



MONASH University

Evaluation in health promotion: Gathering evidence to improve effectiveness

Joanna Schwarzman

Bachelor Physiotherapy

Master Public Health

A thesis submitted for the degree of Doctor of Philosophy at
Monash University in 2019
School of Public Health and Preventive Medicine
Melbourne, Australia

Copyright notice

I certify that I have made all reasonable efforts to secure copyright permissions for third-party content included in this thesis and have not knowingly added copyright content to my work without the owner's permission.

© The author (2019).

Abstract

There is an urgent need to grow the evidence base for use in public health decision making. Evaluation of health promotion and primary prevention programs are a valuable source of this evidence, however, organisations face challenges to conducting rigorous evaluation, and reported evaluation quality is low. Studies from the broader evaluation field have identified components of organisational capacity that affect evaluation practice and use, but there has been limited investigation of influences upon evaluation in health promotion, and models of evaluation capacity have not been validated in this field. In order to plan, implement and evaluate health promotion evaluation capacity building efforts, there is a need to better understand the determinants of evaluation capacity, and their interactions.

The primary aim of this thesis was to determine the strengths and limitations of evaluation practice in the Australian health promotion field, and to identify the factors that need to be addressed to improve evaluation quality. A sequential mixed-methods program of research was undertaken. In this program of research, health promotion was defined as “the process of enabling people to increase control over their health and its determinants, and thereby improve their health.”[1 p1] Government and non-government health promotion agencies (n=116) which seek to promote health and prevent disease and injury, were recruited from four states of Australia to participate in the study.

The first qualitative study used semi-structured interviews to identify the important facilitators and barriers to evaluation practice and use from the perspective of experienced practitioners (n=40). Factors within the organisation, including leadership and culture, and at the prevention system-level, through funding, policy and administrative mechanisms, were identified as pivotal influences upon evaluation.

The second study involved the development and validation of an Evaluation Practice Analysis Survey (EPAS) to examine evaluation practices in health promotion organisations. The EPAS was developed in multiple stages, based on the qualitative study findings. Health

promotion practitioners (n=169) external to the 116 participating organisations were recruited. The psychometric properties of the EPAS were assessed using Principal components analysis and Cronbach's alpha. The final survey included 25 scales measuring aspects of evaluation capacity. Scales representing dedicated resources, leadership, organisational culture and support for evaluation showed promising predictive validity.

The third study assessed the interactions between the aspects of evaluation capacity using EPAS responses from health promotion practitioners (n=219) within participating organisations. This study confirmed the validity of seven important evaluation capacity constructs, using confirmatory factor analysis. Structural equation modelling was then used to assess the validity of a conceptual framework of health promotion evaluation capacity. Good fit was found for a path model that identified the relationships between practitioner, organisational, system-level aspects of health promotion evaluation capacity, and reported impact evaluation.

The final study examined the methods and quality of evaluation conducted in participating organisations, and identified factors associated with evaluation quality. A systematic audit and appraisal of evaluation reports (n=392) was conducted. Overall evaluation quality was found to be low (median score 24.5%). Evaluation quality scores were linked by organisation to EPAS responses from the previous study. Multivariable regression analysis revealed that health promotion budget and conducting state-wide or national prevention programs were independently associated with higher quality evaluation.

This research shows that there is opportunity for gains to evaluation capacity by addressing the pivotal determinants of evaluation. Further, the findings from this study indicate there is great potential to improve evaluation quality in organisations which deliver health promotion on a small scale. Ultimately, this research in this thesis contributes to the evidence base for health promotion evaluation capacity, and provides guidance and tools to those seeking to improve evaluation capacity in the field.

Research outputs related to candidature

Peer reviewed publications

Schwarzman J, Bauman AE, Gabbe BJ, Rissel C, Shilton T, Smith BJ. Understanding the factors that influence health promotion evaluation: the development and validation of the Evaluation Practice Analysis Survey. *Evaluation and Program Planning*. 2019;74:76-83. -- see Chapter Six

Schwarzman J, Bauman A, Gabbe BJ, Rissel C, Shilton T, Smith BJ. The Funding, Administrative, and Policy Influences on the Evaluation of Primary Prevention Programs in Australia. *Prevention Science*. 2019: <https://doi:10.1007/s11121-019-00997-4> -- see Chapter Five

Schwarzman J, Bauman A, Gabbe B, Rissel C, Shilton T, Smith BJ. Organizational determinants of evaluation practice in Australian prevention agencies. *Health Education Research*. 2018;33(3):243–55. -- see Chapter Five

Submitted for publication

Schwarzman J, Bauman A, Gabbe B, Rissel C, Shilton T, Smith BJ. How practitioner, organisational and system-level factors act to influence health promotion evaluation practice: validation of a conceptual framework. *Health Education Research* (under review). -- see Chapter Seven

Schwarzman J, Nau T, Bauman A, Gabbe B, Rissel C, Shilton T, et al. An assessment of program evaluation methods and quality in Australian prevention agencies. *Health Promotion Journal of Australia* (under review). -- see chapter Eight

Conference presentations

J Schwarzman, BJ Smith, A Bauman, B Gabbe, C Rissel, T Shilton, 'The potential for system level change: Addressing political and funding level factors to facilitate health promotion and disease prevention evaluation', oral presentation, Australasian Evaluation Society International Evaluation Conference, September 2018, Launceston, Australia.

J Schwarzman, BJ Smith, A Bauman, B Gabbe, C Rissel, T Shilton, 'Improving evaluation effectiveness: the potential for gains in the Australian prevention system', oral presentation, Public Health Prevention Conference, May 2018, Sydney, Australia

J Schwarzman, BJ Smith, A Bauman, C Rissel, T Shilton, 'Evaluation practice and use in Australian prevention agencies', oral presentation, World Congress on Public Health Conference, April 2017, Melbourne, Australia.

