Author/s:
Stephenson, E;Mikolajczak, G;Ryan, M;Fisher, AN;Hayes, J;Sojo, V;Weaving, M;Tanjitpiyanond, M

Title:
A framework for evaluating women's leadership programmes

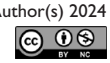
Date:
2024-01-01

Citation:
Stephenson, E., Mikolajczak, G., Ryan, M., Fisher, A. N., Hayes, J., Sojo, V., Weaving, M. & Tanjitpiyanond, M. (2024). A framework for evaluating women's leadership programmes. Evaluation, 31 (1), pp.111-141. https://doi.org/10.1177/13563890241284626.

Persistent Link:
https://hdl.handle.net/11343/354733

Article

Evaluation
1–31



DOI: 10.1177/13563890241284626
journals.sagepub.com/home/evi


A framework for evaluating women's leadership programmes

Elise Stephenson
Australian National University, Australia

Gosia Mikolajczak
Australian National University, Australia

Michelle Ryan
Australian National University, Australia

Alexandra N. Fisher
Australian National University, Australia

Jack Hayes
Australian National University, Australia

Victor Sojo
University of Melbourne, Australia

Morgan Weaving
Stanford University, USA

Mai Tanjitpiyanond
Independent research

Abstract

The increase in women's leadership programmes has been paralleled by a growing demand on behalf of private and public sector stakeholders for pragmatic reporting on impact assessment. Yet, there remains limited research and tools available for conducting evaluation of women's leadership programmes. Based on an extensive review of the literature and co-design with a sub-national Australian government, we develop an evaluation framework for women's leadership

Corresponding author:
Elise Stephenson, Australian National University, Canberra, ACT 2601, Australia.
Email: elise.stephenson@anu.edu.au

programmes. Our framework identifies critical measures and indicators at a micro, meso and macro level that can be used by researchers and practitioners to measure the impact of women's leadership programmes on women (and minoritised genders), organisations and structures – wider gender equality. We argue that evaluation is valuable in understanding the impact that women's leadership programmes have as well as the impact they do not have – providing more rigorous analysis of the sometimes tenuous impacts of women's leadership programmes. Such evaluation is critical to identifying women's leadership programmes' roles in tackling gender inequalities and women's continued under-representation in leadership. It may also help identify other interventions and policy and legislative change needed to bring about a desired change.

Keywords
evaluation, gender equalities, women's leadership, women's leadership programmes

Introduction

Women's leadership programmes (WLPs) are a popular avenue through which the public and private sector attempts to address gendered inequalities in leadership. The increase in gender equality (GE) and gender-specific leadership programmes have been paralleled by a growing demand on behalf of industry stakeholders for pragmatic reporting on impact assessment (Reale et al., 2014). However, the link between increasing individual leadership capacity and the impact on broader organisational and societal outcomes is tenuous (Njah et al., 2021). Indeed, there is limited research on the programmes, workshops and accelerators designed to address the under-representation of women in leadership – including the evaluation of their impacts and effectiveness. Through this article, we problematise this lack of evaluation for WLPs. While it is recognised that WLPs are only one part of a balanced policy response to gender inequality, without evaluation they may be over-relied upon as a panacea for a full range gender inequality-related grievances that they cannot realistically address. Evaluation also helps us to be better reflective of the role they *do* play, and what kinds of complementary policy or tweaks may help to move them from being a primarily 'fix women' approach to inequality to a 'fix systems' approach.

While there is substantial academic literature analysing, evaluating and improving the efficacy of leadership programmes in general (Garman et al., 2021; Mongon and Chapman, 2012; Streeton et al., 2021) and some research on industry-specific GE policies designed to improve representation of women (Kalev et al., 2006; Timmers et al., 2010), this evidence tends not to extend to WLPs. The available literature tends to focus on individual intervention or experience (Chasserio and Bacha, 2023; Debebe, 2011; Isaac et al., 2012; Sinclair, 1997) or a specific sector (Hopkins et al., 2022). For instance, in their review of GE evaluation programmes and policy interventions Schmidt and Cacace (2017) cite a lack of evidence and oversimplification of approaches. There is, therefore, a clear need for insights on evaluation methodology for WLPs from academic, industry and government perspectives. Gardiner et al. (2023) recognise in their systematic review of WLPs that there is an emphasised need for 'enhanced methodological and theoretical rigour to guide the development of future women's leadership programs' (p. 1), a call to which this evaluation framework responds.

In this article, we develop the following methodological framework to measure the individual, organisational and structural GE outcomes of WLPs. We draw on a systematic, qualitative literature review and collaboration with an Australian sub-national (state) government department to co-design a framework for evaluating WLPs. We identify two main issues to be addressed in evaluation frameworks. First, although many WLPs focus on the individual impact on women

as leaders, the multilevel nature of leadership processes and the wider impacts on organisational and structural/societal gender inequalities are infrequently addressed or measured, creating an important omission. Second, without an intersectional gendered lens, evaluation frameworks lack the nuance and context to substantiate and improve the impacts of existing WLPs. We therefore provide a rigorous method of assessing WLPs and offer recommendations for how current and future WLPs may progress beyond 'fixing women' to address systemic barriers and achieve greater representation across gender. We also reinforce others' findings that WLPs are one solution among many policy options organisations can use to address gender inequalities.

In this article, we first cover (1) the ongoing under-representation of women in leadership positions and the consequent rise in WLPs, (2) the assumed theory of change (ToC) behind many WLPs and (3) current best practice evaluation frameworks for leadership programmes more generally as well as for WLPs, specifically. We used learnings from this literature review to develop our WLP evaluation framework, which was tested with data from several WLPs administrated by an Australian sub-national government department, before being further refined for publication. The scope of this article covers the development and key learnings from applying the WLP evaluation framework. This article is not designed to share the specific outcomes of any given programme evaluations, nor to determine whether WLPs are the most efficacious tool in addressing gender equality. Our contribution as such hopes to expand the global knowledge on ways to measure the effectiveness of WLPs at a micro, meso, macro and even meta level. We argue that evaluation is valuable in understanding the impact that WLPs have as well as the impact they do not have – providing more rigorous analysis of the sometimes tenuous impacts of WLPs. Such evaluation is critical to identifying WLPs' strengths and weaknesses/limitations in tackling gender inequalities and women's continued under-representation in leadership. While this article is meta-evaluative, focusing on offering an evaluation framework for WLPs rather than drawing conclusions as to the role of WLPs in addressing gender equality, WLPs cannot be the sole intervention. As such, the article recognises the complexity of interventions for gender equality, and attempts to engrain structural evaluation questions, criteria and indicators that are often absent from WLP evaluation. Through doing so, the framework may also help identify other interventions and additional policy and legislative change needed to bring about a desired change, once key gaps and/or opportunities are revealed.

Literature review: Understanding women's under-representation in leadership and the rise of the Women's Leadership Programme (WLP)

Despite significant progress in women's labour force participation, workplace leadership remains gendered (World Economic Forum, 2023). Improved gender representation in leadership is associated with a range of positive outcomes, including that organisations are more likely to implement policies and practices that further support gender equity (Fine et al., 2020; Terjesen et al., 2009), facilitate workplace cultures that are more receptive to women and gender diverse leadership (Ely et al., 2011) and be more inclusive of other forms of diversity in the workplace more generally (Hoyt et al., 2016). The under-representation of women in leadership roles highlights ongoing gender biases that continue to uphold and reinforce the norms, policies and legislation that facilitate career progression for men over women and other genders. It impedes progress towards workplace GE by limiting the extent to which women leaders can affect structures, policies, and practices in addition to setting workplace culture and acting as role models within the organisation (Cook and Glass, 2015).

Factors at a micro, meso and macro level (otherwise classified for our purposes as individual/interpersonal, organisational and structural levels) challenge women's advancement towards leadership positions. While we recognise there are many ways to define micro, meso and macro levels, according to the literature and our experience co-designing this framework with government, we settled on these definitions as the most widely understood and actionable levels for evaluation (Schmidt and Graversen, 2020). Structurally, psycho-social processes like stereotypes and pervasive gender bias inhibit women. Organisational practices and policies can reinforce these, with challenges often relating to a masculine work culture, gendered workplace norms and exclusionary networks which may be critical for career stability and advancement. At an interpersonal level, there are factors such as caring responsibilities, and individual factors like self-efficacy beliefs and confidence affect women's choice to apply for leadership roles. Factors across micro, meso and macro levels are also interconnected, for example, masculine leadership norms (e.g. think-manager think-male associations, (Schein et al., (1996)), may bias the recruitment process in favour of men leading to an overrepresentation of men in leadership and discouraging women from applying for those positions.

Women from minoritised backgrounds (e.g. from culturally and linguistically minoritised, CALM, backgrounds) face additional challenges and barriers progressing to leadership positions (Skouteris et al., 2023). There is evidence that CALM women do not feel that their cultural identities are valued at work (Women of Colour, 2021) and they perceive more barriers to career success and lower support for professional development compared to non-CALM colleagues (Key et al., 2012). Thus, any initiative towards achieving GE in the workplace must remain cognisant of these intersecting factors.

Given this complex and interrelated context for women's leadership, WLPs rose out of a need to address the systemic under-representation of women in leadership and are a tangible option for public and private sectors alike (Mousa et al., 2021). Many WLPs focus on equipping women with individual skills, networks and intangible characteristics like confidence, based on research that women often underestimate their own abilities and are less likely to assert themselves in the workplace compared to men (Herbst, 2020). While improving skills and individual opportunities is important, there is little evidence that WLPs tackle other systemic factors such as gendered norms and policies limiting women's career progression (Ely et al., 2011). For example, it is harder for women to make the strategic work connections to progress into leadership roles as the existing networks of leaders are dominated by men (McDonald, 2011). For this reason, WLPs have been criticised for their disproportionate focus on 'fixing women' (i.e. addressing individual barriers and skill deficits), which is relatively easy to achieve, instead of focusing on 'fixing the system' (i.e. addressing the systemic or cultural barriers), which is arguably harder to achieve (and, as we argue in this article, assess) but more important in helping women pursue and thrive in leadership roles (e.g. Burkinshaw and White, 2017; Howe-Walsh and Turnbull, 2014; Ryan, 2023). Indeed, Diehl and Dzubinski (2016) categorise common organisational strategies employed to tackle under-representation, with WLPs most aligning to 'technical: fix the women' approaches, as opposed to the other three common organisational strategies that cut across meso and macro levels as well: (1) technical: create equal opportunity; (2) social: celebrate differences; and (3) sociotechnical: revise work culture.

