target group	To be used in conjunction with	Access/administration	Link to measure/permissions/original publication
Profile of Psychological Abuse of Women- Revised (PPAW; Sackett & Saunders, 1999)	The PPAW measures victims'/survivors' experiences of a wide variety of psychological abuse	21 items Four subscales: 1. criticise behaviour 2. ignore 3. ridicule traits 4. jealousy/control	Victims/survivors of violence	Additional measures to capture victim/survivor safety, wellbeing and freedom from coercive control Children's safety, wellbeing and family functioning	Need author permission for use No fee Self-administered Less than 10 minutes for administration	Copyright © 1999, Springer Publishing Company, Inc., New York, NY 10036 Sackett, L. A., & Saunders, D. G. (1999). The impact of different forms of psychological abuse on battered women. <i>Violence and Victims</i> , 14(1), 105-117. https://stacks.cdc.gov/view/cdc/11402/cdc_11402 See a copy of the scale on pp. 70-71
Adult victims'/survivors' safety, wellbeing and freedom						
Warwick-Edinburgh Mental Well-being Scale (WEMWBS; Tennant et al., 2007)	The WEMWBS is a measure of victims'/survivors' mental wellbeing	14 items Three subscales: 1. affective-emotional 2. cognitive-emotional 3. psychological functioning	Victims/survivors of violence	Additional measures to capture victim/survivor experience of violence Children's safety, wellbeing and development Effects on family functioning	Need author permission for use No fee Self-administered Less than 10 minutes for administration	http://www.mentalhealthpromotion.net/resources/user-guide.pdf (See "Frequently asked questions" for conditions of use)
Kessler 6 (K6; Kessler et al., 2003)	The K6 is a simple measure of psychological distress	6 items Targets the following feelings: sad; nervous; restless or fidgety; hopeless; everything is an effort; worthless	Victims/survivors of violence	Additional measures to capture victim/survivor experience of violence Children's safety, wellbeing and development Effects on family functioning	Publicly available (must cite original publication: see link) No cost Self-administered Less than 10 minutes for administration	https://www.hcp.med.harvard.edu/ncs/k6_scales.php

Measure	Overview	Number of items and subscales	Target group	To be used in conjunction with	Access/administration	Link to measure/permissions/original publication
General Self-Efficacy Scale (GSE; Schwarzer & Jerusalem, 1995)	The GSE is a measure of victims'/survivors' wellbeing assessing a general sense of perceived self-efficacy, including coping and adaptation after stressful events	10 items	Victims/survivors of violence	Additional measures to capture victim/survivor experience of violence Children's safety, wellbeing and development Effects on family functioning	Publicly available (must cite original publication) No cost Self-administered Less than 10 minutes for administration	Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy Scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), <i>Measures in health psychology: A user's portfolio</i> . (pp. 35-37). Windsor, UK: NFER-Nelson. http://userpage.fu-berlin.de/~health/engscal.htm
Multidimensional Scale of Perceived Social Support (MSPSS; Zimet et al., 1988)	The MSPSS Scale is a measure of victims'/survivors' social support	12 items Three subscales: 1. family 2. friends 3. significant personal relationships	Victims/survivors of violence	Additional measures to capture victim/survivor experience of violence Children's safety, wellbeing and development Effects on family functioning	Publicly available (must cite original publication: see link) No cost Self-administered Less than 10 minutes for administration	https://www.researchgate.net/publication/311534896_Multidimensional_Scale_of_Perceived_Social_Support_MSPSS_-_Scale_Items_and_Scoring_Information
Rosenberg Self-Esteem Scale (Rosenberg, 1965)	The Rosenberg Self-Esteem Scale is a popular measure of global self-esteem	10 items	Victims/survivors of violence	Additional measures to capture victim/survivor experience of violence Children's safety, wellbeing and development Effects on family functioning	Publicly available (must cite original publication: see link) No cost Self-administered Less than 10 minutes for administration	https://www.yorku.ca/rokada/psycytest/rosenbrg.pdf Rosenberg, M. (1989). <i>Society and the adolescent self-image</i> (Rev. ed.). Middletown, CT: Wesleyan University Press

Measure	Overview	Number of items and subscales	Target group	To be used in conjunction with	Access/administration	Link to measure/permissions/original publication
Children's safety, wellbeing and family functioning						
Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997)	The SDQ is a brief emotional and behavioural screening questionnaire for children and young people. The tool can capture the perspective of children and young people, their parents and teachers	25 items Five subscales: 1. emotional symptoms 2. conduct problems 3. hyperactivity/inattention 4. peer relationship problems 5. pro-social behaviour	Child victims/survivors Victims/survivors of violence	Additional measures to capture victim/survivor experience of violence Adult victim/survivor safety, wellbeing and freedom	Document under copyright (need permission to use) No cost Self-administered Less than 10 minutes to administer	Contact information for permissions: youthinmind@gmail.com https://depts.washington.edu/dbpeds/Screening%20Tools/Strengths_and_Difficulties_Questionnaire.pdf
Personal Wellbeing Index-School Children (PWI-SC; Cummins & Lau, 2005)	The PWI-SC measures health and personal wellbeing in school-aged children and adolescents	7 items Seven domains: 1. standard of living 2. health 3. life achievements 4. personal relationships 5. personal safety 6. community connectedness 7. future security	Child victims/survivors	Additional measures to capture victim/survivor experience of violence Adult victim/survivor safety, wellbeing and freedom	Publicly available (must cite original publication) No cost Self-administered Less than 10 minutes for administration	Instructions for use: http://www.acqol.com.au/uploads/pwi-sc/pwi-sc-english.pdf International Wellbeing Group. (2013). <i>Personal Wellbeing Index</i> (5th ed.). Melbourne: Australian Centre on Quality of Life, Deakin University

Note: ^a Should not be used in isolation.

ANROWS

ANROWS

AUSTRALIA'S NATIONAL RESEARCH
ORGANISATION FOR WOMEN'S SAFETY

to Reduce Violence against Women & their Children