J Schwarzman, BJ Smith, 'Evaluation in health promotion: building capacity for evidence gathering and learning', workshop, 23rd National Conference of Australian Health Promotion Association, June 2016, Perth, Australia.

Other research outputs during candidature

Wright CJC, **Schwarzman J**, Dietze PM, Crockett B, Lim MSC. Barriers and opportunities in the translation of mobile phone and social media interventions between research and health promotion practice in Australia: a qualitative study of expert perspectives. *Health Research Policy and Systems*. 2019;17(1):5.

de Lacy-Vawdon CJ, **Schwarzman J**, Nolan G, de Silva R, Menzies D, Smith BJ. Facilitators of Attendance and Adherence to Group-Based Physical Activity for Older Adults: A Literature Synthesis. *Journal of Aging and Physical Activity*. 2017:1-39.

Thesis including published works declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes three original papers published in peer reviewed journals and two submitted publications. The core theme of the thesis is evaluation capacity and practice in health promotion. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the student, working within the School of Public Health and Preventive Medicine under the supervision of Associate Professor Ben Smith, Professor Belinda Gabbe and Professor Adrian Bauman.

The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.

In the case of Chapters Five to Eight my contribution to the work involved the following:

Thesis Chapter	Publication Title	Status	Nature and % of student contribution	Co-author names Nature and % of Co-author's contribution	Monash student
Five	Organizational determinants of evaluation practice in Australian prevention agencies	Published	65%. Study design, data collection, data analysis, interpretation, manuscript preparation and revision.	Ben Smith, concept, study design, assistance with data analysis, input into manuscript 15%	No
				Belinda Gabbe, assistance with data analysis, input into manuscript 7.5%	No
				Adrian Bauman, concept, study design, input into manuscript 7.5%	No
				Chris Rissel, concept, input into manuscript 2.5%	No
				Trevor Shilton, concept, input into manuscript 2.5%	No

Five	The Funding, Administrative, and Policy Influences on the Evaluation of Primary Prevention Programs in Australia	Published	65%. Study design, data collection, data analysis, interpretation, manuscript preparation and revision.	Ben Smith, concept, study design, assistance with data analysis, input into manuscript 15%	No
				Belinda Gabbe, assistance with data analysis, input into manuscript 7.5%	No
				Adrian Bauman, concept, study design, input into manuscript 7.5%	No
				Chris Rissel, concept, input into manuscript 2.5%	No
				Trevor Shilton, concept, input into manuscript 2.5%	No
Six	Understanding the factors that influence health promotion evaluation: the development and validation of the Evaluation Practice Analysis Survey	Published	65%. Study design, data collection, data analysis, interpretation, manuscript preparation and revision.	Ben Smith, concept, study design, assistance with data analysis, input into manuscript 15%	No
				Belinda Gabbe, assistance with data analysis, input into manuscript 7.5%	No
				Adrian Bauman, concept, study design, assistance with data analysis, input into manuscript 7.5%	No
				Chris Rissel, concept, input into manuscript 2.5%	No
				Trevor Shilton, concept, input into manuscript 2.5%	No
Seven	How practitioner, organisational and system-level factors act to influence health promotion evaluation practice: validation of a conceptual framework	Submitted	65%. Study design, data collection, data analysis, interpretation, manuscript preparation.	Ben Smith, concept, study design, assistance with data analysis, input into manuscript 15%	No
				Belinda Gabbe, assistance with data analysis, input into manuscript 7.5%	No
				Adrian Bauman, concept, study design, assistance with data analysis, input into manuscript 7.5%	No
				Chris Rissel, concept, input into manuscript 2.5%	No
				Trevor Shilton, concept, input into manuscript 2.5%	No
Eight	An assessment of program evaluation methods and quality in Australian prevention agencies	Submitted	60%. Study design, data collection, data analysis, interpretation, manuscript preparation.	Ben Smith, concept, study design, assistance with data analysis, input into manuscript 15%	No
				Tracy Nau, data collection, input into manuscript 7.5%	No
				Belinda Gabbe, assistance with data analysis, input into manuscript 7.5%	No
				Adrian Bauman, concept, study design, input into manuscript 5%	No
				Chris Rissel, concept, input into manuscript 2.5%	No
Trevor Shilton, concept, input into manuscript 2.5%	No				

I have not renumbered sections of published papers within in this thesis. I have renumbered sections of submitted papers in order to generate a consistent presentation within the thesis.

Student signature: [Joanna Schwarzman] Date: 2nd April, 2019

The undersigned hereby certify that the above declaration correctly reflects the nature and extent of the student's and co-authors' contributions to this work. In instances where I am not the responsible author I have consulted with the responsible author to agree on the respective contributions of the authors.

Main Supervisor signature: [Ben Smith] Date: 2nd April, 2019

Acknowledgements

I would firstly like to acknowledge the incredible efforts of my supervisors Ben Smith, Belinda Gabbe and Adrian Bauman. In the earliest supervision meetings I realised how lucky I was to have landed in a team of wonderfully encouraging and generous supervisors. You each possess an uncanny ability to give a motivational speech on request, or at times when I didn't know I needed one. Your commitment, availability, patience and timely support have made all the difference as I have learned to become a researcher.

This research project was funded by an Australian Research Council Discovery Grant. Thank you to the investigator team responsible for the project, including lead investigators Ben Smith and Adrian Bauman, and associate investigators Chris Rissel and Trevor Shilton for their input to critical points of the project.

I'm also grateful to the fabulous research assistants Emily McCluskey, Elizabeth Vella, Keren Shanley and Tracy Nau who worked on various aspects of participant recruitment and data collection. It was a pleasure working with you all, and your persistence with participant recruitment and follow up certainly contributed to the success of this project. Also, thank you to the many others who have provided advice on a range of research topics throughout this project.

To my parents Judy and Peter, thank you for your ongoing practical and emotional support. This PhD would not have been possible without your commitment to my education and encouragement to try new things from an early age.