WLPs are often regarded as more effective for women as compared to all-gender programmes (Debebe et al., 2016); however, some researchers have raised concerns that focusing solely on gender may in fact hinder women's development. This view is based on an intersectional theoretical perspective that highlights how different social categories, such as race and class, intersect with gender to shape an organisation's culture and structure (Wong et al., 2022). By prioritising

gender over other identities, women's programmes may inadvertently reinforce the dominant organisational culture, which is often centred around the norms and experiences of White middle-class men, by encouraging conformity to masculine workplace norms. This perspective has been argued by Debebe and Reinert (2014) as well as Plantenga (2004), while Acker (2012) provides a theoretical basis for understanding the intersectionality of different social categories.

Studies suggest that the effectiveness of leadership development programmes may depend on the size of the participant group, with smaller groups often having more success in achieving learning objectives and greater impact on leadership development outcomes (Avolio et al., 2009). Furthermore, the design and delivery of the leadership development programme can also impact the success of the programme, regardless of the number of participants. For example, customization of the programme to the specific needs of the participants, the use of a range of teaching methods and involving participants in designing the programme are all factors that can contribute to a successful programme (Day and Dragoni, 2015). Logistics aside, WLPs remain a controversial issue with some evidence of stigmatisation and a reluctance of some women to attend (Devillard et al., 2012).

Pedagogical theories have also failed to keep pace with practice. WLPs tend not to be grounded in a coherent, theoretically based, and actionable framework for design and delivery. Lacking such a framework, many adopt an 'add women-and-stir' approach (Meyerson, 1998: 312), simply delivering the same programmes to women or other gender minorities that they deliver to ungendered development programmes. This approach assumes that gender either does not or should not matter for leadership development. Others take a different tack, adopting a 'fix-the-women' approach (Ely and Meyerson, 2000; Ryan and Morgenroth, 2024). These approaches assume that gender matters a great deal, but they locate the problem in women: Women have not been socialised to compete successfully in the world of men, and so they must be taught the skills their male counterparts have acquired as a matter of course. While both approaches may impart some useful skills and tactics, neither adequately addresses the organisational realities women face nor is likely to foster in participants a sustained capacity for leadership (Ely et al., 2011: 475).

However, some recent research suggests a more nuanced view of WLPs' potential impact. Preliminary evidence from STEAM (Science, Technology, Engineering, Arts and Mathematics) fields suggests that when strategically designed and implemented, WLPs, along with mentoring programmes, can potentially serve as 'accelerators' of structural chance (Johnson et al., 2023; Schmidt et al., 2018). This accelerator effect appears particularly pronounced when WLPs are combined with other organisational initiatives aimed at systemic change (Bilimoria et al., 2008). Thus, by moving beyond a simplistic 'fix-the-women' approach, well-designed WLPs may be able to address both individual skill development and broader organisational challenges. However, empirical evidence supporting these claims remain limited and are often context-specific. Moreover, it is crucial to recognise that while WLPs may have the power to initiate change, their effectiveness is likely limited without accompanying structural reforms. The true potential of WLPs may lie in their ability to raise awareness of systemic barriers and catalyse demand for more comprehensive organisational and societal changes. As such, WLPs should be viewed as one component of a broader strategy for achieving gender equity in leadership, rather than a standalone solution to deeply entrenched structural inequalities.

Overall, the demand for teaching leadership to women has far outstripped the pace of research and theorising on women's leadership development. While there are some indications that WLPs excel at improving the skills, networks and confidence of a small number of women, with a privileged group of women benefitting from individual mobility, there remain substantial gaps in how to measure broader impacts that WLPs might have – on organisations,

communities or structural equality, for instance. To create more impactful and sustainable progress towards gender equality, WLPs need to take into consideration the interplay between individual, organisational and structural barriers. By conducting such evaluation, organisations implementing WLPs may also find other gaps for interventions beyond WLPs that may be more effective in bringing about systemic change – or at least be complementary to the WLP. This is a key argument of our article, that while WLPs are a key point of analysis, they may not be the most effective tool to address gender inequality.

Our methodology

We co-designed, developed and delivered an evaluation framework and then pilot tested the impact of six WLPs administered by a sub-national Australian government department over the course of 18 months from January 2022 to June 2023. The WLPs included those focused on different types of leadership (communal, corporate), different ages and career stages (young leaders) and different demographics that could benefit from targeted programming (First Nations, cultural and linguistically diverse women). We conducted a systematic qualitative literature review of evaluation frameworks related to WLPs and Leadership Training Programmes (LTPs), followed by a review of existing evaluation programmes and a consultation with key stakeholders (programme managers of the WLPs, evaluation experts within the government department and third-party peak body organisations). Through this process, we identified a set of meta-dimensions and indicators that WLPs should utilise to understand and assess their impact on individuals (programme participants), the organisations or communities they are representing, and progress towards broader gender equality. We then pilot tested the evaluation framework across an initial four WLPs, which was extended to additional two WLPs and further refined. We discuss this process in greater detail throughout this article, including key learnings and reflections on the development and useability of the framework.

Relevant indicators for each programme were selected based on consultations with the government department and programme coordinators and designers. Although we sought to quantify the impact of WLPs on organisations and at a structural level, significant data gaps and project scope (short timeframe, for instance) hampered our ability to analyse impact in these domains. Specifically, we did not have access to organisational data. Many participants of the WLPs were not affiliated with specific organisations, or if they were, data on the organisations was not known or not shared. Similarly, while we were able to gain a brief outline of data on broader levels of GE in Australia at the time of evaluation, causal complexity and the longer timeframes needed to measure the impact of WLPs at a structural level limit the feasibility of evaluation at this level. Nonetheless, this initial structural level data was gathered to be used as baseline data to evaluate changes across time.

This article focuses on the methodology we used to develop the framework, its application and key learnings we gained in that process, in the hope it can be used by other researchers and practitioners that seek to evaluate WLPs' impact. Due to the confidential nature of the data collected in the pilot, we did not include the details of the specific programmes being evaluated.

What we considered in developing the evaluation framework

There were several key considerations that came to the fore across the literature relevant to developing the evaluation framework. This includes the lack of evaluation frameworks for gender-specific leadership programmes, the need for multi-level evaluation, use of indicators,

the need for a ToC, the importance of establishing a timeline for evaluation, difficulties in establishing causality and other general limitations and challenges raised by the literature.

Creating a framework for evaluating gender-specific programmes

Non-gender specific LTP provide a template of evaluation upon which WLP evaluations may be modelled. These, however, have several limitations including a tendency to focus on short-term (1–2 years post programme) impact at the individual level (Njah et al., 2021). A longer far-sighted time framed analysis (2–5 years post programme) would offer a better understanding of broader institutional changes across multiple institutions and sectors. In implementing our framework, we found that most of the WLPs that we evaluated focused on creating short-term impacts at the individual level, that is, focusing on within-person change among individual women from disparate organisations. This included change within individuals participating in the training, but also that individuals rather than, for example, multiple women from a given organisation or community were targeted by the programmes, limiting the possible organisation-level impacts. This limited our ability to evaluate the WLPs at the organisational and structural level. To better evaluate organisational and structural change, a longer time frame of analysis and focus on creating broader institutional impacts would be required. Implementing a ToC, defined simply as a 'descri[ption] of what the leadership training program is trying to achieve and how it hopes to get there' (Njah et al., 2021: 4) may be a crucial tool in planning how the program can create broader institutional impacts. The time frames of evaluation and a ToC therefore emerge as important features of successful evaluation frameworks for LTPs.

Levels of evaluation

In analysing more specific literature on WLPs, recurring methods of evaluation become evident. Previous research (Goyal et al., 2010; Schmidt and Graversen, 2020) indicates that effective evaluation must be conducted at the micro (individual), meso (organisational) and macro (structural) level to be most effective (this maps onto the individual, organisational and structural indicators in the current framework). While these levels are undoubtably interconnected, and organisational and structural change can take place across the three levels, segregating analysis at these three levels allows for more effective and straightforward evaluation across a range of datasets (Schmidt and Graversen, 2020: 7).

1. Micro (individual level) including individuals and teams
2. Meso (organisational level)
   (a) Institutional rules
   (b) Organisational structures
   (c) Organisational cultures
   (d) Organisational processes
3. Macro (structural level)
   (a) Rules, incentives, structures and processes at the regional, national and supranational level

Evaluation criteria for meta-evaluation

Our view is that an analysis of micro, meso and macro impacts of WLPs is incomplete without a meta-evaluation. An analysis of the philosophical ideas, programmatic ideas and policy

ideas underpinning WLPs, as part of an evaluation process, would provide very rich insights to help understand why and how certain programme objectives have (or have not) been chosen, pursued and achieved. *Philosophical ideas* here represent the broad concepts related to values, moral principles or ideologies about what is desirable and undesirable in a society that frame the thinking and actions of policymakers when they decide what programmes of interventions to develop and how to implement these programmes via policies to address an issue. Importantly, these philosophical ideas are typically deep-seated and slower to change (Schmidt, 2011). For instance, WLPs could be underpinned by values of equity or of equality leading to the development of different programmes of work (Williamson et al., 2024).

These philosophical positions in turn inform the development of *programmatic ideas*. The combination of broad goals, more specific objectives, available instruments and implementation approaches to achieve some ultimate social value represents programmatic ideas (Schmidt, 2011). Here, we are talking about a bundle of policies that are more or less coordinated to achieve common goals (Thomas and Turnbull, 2018). For example, the rollout of WLPs could be one of the policies that organisations implement as part of a set of longer term actions to address gender inequality, alongside other policies such as flexible work arrangements, family friendly policies, targets and quotas, and prevention of sexual harassment. These policies can be implemented as part of a planned or a more emergent process, in an incremental or revolutionary way in certain institutional conditions (Sojo et al., 2022; Thomas and Turnbull, 2018). An evaluation of WLPs requires their consideration in the broader programmatic environment, analysing institutional context and policies with which WLPs could have synergistic effects or that could hinder their implementation.