To my friends who are in the middle of, or have finished, their PhDs. Thank you for the many shared lunches, walks and chats, I'm so glad to have met you all as part of this process. A special mention goes to Anna Nicholson, Christina Heris and Cass Wright - your friendship, encouragement, understanding and generosity of time and advice has been invaluable.

Thanks also to colleagues in CCRE who kept an amazingly stocked lolly jar, introduced me to

Tuesday yoga, and maintained an active social calendar. It was such a great work environment to be part of.

To my siblings and friends, thank you for tolerating my stress and unavailability with kindness. I'm looking forward to spending more time with you all soon.

Finally, an enormous thankyou to Dave Fabri, who has supported me through the stressful moments, and helped to get me through to the end. You have been patient and loving and I can't wait to enjoy post-PhD life with you too.

Table of Contents

Copyright notice.....	2
Abstract.....	3
Research outputs during candidature.....	5
Thesis including published works declaration.....	7
Acknowledgements (placeholder).....	10
Table of Contents	xii
List of Tables	xiv
List of Figures.....	xv
Acronyms and abbreviations	xvi
1 Background	1
1.1 The need for an evidence base of health promotion strategies to prevent disease and injury	1
1.2 The Australian health promotion environment.....	3
1.3 What is health promotion evaluation?.....	3
1.4 What do policy makers, program managers and practitioners need evaluation for?.....	6
1.5 The characteristics and quality of reported evaluation in health promotion	7
1.6 Building capacity to generate quality evaluation. Where to from here?.....	8
1.7 Summary.....	9
2 The potential to improve health promotion evaluation capacity	10
2.1 Barriers and facilitators to health promotion evaluation	11
2.2 Lessons from outside the health promotion field	14
2.3 Measuring elements of evaluation capacity	15
2.4 The reliability and validity of existing evaluation capacity measures	16
2.5 Limitations of evaluation capacity and evaluation capacity building research	17
2.6 Summary.....	18
3 Thesis rationale and aims	20
3.1 Thesis rationale	20
3.2 Thesis aims and objectives.....	21
3.3 Thesis overview	22
3.4 Summary.....	23
4 Study design, participants and methods overview	24
4.1 Study methodology and design	24
4.2 Ethics.....	25
4.3 Organisational identification and screening.....	25
4.4 Organisational recruitment.....	28
4.5 Participating organisations	28

4.6	Data collection and analysis overview	30
4.7	Summary	33
5	The factors that influence health promotion evaluation: findings from a qualitative study	34
6	The development and validation of the Evaluation Practice Analysis Survey.....	60
7	How do the determinants of health promotion evaluation interact? Validation of a conceptual framework	69
7.1	Abstract	71
7.2	Introduction	71
7.3	Methods.....	76
7.4	Results.....	79
7.5	Discussion	85
7.6	Acknowledgements	88
8	An audit and quality appraisal of health promotion evaluation.....	90
8.1	Abstract	92
8.2	Introduction	92
8.3	Methods.....	94
8.4	Results.....	98
8.5	Discussion	106
8.6	Conclusions	109
8.7	Acknowledgements	110
9	Integrated discussion.....	111
9.1	Key findings and implications.....	112
9.2	Strengths and limitations of this research	118
9.3	Future directions for evaluation capacity building.....	119
9.4	Future directions for research.....	121
9.5	Recommendations for evaluation capacity building.....	122
9.6	Conclusion.....	124
	References.....	125
	Appendices.....	138
	Appendix A. Chapter Five Supplementary Material	139
	Appendix B. Chapter Six Supplementary Materials	141
	Appendix C. Chapter Seven Supplementary Material	146
	Appendix D. Chapter Eight Supplementary Material.....	155

List of Tables

Table 1. Levels of evaluation (based on definitions from Bauman and Nutbeam{Bauman, 2014 #1329}).....	5
Table 2. Summary of eligible and participating organisations.....	28
Table 3. Survey respondent characteristics	81
Table 4. Confirmatory factor analysis of practitioner, organisational and system-level latent evaluation capacity variables.....	83
Table 5. Correlations between latent evaluation capacity variables and reported impact evaluation	84
Table 6. Characteristics of participating organisations	101
Table 7. Evaluation characteristics by key organisational attributes	102
Table 8. Summary of data collection methods used at each level of evaluation	103
Table 9. Levels of evaluation by key organisational attributes.....	104
Table 10. Generalised Estimating Equation (GEE) analysis of factors associated with evaluation quality score [†]	106

List of Figures

Figure 1. Stages of participant recruitment	29
Figure 2. Phases of data collection	30
Figure 3. Path model relationship between components of evaluation capacity and reported impact evaluation	85
Figure 4. Organisational response to evaluation report collection	99

Acronyms and abbreviations

AHPA	Australian Health Promotion Association
ARC	Australian Research Council
AUD	Australian Dollars
CEO	Chief Executive Officer
CFA	Confirmatory factor analysis
CFI	Comparative fit index
CI	Confidence interval
df	Degrees of freedom
ECAI	Evaluation Capacity Assessment Index
ECB	Evaluation capacity building
ECI	Evaluation Capacity Instrument
ECOQ	Evaluation Capacity in Organisations Questionnaire
EPAS	Evaluation Practice Analysis Survey
FTE	Full-time equivalent
GEE	Generalised estimating equation
HP	Health promotion
HREC	Human Research Ethics Committee
IQ range	Inter-quartile range

KMO	Kaiser-Meyer-Olkin Measure of Sampling Adequacy
LGA	Local government area
MLE	Maximum likelihood estimation
NGO	Non-government organisation
OR	Odds ratio
PCA	Principal components analysis
R&E	Research and evaluation
RE-AIM	Reach, Efficacy, Adoption, Implementation, Maintenance
Ref	Reference category
RMSEA	Root Mean Square Error of Approximation
SD	Standard deviation
SEF	Standard Evaluation Framework
SEM	Structural equation modelling
SiREN	Western Australian Sexual Health and Blood-borne Virus Applied Research and Evaluation Network
SRMR	Standardized Root Mean Square Residual
Ter	Tertile
UK	United Kingdom
USA	United States of America

Chapter One

1 Background

Chapter One outlines the need to improve the evidence base for health promotion and primary prevention strategies, and provides a brief overview of the Australian health promotion context, which is the setting for the research described in this thesis. Further, an overview of how evaluation is typically undertaken within health promotion, including an examination of the characteristics and quality of evaluation conducted, is provided. The need for a deeper understanding of evaluation practice, in order to improve the capacity for the health promotion field to contribute quality evidence to the evidence base, is also discussed.

1.1 The need for an evidence base of health promotion strategies to prevent disease and injury

In this thesis, health promotion is defined as:

“the process of enabling people to increase control over their health and its determinants, and thereby improve their health. It is a core function of public health and contributes to the work of tackling communicable and non-communicable diseases and other threats to health”;[1 p1]

and primary prevention:

“involves actions that help participants (or to facilitate participants helping themselves) 1) to prevent predictable and interrelated problems, 2) to protect existing states of health and healthy functioning, and 3) to promote psychosocial wellness for identified populations of people”.[2 p9]

Health promotion and primary prevention (from here ‘health promotion’) strives to prevent the onset of disease and injury, and promote the health and well-being of individuals,

groups, communities and the population. In 2016, chronic diseases accounted for 89% of deaths in Australia and 71% of deaths globally.[3] In Australia, up to a third of the burden of disease could be prevented by addressing risk factors such as physical inactivity, diet, alcohol, tobacco and other drug use.[4,5] Chronic disease and ill health disproportionately affect Aboriginal and Torres Strait Islander populations[6] and those from disadvantaged communities.[7] The effect of the social determinants of health, that is the ‘circumstances in which people grow, live, work, and age’,[8 piii] is well documented, where housing, education and income all play a role in health outcomes.[9] However, the complex interaction between the social determinants, risk factors and health outcomes can make it challenging to identify and implement effective strategies.[10]

Health promotion strategies operate within a socio-ecological approach to health, and seek to effect change across legislation and policy, social and physical environments, education or workplace settings, and with small groups and individuals.[5,11] Health promotion strategies intentionally engage diverse and disadvantaged populations, making these strategies essential to addressing achieving health equity.[12] The need to address the determinants of health at multiple levels can result in complicated programs.[11] Initiatives must be adapted to suit the local context,[13] and have strong levels of engagement from the community.[14]

The use of current best evidence is critical in improving the health and wellbeing of the population.[15] Despite the magnitude of current health challenges, the evidence-base to guide the selection and implementation of effective health promotion strategies is limited.[10,16,17] Evidence of program effectiveness (or ineffectiveness), details of strategy implementation, population reach and context are necessary to inform policy, guide resource allocation and support evidence-based decision making in public health.[15,18] There is also a need to generate evidence that addresses, rather than compounds, health inequalities.[19] Evidence generated by traditional research methods has been skewed towards easy to reach populations, and simple interventions that tend to focus on individual behaviours.[10,18,20] Rigorous evaluation of multi-level health promotion programs has the potential to

contribute to the evidence-base that informs policy and practice.[16]

1.2 The Australian health promotion environment

In Australia, health promotion activities are predominantly conducted by government (local, state and federal) and non-government organisations (NGO) such as community health services, Aboriginal community controlled health organisations and disease specific agencies. Prevention activities in some jurisdictions are supported and coordinated through central bodies which have responsibility for commissioning, regional coordination and partnership development. Strategies are frequently implemented in partnership, and health promotion organisations are often the lead agency responsible for the initiation, planning, coordination, delivery and evaluation of these initiatives. Over past decades, the priorities for prevention spending have varied at the national[21] and state levels.[22] A significant national investment in prevention was unexpectedly withdrawn in 2014, leaving organisations facing uncertainty and disinvestment in prevention initiatives.[21]

The health promotion workforce comes from a multitude of professional backgrounds.[23] Those with health promotion and public health qualifications are often taught how to plan, implement and evaluate health promotion initiatives that address the determinants of health at different levels. Others from clinical health, community development, Aboriginal health, education, and sports backgrounds also make a valuable contribution to the health promotion discipline,[24] yet may lack training in program planning and evaluation. Recently, efforts to professionalise the health promotion workforce through a voluntary registration scheme were launched in Australia.[25]

1.3 What is health promotion evaluation?

This thesis focuses on program evaluation in the health promotion field, where programs are defined as:

“any set of organized activities supported by a set of resources to achieve a specific and intended result”;[26 p3]

and evaluation is defined as:

“the systemic collection of information about the activities, characteristics, and results of programs to make judgements about the program, improve or further develop program effectiveness, inform decisions about future programming, and/or increase understanding.”[27 P39]

The health promotion field continues to recognise the contribution of evaluation to planning, implementing, and improving programs, to improve community and population health.[28]

Research, including intervention studies, are conducted with the primary purpose of hypothesis testing and ultimately seek to generate new knowledge. In contrast, the primary purpose of program evaluation is to inform improvements in program effectiveness.[26,29,30] Both evaluation and research in public health and health promotion involve systematic investigation, and can involve similar designs and data collection methods.[29] This thesis focuses on the potential contribution of evidence generated through program evaluation to health promotion policy and practice. While research quality and usefulness for decision making often comes under scrutiny, there is relatively little attention paid to the conduct and quality of the large body of program evaluation conducted within organisations, which deliver prevention strategies in ‘real world’ conditions.