Finally, WLPs could be considered a *policy idea*. Policy ideas can be analysed at many levels. Here, we argue that policy formulation and implementation are two basic dimensions to consider (Schmidt, 2011). Some basic aspects of policy ideas that need to be considered in a meta-evaluation of WLPs include the ToC that was followed, stakeholders consulted and involved in the rollout, allocated resources, guaranteed fidelity strategies, issues that emerged during implementation and how they were resolved, and dimensions of programme evaluation.

This meta-approach to WLPs is consistent with recognising the complexity of delivering GE change in organisations and at a structural level and fits with the broader literature which advocates for reflexivity in the change process (Wroblewski and Palmén, 2022). It also fits with broader findings of the structural difficulties inherent to some GE-related interventions – largely, that 'gender equality continues to be perceived as women's work', drawing attention to who designs, implements and is responsible for the delivery of WLPs may reinforce systemic inequalities (Clavero and Galligan, 2021: 1128).

Indicators

The use of indicators for the purposes of data collection and facilitation of evaluation is another frequently recurring theme. Selection of indicators is key: They determine the critical components in creating a pathway of change (Njah et al., 2021) and allow for the measurement, monitoring and evaluation of the efficacy of GE policies (Fitzsimmons et al., 2020). Previous evaluations have used several different indicators, with Schmidt et al. (2018) using 692 indicators in their analysis across personal to workplace conditions. While using 692 indicators may be comprehensive, the useability of such a long assessment is low, and indeed the categories and indicators used to evaluate WLPs should differ dependent on the desired programme outputs. Njah et al. (2021) conclude that the timeframe of evaluation (1–2 years post programme, or 3–5 years post programme) will determine the number of indicators used, with longer time frames requiring fewer

indicators for mapping institutional change. In attempting to construct indicators that have relevance across a range of WLPs as well as measure against individual, organisational and structural GE goals, we have selected a smaller number of indicators for this research, with a focus on narrowing in on the most important and useable indicators for our purposes.

Theory of change

A ToC focuses on mapping out or 'filling in' what has been described as the 'missing middle' between what a programme or change initiative does (its activities or interventions) and how these lead to desired goals being achieved (Center for Theory of Change, 2022; Mayne and Johnson, 2015; Rog, 2012). A ToC is 'an outcomes-based approach which applies critical thinking to the design, implementation and evaluation of initiatives and programs intended to support change in their contexts' (Vogel, 2012: 3). In essence, a ToC allows organisations to get to the root cause of the issue they are aiming to address and map how their chosen policy programme aims to influence or change that root cause. In the context of WLPs, a ToC should be used to ensure that a clear path can be evidenced between undertaking a WLP and achieving a desired result.

Once a WLP's key components have been identified and its theorised causal pathways mapped and articulated, decisions can be made about which components and pathways of the programme are of most interest (Njah et al., 2021). For example, Schmidt and Graversen (2020) based their ToC along three main axes: concept analysis, implementation analysis and effect assessment, relating to the analysis of the prerogatives of the programme, analysis and evaluation of the implementation of these prerogatives and a post-programme evaluation of outcomes, respectively. Goyal et al. (2010: 7) summarise the benefits of grounding evaluation frameworks in a ToC: in essence, a ToC facilitates the development of a shared understanding of how and why a programme creates change; provides a conceptual framework for monitoring, evaluation and learning around whether a programme works; serves as a communication tool to explain programmes and impacts to stakeholders and enables constructive feedback on programme design and delivery processes.

The lack of ToC can create a challenge for programme evaluation as it is difficult to assess progress without a measurable goal (and a plan for achieving said goal). In addition, if a stated goal of a WLP is to have an impact on organisational or structural inequalities, a WLP may not be the right policy option, unless the WLP has a specific organisational or structural focus. Ultimately, a ToC should guide WLP evaluation.

Establishing a timeline for evaluation

Assessing the impact of a programme is highly dependent on the timing of evaluation (Miles and Cunningham, 2005; Schmidt and Cacace, 2017). As Reale et al. (2014: 37) underline, the relationship between the *impactor* and *impacted* is fairly direct and often ignores the fact that there are many intervening factors/variables. If impact is assessed almost immediately or too late after the finalisation of the programme, stakeholders may not link the achieved effects to the programme itself (Bell et al., 2011). Moreover, data may need to be collected over a longer period so that rigorous and robust impact assessments can be realised – in other words, ideally, evaluation is not a one-off action but is revisited over time for longitudinal impacts. In the context of developing our evaluation framework, we have sought to provide guidance on the most relevant time to conduct evaluation, depending on the type of impact sought to measure, and the indicators used.

Establishing causality

While evaluation is useful for learning and improvement, accountability and enlightenment to guide further action, and should not strictly focus only on controlling variables for establishing and attributing causal links, we still argue that understanding causality is important to evaluation that takes place. This is a key driver behind much critical feminist interventions – that the research and learning done seeks to reduce gender inequalities and move societies one step closer to equality. As such, a consideration of evaluating leadership programmes which aim to increase diversity in leadership is to establish whether programmes indeed *cause* positive changes among the aspiring leaders participating in the programmes, the organisations they work in and society more broadly. Even if not all WLPs aim to address structural-level changes, it remains relevant to consider: Did the WLP itself cause change, was it something else, was it a combination of factors, and regardless – how would one measure causality in a framework? Due to a range of limitations in study designs, existing leadership training studies vary considerably in the extent to which they can establish specific causal effects (Podsakoff and Podsakoff, 2019). This is likely because leadership training studies often occur in an applied context, rely upon client organisations and therefore may have little scope to demand rigorous research design. We acknowledge that these contextual constraints are often outside of researchers' control and believe that well-designed correlational studies that temper conclusions accordingly can make significant contributions to the cumulative body of research. Still, it is important to outline some best practice standards of leadership evaluation studies, to guide researchers who have the goal and resources to investigate the causal role of leadership programmes.

To establish causality, it is first important to establish baseline data – the conditions or status of individuals, organisations and structures pre-intervention. In a best-case scenario, WLPs should also incorporate a control condition (that is, a group of similar participants who do not receive the leadership training) to establish that any positive changes among participants in the leadership programme were not found among participants in the control condition and were thus due to the leadership programme specifically (Hariton and Locascio, 2018; Rubin, 1975). If employing a control condition, participants should be randomly allocated to either the control condition or the leadership programme condition from a broad pool of individuals, to ensure there are no systematic differences between the two groups. In addition, researchers should try to remove contact between leaders and team-members in the leadership programme and in the control condition, to avoid cross-contamination effects (Martin et al., 2021).

We acknowledge these recommendations may be hard to achieve in applied research, and that there is a tension between what researchers wish to do and what study designs they are able to employ. We also acknowledge that there are concerns in the literature regarding what is gold standard or best practice to achieve in evaluation. As such it might be best to think pragmatically about what can be achieved with the evaluation, draw on guidance in terms of any design choices that could improve evaluation, and be upfront with any limitations. Because a key goal of programme evaluations is, in many cases, to establish causal effects, the above design considerations may be useful to inform choices about research design for WLP evaluation – even if not strictly adhered to.

Limitations and challenges

Evaluation of leadership programmes comes with a set of limitations and challenges. The choice of what, when and how in the context of evaluation can affect the evaluated programmes and their participants (Stone, 2016). WLPs face their own series of critical issues

due to their gender specificity, including the risk that stakeholders (programme managers, for instance) do not understand the politicised context in which evaluations take place (Sielbeck-Bowen et al., 2002), there is sometimes limited institutional capacity for gender and gender-sensitive evaluation design and implementation (Espinosa, 2013) and there can be an evaporation of the gender dimension of programmes during the implementation (Moser, 2005). Moreover, assessing impact may suffer from a lack of data and indicators. GE researchers have pointed out the need for more sophisticated frameworks and methodological diversity and suggest going beyond traditional impact indicators and identifying fewer tangible impacts (Bell et al., 2011; Molas-Gallart and Tang, 2011). These considerations have played into the development of our evaluation framework, which we describe in the next section.

The framework

Our framework encompasses the three core levels of impact (structural, organisational and individual) that we believe WLPs should be measured against a range of different indicators and measures, as outlined in Figure 1. A key assumption of the evaluation framework is that, while the three levels of impact are interconnected, they are not always mutually reinforcing, as improvements in one will not necessarily result in improvements in another. The framework also includes overall programme assessment, in order to understand and evaluate programme-specific logistical aspects (completion rates, etc.), as well as meta-evaluative indicators.

It is important to note that while we developed this framework in response to co-design with a sub-national government department and piloting with a series of programmes, this framework can and should be adapted as per requirements. It is thus provided as a useful guide, and an insight into our own thinking when developing a way of evaluating a range of WLPs.

Structural-level indicators

The broader societal context (the 'macro' level) that individuals and organisations operate in can influence their opportunities, available resources and barriers to progress. There is often an implicit, if not directly stated, assumption that WLPs will contribute to broader structural GE (as was the case for the WLPs we evaluated in our pilot). There are many methodological challenges to measuring WLPs' impact at a structural level, as highlighted in the section on establishing causality. However, tracking changes at a structural level of gender equalities alongside the implementation of programmes can be useful at minimum in understanding changes in the wider social context. As such, by measuring these indicators, the intention is not necessarily to attribute causality between WLPs and structural changes, but rather maintain awareness of a wider status quo co-occurring with WLPs, as well as any changes over time.

We created an initial set of GE indicators most pertinent to the WLPs studied. We reviewed existing databases to identify indicators of structural and social gender equality. This was further broken down into three core aspects of women's leadership that we felt most pertinent to measure at an overall, structural level: (1.) Representation of women leaders (in executive leadership roles, board membership, etc.), (2.) Experiences of women leaders and (3.) Recognition of women leaders.