Program evaluation is a systematic process which involves multiple steps.[26] Guidelines and frameworks are available to guide the evaluation of prevention programs delivered in community and population settings. Evaluation approaches that have long been used in the health promotion field include Bauman and Nutbeam’s[31] four levels of evaluation (Table 1), the RE-AIM (Reach, Effectiveness, Adoption, Implementation, Maintenance) framework,[32] and participatory and empowerment approaches.[33,34] Application of economic evaluation to health promotion programs is also important.[18,35-38] More recently, some health promotion practitioners have engaged in developmental evaluation,[30,39,40] and adopted tools associated with systems approaches.[41,42] The

framework of formative, process, impact and outcome evaluation, as described by Bauman and Nutbeam[31], is used throughout this thesis. This framework relates the four levels of evaluation to informational needs within the program cycle[31] (Table 1). This evaluation framework has been used to improve the consistency and quality of evaluation,[43] and forms the basis of health promotion program reporting requirements in many jurisdictions.[44]

Table 1. Levels of evaluation (based on definitions from Bauman and Nutbeam[31])

Level of evaluation	Purpose and description	In practice
Formative	Develop, identify and test program materials and methods.	Formative evaluation occurs before any elements of the program are implemented, and can be used to improve the likelihood of program success. It also includes evaluation of pilot programs.
Process	Assess the success and factors surrounding program implementation.	Process evaluation describes and explains aspects of program implementation. Specifically, this includes the extent of program delivery as planned (delivery or fidelity), participants characteristics and involvement (reach, exposure, satisfaction), and context of program delivery (context, barriers and facilitators).
Impact	Assess short-term effects of an intervention.	Impact evaluation typically involves the measurement of intermediate health outcomes, such as changes in behaviour, service access, knowledge, skills or attitudes.
Outcome	Assess whether or not the program successfully achieved its goals; usually implies longer term changes.	Outcome evaluation involves the measurement of changes to health status or behaviour.

Health promotion practitioners are often responsible for planning and carrying out program evaluation;[17] evaluation conduct and use is a documented core competency of health promotion practitioners,[45] and a vast body of additional guidance material is available to assist practitioners at each stage of the evaluation process.[14,26,46-49] Having health promotion practitioners responsible for the evaluation of their own programs has been questioned, given the wide range of other duties they perform in relation to program delivery.[50,51]. The diversity of professional backgrounds and knowledge among health

promotion practitioners can also create challenges in establishing a shared language and understanding about consistent approaches to evaluating health promotion programs.

1.4 What do policy makers, program managers and practitioners need evaluation for?

Evaluation of programs conducted in 'real world' settings has the potential to make significant contributions to the health promotion evidence base.[16,18] Policy and decision makers require evidence of effectiveness, successful implementation, feasibility and cost-effectiveness.[14,52,53] Program transferability and scalability requires relevant and timely evidence with sufficient contextual information and detail of the implementation process.[13,16,18,52] Further, evidence from evaluation of program reach is needed in order to understand which sections of the population are impacted by health promotion strategies, and to ensure programs do not contribute to widening health inequalities.[18,54]

In health promotion organisations, practitioners and managers use and value multiple sources of evidence for practical or strategic purposes.[14,55] Practitioners and managers need evidence to inform program planning, stay up-to-date with current practice,[13,51] justify funding, or influence managerial decisions.[55] To inform program planning and design, practitioners have been found to place greater value on local program evaluations that are highly contextualised[55], particularly in rural or remote settings or in Indigenous communities.[13] Within organisations, evaluation evidence is needed to improve strategies and build relationships with the community stakeholders.[14,31,56]

Health promotion evaluation is often essential to demonstrate compliance with funding body requirements and accountability for funding and to the community.[14,44,57-59] However, there is a perception that findings of evaluation conducted for compliance purposes are not often used by the funding body.[44,56,60]

1.5 The characteristics and quality of reported evaluation in health promotion

Despite the well understood need for program evaluation to inform the evidence base, and the widely available guidance materials, there remain gaps in the availability and quality of evidence to meet the needs of the health promotion field.[10,16,17] As outlined in section 1.4, evidence of program efficacy, effectiveness and program implementation is important to policy makers and program managers, yet this information is often reported as lacking.[13,52,54,55]

A number of studies have examined the extent to which evaluations contain sufficient and valid evidence to inform decision making.[43,44,60-63] These studies have examined evaluation published in the peer-reviewed literature or collected reports directly from health promotion organisations. Three of the studies reviewed the peer-reviewed literature, focusing on mass media campaigns targeting physical activity in adults,[61] health promotion initiatives delivered via social networking sites[62] and Australian health promotion programs.[63] The remaining studies sourced evaluation reports directly from Australian community health organisations,[44,60] or United Kingdom (UK) sites funded to deliver a physical activity initiative.[43]

Two studies described the characteristics of the evaluation conducted, and found impact and process evaluation was commonly included in health promotion evaluations.[44,63] However, both studies described limitations of the data collection methods used.[44,63] Chambers[63] also identified a lack of clearly reported evaluation purpose as a limitation to the utility of evaluation findings for policy makers.[63] Francis and Smith[44] found that there was close alignment between the evaluation levels reported and the requirements of the funding body, indicating that evaluation was driven by the need to meet funding body requirements.[44] Further studies reviewed the peer-reviewed published literature and examined the suitability of evaluation designs and methods to evaluate mass media, or social media interventions.[61,62] Both studies concluded that optimal designs to determine

campaign impact or effectiveness were infrequently used.[61,62]

Only two studies were identified that sought to appraise the quality of evaluations collected directly from health promotion organisations, that is evaluation reports predominantly not published in the grey, or peer-reviewed literature.[43,60] The first study aimed to understand the potential of evaluation, conducted by community health services in metropolitan South Australia, to contribute evidence for health promotion in this setting.[60] The research team developed a 16-item tool to appraise the degree of reporting on program planning, logic models and evaluation methodology in 93 evaluations.[60] The study found lowest scores for items relating to measurement of long-term outcomes, and assessment of sustainability and transferability.[60] The second study also examined evaluation documentation collected directly from six sites funded to deliver ‘Walking Cities’ in the UK.[43] This study examined planning and evaluation documentation and compared it against the mandatory Standard Evaluation Framework (SEF)[64] provided to funded organisations.[43] The SEF was based on the framework by Bauman and Nutbeam,[31] and aimed to improve the consistency and quality of obesity related interventions in the UK.[43] The 52-item SEF assessed program details, evaluation methodologies, participant details and characteristics, collection of baseline and follow-up data, process evaluation, analysis and interpretation.[43] Overall, the authors found few of the criteria to be reported or met, despite the SEF being mandatory.[43]

1.6 Building capacity to generate quality evaluation; where to from here?

The studies described in section 1.5 largely support the need for evaluation quality to be defined based on its ability to inform policy and program decision making in health promotion and public health. These review studies and evaluation report appraisals highlight the persistent gaps in the potential of evaluation, across a range of health promotion settings, to sufficiently inform the evidence base.[43,44,60-63] However, the scope of recommendations arising from these studies highlight the lack of agreement about what needs to be addressed to improve evaluation quality in health promotion. The authors offer

suggestions ranging from the provision of explicit guidance about the designs appropriate for specific settings,[61,62] through to calls for increased resourcing for health promotion research and evaluation,[60] increased collaboration between researchers and practitioners,[43,44] and better communication between funding agencies and organisations about reporting requirements.[44] The health promotion field must first understand the influences upon evaluation practice and capacity, in order to begin strengthening the availability and quality of evidence that is needed to inform policy and practice. Gaining a deeper understanding of the factors that affect health promotion program evaluation is a crucial initial step towards building evaluation capacity.

1.7 Summary

In this chapter, the critical need for evidence of health promotion and prevention program implementation and effectiveness was described. Within health promotion, practitioners, program managers and policy makers are aware of the benefits of conducting and using evaluation. However, the literature reviewed also highlights that despite the known importance, and the availability of evaluation frameworks and guidance materials, documented health promotion evidence is often of insufficient quality to inform decision making, or contribute to program improvement. The challenge of improving health promotion evaluation quality hinges on understanding what affects evaluation practice and capacity within the field. The next chapter reviews literature that addresses what is already known to influence health promotion evaluation, and describes efforts to conceptualise and measure evaluation capacity from the evaluation field.

Chapter Two

2 The potential to improve health promotion evaluation capacity

Over the past two decades, a wide range of organisations, including community-based organisations and government agencies, have engaged in strategies that aim to improve the capacity to conduct and use evaluation.[65] Evaluation capacity building (ECB) is a recognised field of research and practice within the evaluation discipline.[66] While a number of working definitions of ECB focus on improving evaluation capacity within an organisation,[67-70] this conceptual definition of ECB by Stockdill[70] allows greater scope to define the challenges of improving evaluation quality across the health promotion field:

“ECB is a context-dependent, intentional action system of guided processes and practices for bringing about and sustaining a state of affairs in which quality program evaluation and its appropriate uses are ordinary and ongoing practices within and/or between one or more organizations/programs/sites.”[70 p8]

In order to identify, plan and implement successful ECB activities within health promotion organisations, and the wider prevention system, it is necessary to understand the key determinants of evaluation practice and how they act within health promotion. The challenges of evaluating health promotion strategies have been described in the literature.[11,49] As indicated in Chapter One, health promotion strategies typically address complex health issues, using multi-faceted approaches. A barrier to evaluation identified by practitioners is the collection of data from particular groups, for example young people, transient and vulnerable populations,[71,72] where building relationships can be perceived to be at odds with conducting evaluation.[73]

The literature presented in this chapter was drawn from peer-reviewed sources in the fields of health promotion, public health, and evaluation. Online databases Ovid MEDLINE, CINAHL Plus, ProQuest, EBSCOhost, Google Scholar and PsychINFO were searched for terms including (health promotion OR public health OR prevention) AND (evaluation capacity OR evaluation practice). Searches were further refined using terms (barriers OR challenges OR enablers OR facilitators). Key journals from the health promotion and evaluation fields were searched separately. Reference lists of key papers were also examined for additional articles.

In this chapter, the barriers and facilitators to evaluation practice identified within the health promotion literature are described. Next, conceptual and empirical models of evaluation capacity from other fields of practice are reviewed as these can contribute learnings to the health promotion context. Finally, in this chapter, the literature that describes the development and validation of evaluation capacity measurement instruments and conceptual models is reviewed, and the limitations of applying this body of work to the health promotion field are identified.

2.1 Barriers and facilitators to health promotion evaluation

A review of the peer-reviewed health promotion literature identified seven studies that sought to identify the barriers and facilitators to evaluation practice in a range of health promotion settings. Four studies were conducted in Australia within community health settings,[44,60] peer-led programs,[72] or government public health agencies.[74] Two studies were conducted in the Netherlands: in local government obesity programs,[75] and government funded national health campaigns.[59] A final study explored evaluation practices in community-based HIV prevention organisations in the United States of America (USA).[71] The diverse contexts of these studies reflect the spectrum of health promotion practices and organisations. While all studies drew on the experiences of multiple sites and both government agencies and NGOs, these studies were limited to single organisation types, addressing a single health issue, or in one jurisdiction.

All seven studies used qualitative data collection techniques, and the majority use thematic analysis[44,59,60,71,72,74] to explore barriers and facilitators to evaluation practice. Specifically, semi-structured interviews with individuals or small groups[44,59,71,74] or focus groups in combination with interviews were used.[72,75] One study conducted workshops to identify and agree upon the important barriers and facilitators to evaluation.[60] Purposive sampling was common; participants were recruited for a minimum level of experience, based on their role as program manager,[44,60,74,75] other senior roles, or those with knowledge of, or responsibility for evaluation.[59,71,74,75] The total participant numbers in interview or focus group studies ranged from 13[59] to 54.[71] One study indicated that data saturation was not reached,[75] however the remaining studies did not discuss saturation of themes.