1. *Structural equality*. These indices capture the proportion of women in formal leadership positions (e.g. executives, CEOs, managers), on boards, in overall employment, at different levels of leadership and across different portfolios and types of work. Our

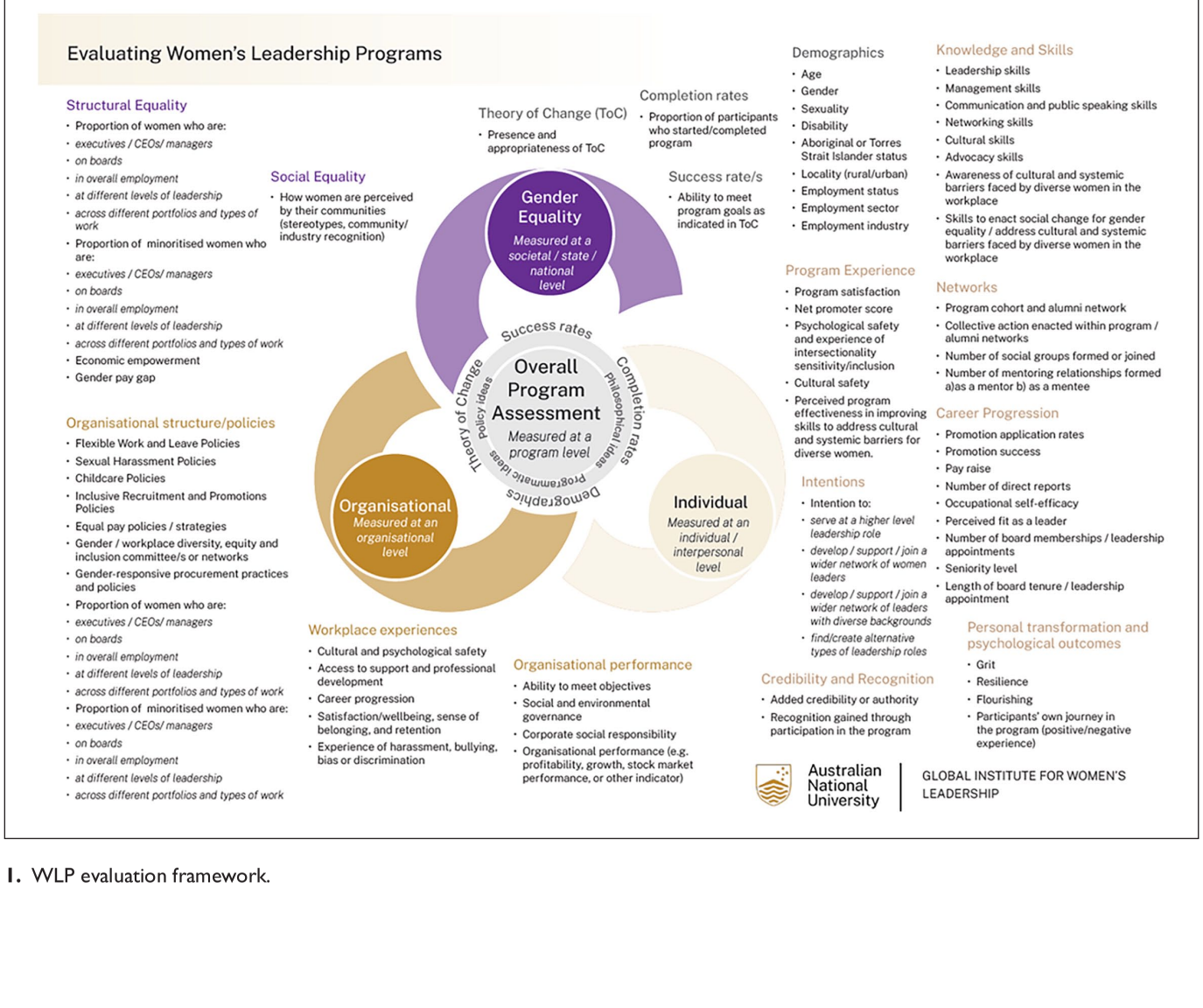


Figure 1. WLP evaluation framework.

framework recommends attention be paid to indicators of GE for women generally and women and gender minorities specifically. Longitudinal data on economic empowerment and the gender pay gap provide additional indicators to trace improvements in GE at the societal, state and national level. The under-representation of women in leadership positions has a compounding effect: a lack of gender diversity contributes to the 'token status' of relatively few women in leadership, which further impedes the improvement of gender equity in these positions (Chang and Milkman, 2020). As more women complete WLPs and become agents of change for GE, this should be reflected both in an increased representation of diverse women in leadership roles and in more equitable leadership when this framework is used, and impacts measured over time.

2. *Social equality*. These indices assess women's experiences working towards more senior leadership roles. While tracking changes in the proportion of women in leadership allows for the assessment of the rate of progress towards GE, it does not capture the barriers and challenges faced by diverse women aspiring to be leaders and those occupying leadership positions. For instance, women often face the 'glass ceiling' that prevents them from progressing to leadership positions (Cotter et al., 2001), followed by the 'glass cliff' that may impact their experiences of precarity once they reach leadership (Ryan and Haslam, 2007). For these reasons, we have included indicators measuring changes in the prevalence of gender stereotypes related to work and leadership, to monitor whether there are improvements in experiences of women working towards higher roles. Relatedly, the framework includes indicators for understanding recognition of women in leadership. Awards and credentials recognising women in leadership have been included as useful measures of social equality, both in how these awards are valued by the recipients and by the broader industry or community. It should be noted, however, that there is an important distinction between increasing the number of 'awards for women' and 'awards granted to women'. Context is therefore important in evaluated changes. We also recognise that there are other ways to measure women's recognition in leadership that could be included as specific, tailored indicators per the WLP analysed. Rather than being limited or prescriptive to these indicators solely, we encourage users of this framework to adapt social equality measures relevant for them. For instance, taking into account awards and credentials was important for some of the WLPs we evaluated as that was an explicit aim of the programme – to see more women awarded and recognised.

Organisational-level indicators

Organisational-level indicators assess changes at the organisational level such as the reduction of gender bias in organisational policies and procedures. This can be considered the 'meso' level of analysis. It is a relevant component of the framework as, if more women participants are promoted to leadership roles in their workplaces, it is often assumed that they will become agents of change, driving their organisations to become more inclusive. That is, women's formal and descriptive representation is often assumed to have an impact on substantive representation and symbolic representation (Pitkin, 1972). While the literature is mixed on the degree to which this is true, the *belief* that women's participation in WLPs will have impacts on their organisations is relevant to test, in part to determine whether it is the case, and in part, to determine whether there are ways to strengthen WLPs so that they *do* have an impact on an

organisational level. Therefore, tracking changes at an organisational level alongside the implementation of programmes can be useful at minimum in understanding circumstances and/or changes in the wider organisational context in which women are operating.

The following organisational-level indicators are included in our framework:

1. *Organisational structure/policies*. These measures assess existing organisational policies, processes and practices which perpetuate gender inequity at different stages of the work cycle including recruitment, training, promotion, performance evaluation, pay and termination. For instance, women might be underrepresented in leadership roles as a result of organisational promotion policies that fail to consider caregiving labour outside of the workplace. Such policies are more likely to disadvantage women (Starmarski and Son Hing, 2015). Given many of the WLPs we evaluated sought to have an impact at an organisational level, under a theory that more empowered women might change their organisations for the better, we sought to use these indicators to determine over time whether WLPs made a difference to: flexible work and leave, sexual harassment, childcare, inclusive recruitment and promotion, equal pay, GE in the supply chain as well as whether the organisations have established Equity, Diversity and Inclusion Committees. Structural data on the presence of women in leadership is also a necessary pillar of this measurement, again under the assumption that participation in WLPs, over time, improves women's representation in leadership.
2. *Organisational performance*. These measures capture organisational performance such as the level of profitability, growth and stock market performance. Research suggests that gender equity (and diversity in general) is beneficial to organisations boosting productivity, innovation and, in some instances, overall profitability (Erhardt et al., 2003). While many organisations have a focus on the 'business case' for diversity and investing in WLPs or other initiatives for the benefits more women might provide in an organisation, it is important to note that measuring changes in organisational performance should be irrelevant to the goal of addressing women's systematic under-representation in leadership. We note that this information can be critical and demanded from organisations, however where it is used, we encourage the collection and use of this information with caution.
3. *Workplace experiences*. These measures assess participants' experiences in relation to gender bias, discrimination or a sense of cultural safety in their workplaces. These measures are proxies for assessing an organisational culture, that is beliefs, assumptions and values shared by members of an organisation. Organisational culture is important in shaping perceptions and behaviours that promote GE and affects the efficacy of organisational policies. Participants who completed WLPs might be expected to contribute to improving organisational culture that is more equitable and inclusive. Thus, they should report lower perceived gender bias and discrimination but a higher sense of cultural safety (for more reading on cultural safety and its relevance to WLPs, see, for instance: Debebe, 2011; Ryan, 2020).

At a practical implementation level, to include organisational indicators, WLP coordinators should consider partnering with specific organisations, ideally recruiting multiple participants per organisation, to create the conditions for a 'critical mass' and broader network opportunities enhancing collective action.[1] Data on participants' organisational affiliations may need to be collected and tracked over time, which can present some significant limitations. This

reinforces the importance of WLP design that considers what kinds of impact a programme seeks to have, and map appropriate pathways to achieve that impact through a ToC. Measuring organisational impact of WLPs without effectively building in pathways to create organisational impact can therefore be a major shortcoming of WLPs.

Individual-level indicators

These measures assess changes at the individual participant ('micro') level during and after the programme. These measures are those most commonly found in the literature and most often used in practice (as opposed to organisational or structural measures, given the individualised nature of many WLPs).

1. *Programme Experience*. These measures assess perceived programme quality and effectiveness, such as satisfaction with the programme and participants' willingness to recommend it to future cohorts. Measures such as 'psychological safety' are important for assessing whether participants felt comfortable during the programme (Nash and Moore, 2019), which may also be linked to the efficacy of the programme. That is, particularly when considering intersectionality and First Nations women (for instance) a lack of psychological and cultural safety may render the programmatic contents less or ineffective (for instance, see: Debebe, 2011; Ryan, 2020).
2. *Knowledge and Skills*. These measures are meant to be programme specific, reflecting the individual programme's content and the underlying ToC. For example, WLPs designed to get women into board leadership should focus on knowledge and skills necessary in the role of a board member. If a WLP is expected to have organisational and structural impact, knowledge and skills development (and the evaluation) might centre around the ability to influence policy, campaign, network with others or develop collective actions. This was a key reflection from our piloting – there was an express aim that WLPs foster organisational and structural change, and so as a result, programmatic elements also focused on building coalitions of support, how to make policy change and so on.
3. *Behavioural Intentions*. These measures are useful in evaluating whether programme participation encourages participants to seek opportunities crucial to their development as leaders (e.g. apply for promotion or for a board position).
4. *Networks*. This measure assesses whether participation in a programme has increased women's social network. Research suggests that women have more limited leadership social networks compared to men (Nash and Moore, 2019). These networks are important in providing support and guidance for women to develop leadership skills and creating leadership opportunities, as well as facilitating more organisational and structural-level impacts.
5. *Credibility*. This measure assesses whether participation in the programme enhances individual's professional credibility as a leader. Research suggests that women leaders are evaluated more critically than men, that is they need to be significantly more qualified to be seen as equally competent as men (Boldry et al., 2001; Yang and del Carmen Triana, 2019).
6. *Personal Transformation and Psychological Outcomes*. These measures assess grit or perseverance which was identified as an important attribute required for women to navigate unequal systems through the literature review and stakeholder interviews

(Koekemoer et al., 2023). Although these gender biases exist on a system-level and can't be 'fixed' by individuals, women's willingness to persist and maintain a positive perception of themselves are crucial in allowing them to seek leadership opportunities and thrive as leaders.