A number of studies found practitioners had limited skills and knowledge which acted as a barrier to evaluation.[44,60,71,72,74,75] However, those with health promotion training were able to better apply evaluation knowledge than practitioners from non-health promotion backgrounds, including dietetics or occupational therapy.[44] Two studies also identified that those employed to provide expertise for evaluation lacked the required skills in the evaluation of community-based approaches,[75] or for HIV community-based organisation context.[71] Attitudes and beliefs about evaluation were cited as barriers, specifically that evaluation was not worth investing time in, that evaluation required specialists, and that evaluation was a threat to the program or organisation.[60,71,72]

Limited resourcing was frequently listed as a barrier to evaluation. A lack of time, funding and human resources that could or should be dedicated to evaluation were key themes in a number of studies.[44,59,71,72] Internal or external pressure to prioritise funding and time for program delivery was also identified to hinder allocation of sufficient funding to evaluation.[59,71,72,75] However, an organisational culture that encouraged evaluation, the use of findings[60,71,74] and leaders who could champion evaluation within the organisation[74] could facilitate evaluation. Collaboration that enabled access to expertise,

resource sharing and support was cited as a facilitator to evaluation in a number of studies,[59,71,75] especially if those providing technical expertise had good understanding of the program and organisational context.[71] Access to tools, data collection and reporting systems, and consistent use of evaluation frameworks were identified as facilitators of evaluation in several studies.[44,60,72,74,75]

Studies of the barriers and facilitators to evaluation in health promotion also identified external administrative or funding factors that could both drive and hinder evaluation. Reporting obligations to funders that were perceived to have little relevance or use were identified as a barrier to evaluation, and a burden on practitioner time.[44,71,72] However evaluation could also be facilitated if the evaluation requirements specified by a funder were clear and appropriate,[60,71,72] .In one study, the same reporting template was perceived as both a facilitator or barrier to evaluation by different participants, where a template was considered to either promote rigour of evaluation or conversely, create confusion and waste time.[44] Most studies identified low overall levels of funding for program delivery, let alone allocation of funds to evaluation.[59,71,72] Program funding cuts were identified as a barrier to evaluation, which was compounded by the imperative to demonstrate program success.[44,75]

The studies outlined above identify an array of factors that appear to influence evaluation conduct within health promotion organisations. However, the ability to apply these findings to the wider health promotion field may be limited, given that evaluation practice and capacity to date has only been examined in discrete practice settings. Further, these studies all used qualitative methods to explore evaluation practice in health promotion, restricting the potential to generalise findings to the wider field. The studies described above also varied in quality: while purposive recruitment methods were generally well described and sample sizes appear to meet rule-of-thumb guidelines, only one study discussed data saturation and this reported that saturation was not reached.[75] Additionally, there was limited or no attention to conceptual frameworks,[44,59,60,71,72,74] and some studies utilised only one

researcher for data analysis.[44,59,60,72] Consequently, exploring evaluation practice within health promotion requires consideration of the broader literature regarding evaluation practice and capacity.

2.2 Lessons from outside the health promotion field

The field of evaluation has contributed theoretical and empirical work to deepen understanding of the key components of organisational evaluation capacity, and frameworks for their interactions. This section introduces the conceptual frameworks that form the basis for empirical studies of evaluation capacity conducted in the evaluation field.

Two collections of studies conducted in government agencies and NGOs in Canada conceptualise evaluation capacity as the capacity to ‘do’ and ‘use’ evaluation for organisational learning.[76-82] Organisational learning capacity is also the focus of Preskill and Boyle’s[83] model of ECB, which highlights the importance of addressing individual knowledge, skills and attitudes to improve organisational learning capacity through ECB. Labin et al.[68] extended this work by integrating additional literature and outlining potential outcomes of ECB at the individual and organisational levels. The interaction between the individual and the organisation in determining evaluation capacity outcomes was also the focus of the synthesis model of evaluation capacity and an associated validation study.[84] The study by Taylor-Ritzler et al., conducted in not-for profit organisations based in Chicago, USA, found that the organisational factors, comprising leadership, learning climate and resources, appeared to fully mediate individual determinants of evaluation capacity.[84] More recently, Fierro[85] conducted a literature review and interviewed four ECB researchers to develop a framework distinguishing 11 evaluation capacity constructs from eight evaluation practice outcomes that may result from an organisation with effective evaluation practice. Further, a study of local government evaluators in Denmark tested the hypothesis that organisational evaluation capacity resulted from a combination of evaluation supply (technology, human capacity) and demand (objectives, structure and processes).[86] The authors found the factors representing the demand for evaluation was a stronger

contributor to overall evaluation capacity, compared to evaluation supply.[86]

Despite the different emphasis of each conceptual framework, these models incorporate several common determinants of evaluation capacity, or intended outcomes of ECB.