7. *Career Progression*. These measures assess women's progression in their careers and into leadership roles and include objective indicators such as promotion application and success, salary progression, number of direct reports, number of board membership appointments (as well as the length of appointments and seniority in the organisation). More subjective measures include perceived occupational self-efficacy and fit as a leader, which were particularly important in providing insights into cultural and psychological safety – as referenced above.

Overall programme indicators

These indicators assess programme goals, activities, delivery, participation rates and completion rates. These measurements are important in evaluating whether the intervention is successful in its setup and execution, taking account of the demographic profile of the individuals it is aimed towards. It is assumed that the effectiveness of a programme will have links to micro-, meso- and macro-level outcomes, justifying the need to capture the overall programme evaluation as part of the framework.

Four key elements of the overall programme indicators include:

1. *ToC*. Assesses whether a programme has a ToC, and if so, what and how appropriate its specified goals are. Users are encouraged to adapt evaluation measures/indicators as necessary to ensure meaningful measurement of programme against ToC goals. Evaluators are encouraged to think of a clearly articulated ToC as foundational to the following three overall programme indicators.
2. *Demographics*. Participant demographic information is necessary for assessing participant diversity and understanding the intersectional nature of challenges/opportunities/success/failure experienced by participants.
3. *Completion Rates*. These measures help in understanding who started and who completed the programme. Low completion rates may suggest a mismatch between the participants and the programme.
4. *Success Rates*. Success rate/s will depend on how 'success' is articulated in a programme's ToC. Low success rate/s may suggest a mismatch between the participants and the programme, or programme and ToC.

In addition, the three levels of meta-evaluation can be considered under this dimension: analysis of the philosophical ideas, programmatic ideas and policy ideas underpinning WLPs helps understand why and how certain programme objectives have (or have not) been chosen, pursued and achieved.

1. *Philosophical ideas*. Values, moral principles or ideologies about what is desirable and undesirable.
2. *Programmatic ideas*. The suite of interventions or policies that are more or less coordinated to achieve common goals.

3. *Policy idea*. Formulation and implementation, for example, the ToC that was followed, stakeholders consulted and involved in the rollout, allocation of resources, guaranteed fidelity strategies, issues that emerged during implementation and how they were resolved.

A note on intersectionality

WLPs must embed an intersectional approach in their very design, one that considers and accounts for the multiple, overlapping identities (e.g. race, class, religion, culture, etc.) that women hold – an approach that ignores these intersections of identity is unlikely to address depth and breadth of barriers that different groups of women face. Thus, intersectional considerations should be part and parcel of WLP evaluation. For instance, evaluation can include assessing whether and to what extent programmes are inclusive of different groups of participants as well as their impact on different groups (e.g. are they more likely to deter or encourage certain groups of women to seek leadership positions). To this end, indicators can be split to assess outcomes for different groups of women, whenever possible.

We recognise that some WLPs will be designed to address a specific minoritised community or demographic, with a focus towards race, class, sexuality, religion, culture, nationality or disability. Considerations of lived experience and minoritised context will therefore inform the importance of some indicators, or require bespoke design in collection, monitoring and tracing of these indicators. For example, collecting data from participants in WLPs that are designed for women living with disabilities, or sexual minorities, may require additional sensitivities and transparency regarding storage of this data. For evaluative purposes, successful mapping of progress at the meso and macro level should integrate intersectional context; for example, indicators at the meso level examining organisational structure and experience will be perceived differently by different backgrounds, for example, a First Nation woman might experience the double minoritisation of their gender and ethnicity in an organisational culture. Structural equality indicators should consider not only the proportional number of women in leadership, but whether there is a diversity of race, class, sexuality, religion, culture, nationality and disability among this cohort. Intersectionality must therefore inform the evaluative baseline for the use of this framework.

Table of measures and indicators

The following five pages lay out in table form the level of impact, indicators used, measures, time of measurement, and measurement type and methodology needed to evaluate the WLPs. Depending on the programme, some or all of the indicators and measures may be used. Tables include:

1. Programme Assessment
2. Impact on Gender Equality
3. Impact on Organisations
4. Impact on Individuals

It should be noted that the framework has been designed with adaptability and useability in mind, and so researchers and practitioners are encouraged to hone measures and indicators as per their WLP's ToC and evaluation needs.

	Measure			Time of measurement				Measurement type and method	
	Level of impact	Indicators	Measures	Before programme	After programme	6 months follow up	1 + Years follow up	Type of measurement	Data collection method
1	**Overall Programme Assessment**	Success rate/s	Ability to meet programme goals as indicated in ToC		✓			Result	Programme data
		Completion rates	Proportion of participants who started/completed programme					Result	Programme Data
		Demographics	Age					Result	Survey
			Gender					Result	Survey
			Sexuality					Result	Survey
			Disability					Result	Survey
			Aboriginal or Torres Strait Island status					Result	Survey
			Locality	✓				Result	Survey
			Employment status					Result	Survey
			Employment sector					Result	Survey
			Employment industry					Result	Survey
		Theory of Change (ToC)	Presence and appropriateness of ToC		✓			Result	Programme Data
2	**Gender Equality** *Measured at a societal / state / national level*	Structural Equality	Proportion of women / minoritised women who are: • *executives / CEOs/ managers* • *on boards* • *in overall employment* • *at different levels of leadership* • across different portfolios and types of work	✓			✓	Result	Public Data
			Economic empowerment					Result	Public data
			Gender pay gap					Result	Public data
		Social Equality	How women are perceived by their communities (stereotypes, community/industry recognition)	✓			✓	Result	Public data / Surveys
3	**Organisational** *Measured at an organisational level*	Workplace experiences	Cultural and psychological safety					Process	Survey
			Access to support and professional development					Result	Survey
			Career progression	✓	✓			Result	Survey
			Satisfaction/wellbeing, sense of belonging and retention					Process	Survey
			Experience of harassment, bullying, bias or discrimination					Result	Survey
		Organisational structure /policies	Flexible work and leave policies					Result	Organisation's data
			Sexual harassment policies					Result	Organisation's data

		Measure		Time of measurement				Measurement type and method	
			Childcare policies					Result	Organisation's data
			Inclusive recruitment and promotions policies					Result	Organisation's data
			Equal pay policies / strategies					Result	Organisation's data
			Gender / workplace diversity, equity and inclusion committee/s or networks	✓				Result	Organisation's data
			Proportion of women / minoritised women who are: • *executives / CEOs/ managers* • *on boards* • *in overall employment* • *at different levels of leadership* • *across different portfolios and types of work*					Result	Organisation's data
			Gender-responsive procurement targets/practices	✓			✓	Result	Organisation's data
		Organisational Performance	Ability to meet objectives					Result	Organisation's data
			Social and environmental governance					Result	Organisation's data
			Corporate social responsibility	✓			✓	Result	Organisation's data
			Organisational performance (e.g. profitability, growth, stock market performance or other indicator)					Result	Organisation's data
4	**Individual (part 1)** *Measured at an individual level*	Programme Experience	Programme satisfaction					Process	Survey
			Net promoter score		✓			Process	Survey
			Psychological safety and experience of intersectionality sensitivity/inclusion					Process	Survey
			Cultural safety					Process	Survey
			Perceived programme effectiveness in improving skills to address cultural and systemic barriers for diverse women					Process	Survey
		Knowledge and Skills	Leadership skills					Result	Survey
			Management skills					Result	Survey
			Communication and public speaking skills					Result	Survey
			Networking skills					Result	Survey
			Cultural skills	✓	✓	✓	✓	Result	Survey
			Advocacy skills					Result	Survey
			Awareness of cultural and systemic barriers faced by diverse women in the workplace					Result	Survey
			Skills to enact social change for gender equality/address cultural and systemic barriers faced by diverse women in the workplace					Result	Survey

		Measure		Time of measurement				Measurement type and method	
5	**Individual (part 2)** *Measured at an individual level*	Intentions	Intention to: • *serve at a higher level leadership role (short, medium, long term)* • develop / support / join a wider network of women leaders • develop/support/join a wider network of leaders with diverse backgrounds • find/create alternative types of leadership roles	✓	✓	✓	✓	Result	Survey
		Networks	Programme cohort and alumni network					Result	Survey
			Collective action enacted within programme/alumni networks					Result	Survey
			Number of social groups formed or joined (social identity mapping)		✓	✓	✓	Result	Survey
			Number of mentoring relationships formed (1) as a mentor (2) as a mentee					Result	Survey
		Credibility and Recognition	Added credibility or authority	✓	✓	✓	✓	Result	Survey
			Recognition gained through participation in the programme					Result	Survey
		Personal transformation and psychological outcomes	Grit	✓	✓	✓	✓	Process	Survey
			Resilience					Process	Survey
			Participants' own journey in the programme	✓ Motivations to enrol	✓ Programme experience	✓ Perceived career impact	✓ Perceived career impact	Process	Interview / Case Study / Survey
			Flourishing	✓	✓	✓	✓	Process	Survey
		Career Progression	Promotion application rates					Result	Survey
			Promotion success					Result	Survey
			Pay raise					Result	Survey
			Number of direct reports					Result	Survey
			Occupational self-efficacy		✓	✓	✓	Result	Survey
			Perceived fit as a leader	✓				Result	Survey
			Number of board memberships/leadership appointments					Result	Survey
			Seniority level					Result	Survey
			Length of board tenure / leadership appointment					Result	Survey

How to use the framework

This framework relies on mixed method data collection and analysis. The primary data collection tools used are programme surveys as well as longitudinal comparisons across wider in-country datasets. Such datasets in the Australian case might include data from the Workplace Gender Equality Agency, Australian Bureau of Statistics, the Organisation for Economic Co-operation and Development (OECD) and other government databases, however researchers and practitioners are encouraged to map the relevant data sources for their context and WLP. This section outlines two key elements of data collection for our framework: a pre-programme survey and a post-programme survey.

Pre-programme survey

Pre-programme surveys assess participants' baseline in relation to their perceptions of their own abilities and leadership aspirations. Responses on these measures are used to compare to participants' responses in the post-programme survey to determine any changes in their perceptions and aspirations after programme completion. In general, we recommend the pre-programme survey is undertaken any time from a week to 1 day before the WLP commences. The survey should include:

- *Demographic Questions.* To understand how participants' diverse backgrounds may shape their experiences with the programmes, it is important to measure demographic questions. Participants' age, education, gender, LGBTQIA+, disability and other cultural identities.
- *Knowledge and Skills.* This is an initial assessment of participants' perceptions of their own skills in relation to leadership, management, communication and public speaking, networking, cultural competency, advocacy, awareness of cultural and systematic barriers faced by diverse women in the workplace as well as skills to enact social change to address gender equity. Participants can be asked questions like: 'How would you rate yourself on the following. . . [e.g. leadership skills]' and provided their response on a 7-point scale (1 = Very weak to 7 = Very strong).
- *Behavioural Intentions.* These measures assess participants career and professional aspirations before the programme commencement such as their intentions to apply for leadership roles, join or develop a wider network of (diverse) women leaders or to create an alternative type of leadership role for themselves. The latter intention is based on existing findings that leadership roles are often masculine in nature and thus women leaders need to craft their own leadership styles in the workplace (Eagly and Carli, 2018).
- *Social Network.* These questions ask participants if they have (and how many) sponsors and mentors to support them in their careers and their own role as mentors to potential mentees (e.g. Do you have a sponsor?).
- *Psychological Outcomes.* These measures relate to psychological transformation as a result of programme participation. This includes participant's perceived fit as a leader, their sense of occupational self-efficacy (or their confidence in their ability to navigate their job) and resilience.

Post-programme survey

The post-programme survey has questions from the pre-programme survey such as knowledge and skills, behavioural intentions, social network and psychological outcomes. In addition, there are new questions assessing participants' programme experience. In general, we recommend the post-programme survey is undertaken immediately, from 1 day to 1 week after the WLP commences, and then again longitudinally at specified increments (e.g. 6 monthly, yearly, 5-yearly).

Cultural safety and belonging. These measures assess whether the programmes were a safe space for participants. In particular, the measures ask how inclusive the programmes were for people from different minoritised groups based on gender, culture, sexuality and disability. Participants were also asked if they felt their voice was being heard in the programme. Responses were recorded on a 7-point scale (1 = Not at all to 7 = Very much so).

Perceived programme effectiveness. These indicators assess participants' perceptions of the programme effectiveness in addressing the following topics: (1) Leadership skills, (2) Networking, (3) Issues of gender inequality, diversity and inclusion in your workplace, (4) Systemic barriers in your workplace, (5) Organisational culture. Responses were recorded on a 7-point scale (1 = Not at all to 7 = Very much so).

Social network. In addition to the questions asked in the pre-programme survey, participants were asked additional questions about the relationships they developed with other women in the programme. In particular, they were asked about whether they still kept in touch with their programme's peers, their plans to engage in any activities to improve GE in Victoria together with other members of the programme cohort or alumni network as well as whether participants joined any working groups after programme completion.

Career impacts. These measures assess participants' perceptions of how the programme may have given them additional credibility in their career as well as greater recognition (e.g. Do you feel that the programme has given you additional credibility/authority as a leader?). Participants responded to a 7-point scale (1 = Not at all to 7 = Very much so). Participants were also asked to write an open-ended response about how they think the programme can help them with their career aspirations.

Organisational/community impacts. These are open-ended questions which ask participants to reflect on how the programmes can help them (1) make a stronger contribution to their organisation and/or local community as well as (2) progress GE, diversity and inclusion in their organisation and/or local community.

Bringing it all together: Insights and reflections from developing and applying the framework

In applying the Global Institute for Women's Leadership (GIWL) WLP Evaluation Framework to six WLPs across Australia, we gained several important insights and reflections which we describe in detail below. While there is no singular approach to WLPs, as WLPs should be

responsive to local needs, this meta-evaluation identified some key considerations for WLPs moving forward, including the following: WLPs must articulate a clear ToC and plan for evaluation; data collection must be mainstreamed into programme delivery; intersectionality must be mainstreamed throughout WLPs and evaluation to understand more specific and nuanced outcomes and experiences; thoughtful programme design embedding organisational and wider institutional partnerships, and longer evaluation timeframes, and complementary social interventions may have more success at creating and sustaining impact at organisational and structural levels.

Articulate a clear ToC and plan for meta-evaluation

Both a ToC and plan for evaluation are critical for achieving programme goals and greater alignment between individual, organisational and structural levels of impact. While all the WLPs we analysed aimed to support women into leadership roles, each programme differed in its approach on how this could be achieved. Outlining a clear ToC (either across programmes, or per programme) would have helped clarify how each programme contributes to women's leadership or wider GE. It would also allow for better evaluation of programme effectiveness. Relatedly, from the material available to us during our evaluation, it was not clear whether the WLPs themselves followed an evidence-based approach to programme content, which may impact on the degree to which they can fulfil their ToC and meet the stated goals. In contrast, an evidence-based approach, which involves women and other key stakeholders in programme planning stages, would allow for the identification of key barriers and programming to address those specific barriers, supporting overall programme development and responsiveness to local needs. In some cases, such a co-designed and evidence-based process may result in the realisation that WLPs are not the appropriate mechanism to achieve some of the desired impacts (structural or organisational level impacts in particular) and may necessitate the creation of additional policy or programmes to target under-serviced goals. Our analysis of WLPs and the relevant literature indicates that they are one of many policy options that – when working in tandem with other programmes and policies such as legal and policy reforms – can deliver more tangible impact across a range of indicators than they can deliver alone.

Mainstream data collection in programme delivery

WLPs need more comprehensive data and better designs to accurately assess their impact. One of the main limitations we found with our evaluation was that only small numbers of participants took part in evaluation (evaluations were not compulsory). In addition, a lack of organisational data hampered the ability of WLPs to connect programme implementation and participation to measures of organisational impact. A lack of causal clarity hampered the attribution of causation/correlations between WLPs and wider measures of GE and GE in organisations. Embedding WLPs within organisations via organisational partnerships is one potential strategy to address this issue (discussed below).

Furthermore, WLPs are often delivered to a relatively small number of women and evaluation of their impact typically involves a pre- and post-programme surveys whereby participants' average self-reported survey scores (e.g. satisfaction, confidence, skills) at each assessment point are compared. Such small sample sizes mean that statistical analyses of impact at the individual level are chronically underpowered, making it near impossible to

detect change with statistical significance, though perceptible changes in scores may still be considered meaningful. With only one pre- and one post-programme survey, it is also impossible to determine whether any observed change is lasting and attributable to the programme itself or the various external factors. Improving survey uptake, partnering with organisations and understanding causal limitations may improve programme evaluation and outcomes.

One way in which WLPs can improve their ability to assess impact is by including more assessment points in their evaluation plan. Including multiple assessment points before and after the programme can improve WLPs ability to detect meaningful individual change. It can also help to rule out natural fluctuation in participants' scores that may otherwise be misattributed to the programme by allowing natural variation to be accounted for in the statistical analysis (Lagarde, 2012). Another way WLPs can improve their ability to measure impact is by including a control or comparison group of women who do not take part in the programme. Observing increases in the self-efficacy of women in the WLP compared to a matched control group of women who are not enrolled in the programme would lend credibility that the WLP has the intended effect. In practice, however, it can be challenging to find an appropriate comparison group of women who are matched in terms of their demographics and organisational considerations. In such cases, comparisons to other available social metrics and databases can be a useful tool towards building an evidence base of programme impact.

Mainstream intersectionality throughout WLPs and evaluation

WLPs must consider how women's leadership experiences may be affected by identities beyond gender (e.g. race, class, sexuality, religion, culture, nationality, disability, etc.) and how some of these identities may be privileged and others marginalised within organisations and society more generally. Indeed, a one-size fits all approach is unlikely to benefit all women. We therefore recommend that WLPs expand their focus beyond gender to include additional dimensions of identity. To do this, WLPs need to collect a more comprehensive range of demographic information from their participants. Importantly, however, there needs to be caution in how such sensitive data is collected, stored and analysed, as WLPs with smaller cohorts risk identifying participants with unique combinations of identities (Global Partnership for Sustainable Data, 2022).

Embed research and programme partnerships

WLPs can increase their impact, and impact assessment, by partnering with organisations. Organisational partnerships would allow WLPs to identify the shared barriers that exist for different groups of women within specific organisations. For example, WLPs could be piloted within an organisation to gauge impact (Kempster et al., 2014). From there, WLPs could develop initiatives that not only aim to empower women but also transform organisations. Organisational initiatives may include training and education for managers, leadership development coaching, climate surveys, support networks and advisory councils for women and minorities and policies the support work–life integration (Bilimoria et al., 2008; Bilimoria and Liang, 2014). While this kind of integrated approach requires organisational buy-in and funding, it is probably one of the most promising strategies for facilitating long-term change (Bilimoria et al., 2008).

Partnering with organisations can also improve WLPs' ability to assess impact. For instance, if an organisation were to run WLPs annually, data from each cohort could be collated and analysed together to increase total sample size and the ability to map change. This approach would also allow for greater anonymity of individual participants during analysis and the sharing of results. Moreover, organisational partnerships offer a promising avenue through which organisational change can be measured via human resources data like performance rankings, salary and rates of turnover and promotion. It would also facilitate the within-organisation comparison of outcomes for women who participate in WLPs and those that don't.

In addition, when partnered with knowledge-partners, such as researchers, WLPs programme managers can be more certain that the programmes developed are evidence-based and best practice, to enable the highest chance of impact across desired measures.

Identifying intermediate effects and paths

It is noted that this framework may be useful for identifying intermediate effects and pathways to change. For instance, in the early years of analysis in a programme, it might be expected that only a limited number of impacts are witnessed. This might lead to iterative changes in programming that enable WLPs over time to get more effective. It might therefore be useful to think of a maturity curve when analysing WLPs. For instance, in the aims of getting a WLP off the ground and an initial evaluation conducted, organisations may seek to target a limited set off goals or objectives first. In subsequent years, both the programming depth and evaluation depth could be increased. Over time, the evaluation could be used to understand what design choices lead to stronger pathways to organisational and structural change, versus those that predominantly maintain impact at an individual level. For instance, feelings of safety and belonging, in our analysis, were critical particularly for First Nations women and women of colour, for whom the broader WLP, LP and societal contexts may be additively exclusionary. In this case, part of ensuring programme effectiveness for First Nations women and women of colour is cultural and psychological safety – that they feel able to participate, are heard and are supported for the duration of the programme. Although we are constrained by the space limitations of this article, this is a further suggested area for research.

WLP must be one tool among many in addressing societal gender inequity

As has been articulated throughout this article and in similar WLP academic literature, we reiterate that WLP should be one intervention among many in addressing social gender inequity. While improvements to structural and social (meso and macro) gender inequality may be mapped over time, it is unrealistic to expect significant improvement without WLP initiatives as complementary to other more structural-orientated interventions (see Palmén and Schmidt, 2019). This may only be achievable in collaboration between public and private sector interventions, and plottable over longer periods of time. Undertaking or tying in a meta-evaluation may better allow this analysis of the philosophical ideas, programmatic ideas and policy ideas underpinning WLPs, to provide insights to help understand why and how certain programme objectives have (or have not) been chosen, pursued and achieved.

Conclusion

Overall, this article has shared our development of the GIWL's Evaluation Framework for WLPs. Developed from a co-design process with a sub-national Australian government and extensive review of the literature, the GIWL WLP Evaluation Framework identifies critical measures and indicators at micro, meso and macro level that can be used by researchers and practitioners to measure the impact of WLPs on women (and minoritised genders), organisations and structures – wider GE. We argue that attention at design and delivery stages must be matched with attention at evaluation stages for WLPs to be most successful. Having dedicated theories of change would enable the organisations to maximise the role that WLPs do play and streamline evaluation.

We argue that evaluation is valuable in understanding the impact that WLPs have as well as the impact they do not have – providing more rigorous analysis of the sometimes tenuous link between WLPs and impacts. Such evaluation is critical to identifying WLPs' roles in tackling gender inequalities and women's continued under-representation in leadership. It may also help identify additional policy and legislative change, or other interventions, needed to bring about a desired change.

Declaration of conflicting interests

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

Funding

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

ORCID iD


Elise Stephenson https://orcid.org/0000-0002-3977-8464

Note

1. According to McKinsey and Company, at least 7% of the organisation needs to be involved in organisational change effort to produce successful outcomes (London et al., 2021). By creating this, the programmes will be better positioned to elicit change at the organisational level. In addition, it is recommended that the organisation is to assess other women employees (who did not complete the WLPs) or all employees in the organisation as a baseline to assess changes in participants who completed WLPs.

References

Acker J (2012) Gendered organisations and intersectionality: Problems and possibilities. *Equality, Diversity and Inclusion: An International Journal* 31(3): 214–24.

Avolio B, Walumbwa F and Weber T (2009) Leadership: Current theories, research, and future directions. *Annual Review of Psychology* 60: 421–449.

Bell S, Shaw B and Boaz A (2011) Real-world approaches to assessing the impact of environmental research on policy. *Research Evaluation* 20(3): 227–37.

Bilimoria D and Liang X (2014) Effective practices to increase women's participation, advancement and leadership in US academic STEM. In: Bilimoria D and Lord L (eds) *Women in STEM Careers:*

International Perspectives on Increasing Workforce Participation, Advancement and Leadership. Cheltenham: Edward Elgar Publishing, 146–65.

Bilimoria D, Joy S and Liang X (2008) Breaking barriers and creating inclusiveness: Lessons of organisational transformation to advance women faculty in academic science and engineering. *Human Resource Management* 47(3): 423–41.

Boldry J, Wood W and Kashy DA (2001) Gender stereotypes and the evaluation of men and women in military training. *Journal of Social Issues* 57(4): 689–705.

Burkinshaw P and White K (2017) Fixing the women or fixing Universities: Women in HE Leadership. *Administrative Sciences* 7: 30.

Center for Theory of Change (2022) What is theory of change. Available at: https://www.theoryofchange.org/what-is-theory-of-change (accessed 14 October 2022).

Chang EH and Milkman K (2020) Improving decisions that affect gender equality in the workplace. *Organizational Dynamics* 49(1): 100709.

Chasserio S and Bacha E (2023) Women-only training programmes as tools for professional development: Analysis and outcomes of a transformative learning process. *European Journal of Training and Development* 48: 3455–77.

Clavero S and Galligan Y (2021) Delivering gender justice in academia through gender equality plans? Normative and practical challenges. *Gender, Work & Organization* 28: 1115–1132.

Cook A and Glass C (2015) Diversity begets diversity? The effects of board composition on the appointment and success of women CEOs. *Social Science Research* 53: 137–47.

Cotter DA, Hermsen JM, Ovadia S, et al. (2001) The glass ceiling effect. *Social Forces* 80(2): 655–81.

Day D and Dragoni L (2015) Leadership development: An outcome-oriented review based on time and levels of analyses. *Annual Review of Organizational Psychology and Organizational Behavior* 2: 133–156.

Debebe G (2011) Creating a safe environment for women's leadership transformation. *Journal of Management Education* 35(5): 679–712.

Debebe G, Anderson D, Bilimoria D, et al. (2016) Women's leadership development programs: Lessons learned and new frontiers. *Journal of Management Education* 40(3): 231–52.

Debebe G and Reinert KA (2014) Leading with our whole selves: A multiple identity approach to leadership development. In: Ferguson AD and Miville ML (eds) *Handbook of Race-Ethnicity and Gender in Psychology*. New York: Springer, 271–93.

Devillard S, Graven W and Lawson E (2012) *Making the Breakthrough*. McKinsey & Co. https://www.mckinsey.com/~/media/mckinsey/dotcom/client_service/organization/pdfs/women_matter_mar2012_english.ashx

Diehl AB and Dzubinski LM (2016) Making the invisible visible: A cross-sector analysis of gender-based leadership barriers. *Human Resource Development Quarterly* 27(2): 181–206.

Eagly AH and Carli LL (2018) Women and the labyrinth of leadership. In: Rosenbach WE (ed.) *Contemporary Issues in Leadership*. New York: Routledge, 147–62.

Ely RJ and Meyerson DE (2000) Theories of gender in organisations: A new approach to organisational analysis and change. *Research in Organizational Behavior* 22: 103–51.

Ely RJ, Ibarra H and Kolb DM (2011) Taking gender into account: Theory and design for women's leadership development programs. *Academy of Management Learning and Education* 10(3): 474–93.

Erhardt N, Werbel J and Shrader C (2003) Board director diversity and firm financial performance. *Corporate Governance: An International Review* 11: 102–11.

Espinosa JF (2013) Towards a gender sensitive evaluation? Practices and challenges in international development evaluation. *Evaluation* 19(2): 171–82.

Fine C, Sojo V and Lawford-Smith H (2020) Why does workplace gender diversity matter? Justice, organisational benefits, and policy. *Social Issues and Policy Review* 14: 36–72.

Fitzsimmons TW, Yates MS and Callan VJ (2020) *Employer of choice for gender equality: Leading practices in strategy, policy and implementation*. Report. Brisbane, QLD, Australia: AIBE Centre for Gender Equality in the Workplace.

Gardiner A, Chur Hansen A, Turnbull D, et al. (2023) Qualitative evaluations of women's leadership programs: A global, multi-sector systematic review. *Australian Journal of Psychology* 75: 1–20.

Garman AN, Standish MP, Carter C, et al. (2021) NCHL's 'best organisations for leadership development' program: A case study in improving evidence-based practice through benchmarking and recognition. *Advances in Health Care Management* 20: 221–30.

Global Partnership for Sustainable Data (2022) Inclusive Data Charter. https://www.data4sdgs.org/initiatives/inclusive-data-charter

Goyal R, Pittman A and Workman A (2010) *Measuring change: Monitoring and evaluating leadership programs*. A Guide for Organisations, Report. Bethesda, MD: Women's Learning Partnership.

Hariton E and Locascio JJ (2018) Randomised controlled trials – The gold standard for effectiveness research. *BJOG: An International Journal of Obstetrics and Gynaecology* 125(13): 1716.

Herbst TH (2020) Gender differences in self-perception accuracy: The confidence gap and women leaders' underrepresentation in academia. *SA Journal of Industrial Psychology* 46(1): 1–8.

Hopkins K, Meyer M, Afkinich J, et al. (2022) Impact of leadership development and facilitated peer coaching on women's individual, collective, and organizational behaviors in human services. *Nonprofit Management and Leadership* 32(3): 387–408.

Hoyt CL, Murphy SE and Hackett JD (2016) Gender diversity in the boardroom and firm financial performance. *Journal of Management* 42(3): 477–510.

Howe-Walsh L and Turnbull S (2014) Barriers to women leaders in academia: Tales from science and technology. *Studies in Higher Education* 41(3): 415–428.

Isaac C, Kaatz A, Lee B, et al. (2012) An educational intervention designed to increase women's leadership self-efficacy. *CBE – Life Sciences Education* 11(3): 307–22.

Johnson SRH, Benjamin C, Miksys C, et al. (2023) A pathway to systemic changes in STEM leadership: Increasing representation of women trough the external mentor program. *Journal of Women and Minorities in Science and Engineering* 29(4): 79–99.

Kalev A, Dobbin F and Kelly E (2006) Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies. *American Sociological Review* 71(4): 589–617.

Kempster S, Higgs M and Wuerz T (2014) Pilots for change: Exploring organisational change through distributed leadership. *Leadership & Organization Development Journal* 35(2): 152–67.

Key S, Popkin S, Munchus G, et al. (2012) An exploration of leadership experiences among white women and women of color. *Journal of Organizational Change Management* 25(3): 392–404.

Koekemoer E, Olckers C and Schaap P (2023) The subjective career success of women: The role of personal resources. *Frontiers in Psychology* 14: 1121989.

Lagarde M (2012) How to do (or not to do) . . . Assessing the impact of a policy change with routine longitudinal data. *Health Policy and Planning* 27(1): 76–83.

London L, Madner S and Skerritt D (2021) How many people are really needed in a transformation? Available at: https://www.mckinsey.com/capabilities/transformation/our-insights/how-many-people-are-really-needed-in-a-transformation (accessed 14 October 2022).

McDonald S (2011) What's in the 'old boys' network? Accessing social capital in gendered and racialized networks. *Social Networks* 33(4): 317–30.

Martin R, Hughes D, Epitropaki O, et al. (2021) In pursuit of causality in leadership training research: A review and pragmatic recommendations. *The Leadership Quarterly* 32(5): 101375.

Mayne J and Johnson N (2015) Using theories of change in the agriculture for nutrition and health CGIAR research program. *Evaluation* 21(4): 407–28.

Meyerson D (1998) Feeling stressed and burned out: A feminist reading and re-visioning of stress-based emotions within medicine and organization science. *Organization Science* 9(1): 103–118. https://doi.org/10.1287/orsc.9.1.103

Miles I and Cunningham P (2005) Smart innovation: A practical guide to evaluating innovation programmes. Brussels and Luxembourg: European Commission.

Molas-Gallart J and Tang P (2011) Tracing 'productive interactions' to identify social impacts: An example from the social sciences. *Research Evaluation* 20(3): 219–26.

Mongon D and Chapman C (2012) *High-Leverage Leadership: Improving Outcomes in Educational Settings*. New York: Routledge.

Moser C (2005) Has gender mainstreaming failed? *International Feminist Journal of Politics* 7(4): 576–90.

Mousa M, Boyle J, Skouteris H, et al. (2021) Advancing women in healthcare leadership: A systemic review and meta-synthesis of multi-sector evidence on organisational interventions. *EClinicalMedicine* 39: 101084.

Nash M and Moore R (2019) 'I was completely oblivious to gender': An exploration of how women in STEMM navigate leadership in a neoliberal, post-feminist context. *Journal of Gender Studies* 28(4): 449–61.

Njah J, Hansoti B, Adeyami A, et al. (2021) Measuring for success: Evaluating leadership training programs for sustainable impact. *Annals of Global Health* 87(1): 63.

Palmén R and Schmidt E (2019) Analysing facilitating and hindering factors for implementing gender equality interventions in R&I: Structures and processes. *Evaluation and Program Planning* 77: 101726.

Pitkin HF (1972) *The Concept of Representation*, 1st edn. Berkeley, CA: University of California Press.

Plantenga D (2004) Gender, identity, and diversity: Learning from insights gained in transformative gender training. *Gender and Development* 12(1): 40–6.

Podsakoff PM and Podsakoff NP (2019) Experimental designs in management and leadership research: Strengths, limitations, and recommendations for improving publishability. *The Leadership Quarterly* 30(1): 11–33.

Reale E, Nedeva M, Thomas D, et al. (2014) Evaluation through impact: A different viewpoint. *Fteval Journal* 39: 36–41.

Rog DJ (2012) When background becomes foreground: Toward context-sensitive evaluation practice. *New Directions for Evaluation* 135: 25–40.

Rubin J (1975) What the 'good language learner' can teach us. *TESOL Quarterly* 9: 41–51.

Ryan MK (2023) Addressing workplace gender inequality: Using the evidence to avoid common pitfalls. *British Journal of Social Psychology* 62(1): 1–11.

Ryan MK and Haslam SA (2007) The glass cliff: Exploring the dynamics surrounding the appointment of women to precarious leadership positions. *Academy of Management Review* 32(2): 549–72.

Ryan MK and Morgenroth T (2024) Why we should stop trying to fix women: How context shapes and constrains women's career trajectories. *Annual Review of Psychology* 75: 555–72.

Ryan T (2020) The intersectional challenges of indigenous women's leadership. *AB-Original* 3: 149–71.

Schein VE, Mueller R, Lituchy T, et al. (1996) Think manager—think male: A global phenomenon? *Journal of Organizational Behavior* 17(1): 33–41.

Schmidt EK and Cacace M (2017) Addressing gender inequality in science: The multifaceted challenge of assessing impact. *Research Evaluation* 26(2): 1–13.

Schmidt EK and Graversen EK (2020) Developing a conceptual evaluation framework for gender equality interventions in research and innovation. *Evaluation and Program Planning* 79: 101750.

Schmidt EK, Bührer S, Schraudner M, et al. (2018) A conceptual evaluation framework for promoting gender equality in research and innovation. *European Project, European Union*. Available at: https://efforti.eu/files/sites/default/files/2018-03/efforti%20d3.3%20final%20report%2027032018.pdf

Schmidt V (2011) Speaking of change: Why discourse is key to the dynamics of policy transformation. *Critical Policy Studies* 5(2): 106–26.

Skouteris H, Ananda-Rajah M, Blewitt C, et al. (2023) 'No one can actually see us in positions of power': The intersectionality between gender and culture for women in leadership. *BMJ Leader* 8: 794.

Sielbeck-Bowen K, Brisolara S, Seigart D, et al. (2002) Exploring feminist evaluation: The ground from which we rise. *New Directions for Evaluation* 2002: 3–8.

Sinclair A (1997) The MBA through women's eyes: Learning and pedagogy in management education. *Management Learning* 28(3): 313–30.

Sojo V, Ryan M, Fine C, et al. (2022) *What works, what's fair? Using systematic reviews to build the evidence base on strategies to increase gender equality in the public sector*. Research Report for the Victorian Commission for Gender Equality in the Public Sector. Melbourne, VIC, Australia: The University of Melbourne, The ANU, Swinburne University.

Starmarski CS and Son Hing LS (2015) Gender inequalities in the workplace: The effects of organisational structure, processes, practices and decision makers' sexism. *Frontiers in Psychology* 6: 1400.

Stone D (2016) Quantitative analysis as narrative. In: Bevir M and Rhodes RA (eds) *Routledge Handbook of Interpretative Political Science*. New York: Routledge, 169–182.

Streeton A-M, Kitsell F, Oo M, et al. (2021) The improving global health programme – Leadership development in the NHS through overseas placement. *BMJ Global Health* 6: e004533.

Terjesen S, Sealy R and Singh V (2009) Women directors on corporate boards: A review and research agenda. *Journal of Business Ethics* 88(1): 1–18.

Thomas H and Turnbull P (2018) From horizontal to vertical labour governance: The International Labour Organization (ILO) and decent work in global supply chains. *Human Relations* 71(4): 536–59.

Timmers TM, Willemsen TM and Tijdens KG (2010) Gender diversity policies in universities: a multiperspective framework of policy measures. *Higher Education* 59(6): 719–735.

Vogel I (2012) *Review of the use of 'theory of change' in international development*. Report. London: Department for International Development.

Williamson A, Smith K, Sojo V, et al. (2024) *Equitable Foundations: A Framework for Gender-Wise Philanthropy in Partnerships*. Melbourne, VIC, Australia: Australians Investing In Women and Melbourne Social Equity Institute, The University of Melbourne.

Women of Colour (2021) *Workplace Survey Report 2021, Report: Women of Colour in partnership with Dr. Catherine Archer*. Murdoch, WA, Australia: Murdoch University.

Wong CYE, Kirby TA, Rink F, et al. (2022) Intersectional invisibility in women's diversity interventions. *Frontiers in Psychology* 13: 791572.

World Economic Forum (2023) Global gender gap report 2023. Available at: https://www.weforum.org/reports/global-gender-gap-report-2023/ (accessed 12 October 2023).

Wroblewski A and Palmén R (2022) A reflexive approach to structural change. In: Wroblewski A and Palmén R (eds) *Overcoming the Challenge of Structural Change in Research Organisations – A Reflexive Approach to Gender Equality*. Leeds: Emerald Publishing Limited, 15–32.

Yang T and del Carmen Triana M (2019) Set up to fail: Explaining when women-led businesses are more likely to fail. *Journal of Management* 45(3): 926–54.

Elise Stephenson is the Deputy Director of the Global Institute for Women's Leadership, Australian National University, founded and chaired by former prime minister Julia Gillard. Elise's research focuses on gender, sexuality and leadership in policy frontiers, from researching the space sector, to diplomacy, national security, intelligence, climate action, international relations, and the Asia Pacific.

Gosia Mikolajczak is a Research Fellow at the Global Institute for Women's Leadership. Her research focuses on gender and gender-related policies, well-being of vulnerable groups, ideology, and social change.

Michelle Ryan is a World-Renowned Gender Equality Expert, Professor of Social and Organisational Psychology, and the inaugural Director of the Globval Institute for Women's Leadership. Her work centres on understanding the psychological processes underlying workplace gender inequality, and designing and implementing innovative and evidence-based interventions to increase gender equality.

Alexandra N. Fisher is a Research Fellow and Social Psychologist at the ANU Global Institute for Women's Leadership. Dr Fisher's research aims to challenge dominant norms and social scripts that prevent progress toward gender equality, both in the context of work and close relationships.

Jack Hayes is a PhD Candidate in the Department of International Relations in the Coral Bell School of Asia Pacific Affairs at the Australian National University. His research expertise includes LGBTI+ representation in Australian politics, democratic and electoral health, and good governance in public service.

Victor Sojo is an Associate Professor in Leadership in the Department of Management and Marketing, FBE. He works in several interdisciplinary leadership, diversity management and equity research projects with government/industry partners.

Morgan Weaving is a Postdoctoral Researcher, working with Dr. Michele Gelfand on research relating to culture, norms, diversity, and stigma at the Stanford Graduate School of Business. Morgan obtained her PhD in social psychology at the University of Melbourne, where she explored the behaviours and psychological processes that reinforce hierarchical gender relations.

Mai Tanjitpiyanond expertise is in the Psychology of Leadership and Group Dynamics. She completed her PhD in social psychology at the University of Queensland (UQ) where she won multiple awards for her research on economic inequality and stereotyping.