Leadership, organisational culture and systems, and evaluation resources are consistently recognised as important to organisational evaluation capacity.[68,76,78,80,83-87]

Organisational culture and leadership can also play a role in facilitating ECB with staff,[68,84,87] and access to support from within or outside the organisation can have a mediating effect on strategies to improve evaluation capacity.[87] Individual characteristics are also highlighted in models of evaluation capacity; for example, skills, knowledge and beliefs about evaluation are important.[68,76,83-86]

2.3 Measuring elements of evaluation capacity

Academics within the evaluation field have called upon ECB researchers to build on conceptual frameworks and develop valid measures and models of evaluation capacity.[65,66,87] There is a need for validated measurement instruments and frameworks to guide the planning, implementation and evaluation of ECB initiatives,[65,87] particularly to establish evidence of ECB effectiveness or impact.[88]

To date, the measurement development in evaluation capacity has focused on organisational or individual components of evaluation capacity, or outcomes of ECB. Specifically, at the individual level, evaluation capacity constructs for which scales or items have been developed include awareness, motivation and competence for evaluation[84] and human capital (including experience and volume of evaluation production).[86] Similarly, at the organisational level, scales have been developed for measurement of evaluation capacity constructs include evaluation objectives, structure and processes, technology,[86] leadership, learning climate and resources, and the two ECB outcomes of mainstreaming and use of findings.[84] In Canada, Cousins et al.[80] define six domains of evaluation capacity for which measurement tools have been developed: capacity for organisational learning,

organisational support structures (formalisation and training support), capacity to do evaluation, sources of knowledge, skill and ability, evaluative inquiry, capacity to use evaluation and conditions that mediate the use of evaluation.[80]

A separate survey instrument was developed to collect detailed information about the drivers for evaluation, evaluation conduct and use of findings in not-for profit organisations in social and community services.[89-91] This body of work also sought to capture information about the barriers and facilitators to evaluation and identified a lack of resources, limited access to evaluation expertise, implementation issues, disinterest from leadership and lack of support as challenges to evaluation.[91] Similarly, Fierro[85] sought to measure 19 elements of evaluation practice and evaluation capacity. In this instrument, aspects of evaluation practice included conduct, dissemination and use, and incorporation into organisational activities.[85] The evaluation capacity components of Fierro's survey comprised organisational learning, policies, evaluation frameworks, access to resources and information, engagement and support of supervisors, collective attitudes and knowledge of staff.[85] Self-report surveys were used in the majority of studies,[80,84-86,91] where the perspectives of internal evaluators,[80,86] program managers or directors[84,91] or a mix of program managers and evaluators[85] were captured.

2.4 The reliability and validity of existing evaluation capacity measures

A number of the instruments were developed in multiple stages, including combinations of literature review, qualitative research, key informant advice, and piloting with the target population,[78,80,84-86] contributing to the face and content validity of these tools. However, the psychometric properties of only three instruments have been evaluated.[80,84,86]

In the first study, eight of the nine Evaluation Capacity in Organisations Questionnaire (ECOQ) scales were found to have good internal consistency (alpha >0.8).[80] To assess the construct validity of the ECOQ scales and overall conceptual framework of evaluation

capacity, the authors used a combination of exploratory factor analysis and path analysis.[82] After some modifications, evidence of acceptable fit for the data was found in the model that included constructs of organisational support, stakeholder participation, evaluative inquiry, capacity to do evaluation, capacity to use evaluation, and organisational learning capacity.[82] Secondly, a Danish study that aimed to assess the construct validity of the Evaluation Capacity Index used confirmatory factor analysis (CFA) to assess the relationship between evaluation demand factors (objectives, structure and processes), and supply factors (technology, human capacity).[86] The study found good fit for the model of evaluation capacity, although the demand side of the model appeared to have a stronger effect compared to supply.[86] In the third study, the Evaluation Capacity Assessment Index (ECAI), developed by Taylor-Ritzler et al.[84] also demonstrated good internal reliability of all ten scales ($\alpha > 0.8$). Again, using CFA and structural equation modelling (SEM), the study confirmed the relationship between awareness, motivation, competence constructs and the latent variable of individual factors, and between leadership, learning climate and resources and a latent variable of organisational factors.[84]

As described in section 2.3, the need for further understanding of evaluation capacity constructs and the assessment of ECB initiatives in practice has driven the development of evaluation capacity measurement instruments.[66,87] However, despite the intentions of many instruments to be used in the evaluation of ECB initiatives,[84-86] there is limited evidence in the peer-reviewed literature of these instruments being used for this purpose.

2.5 Limitations of evaluation capacity and evaluation capacity building research

Despite some good progress in the development of validated instruments to measurement of evaluation capacity constructs,[82,84,86] methodological and practical limitations hinder the application of this research to health promotion. All three of the studies that used statistical techniques to validate a conceptual model of evaluation each used only one survey sample for stages of analysis.[82,84,86] One study acknowledged the potential risk of distortion of effects by their use of one sample for both measurement and model

validation.[84] Additionally, the evaluation literature frequently defines evaluation capacity as occurring within the organisation[68,76,78,80,83-87] and subsequently, the determinants of evaluation capacity described and measured encompass factors only modifiable within an organisational setting.

In contrast to existing models of evaluation capacity, the literature describing the barriers and facilitators to evaluation in health promotion indicate that there are several external factors within the prevention system that may bear directly or indirectly on evaluation capacity, such as funding availability and administrative requirements.[44,59,60,71,72,75] Further, while many of the studies of evaluation capacity emerging from the evaluation field do not describe the setting and context for their research in detail, it is likely that the nature of health promotion programs and organisations pose additional challenges to evaluation. For example, challenges described in the literature include the political nature of prevention program funding and approaches,[21] misalignment of funding models to expected program achievements[92] and multi-level strategies delivered concurrently and to diverse populations.[11,49]

Finally, another challenge for health promotion evaluation is that it is frequently expected to be completed in-house, often unrealistically, by health promotion practitioners.[17,50,51] This is particularly relevant given that two of the three existing validated models of evaluation capacity sought solely the perspectives of professional evaluators.[82,86] While models of evaluation capacity developed from within the evaluation field can offer important insights, there is a need to conceptualise models of evaluation capacity that specifically incorporate the determinants of evaluation relevant to health promotion.

2.6 Summary

The potential to improve evaluation capacity and quality in health promotion is reliant on understanding the important determinants of evaluation practice, and how they interact. In this chapter, an overview of the evaluation practice and capacity from the health promotion

and evaluation fields was provided. Findings from qualitative studies from the health promotion field provide some insight into the potential range and scope of practitioner-, organisational-, and system-wide factors that may facilitate or hinder health promotion evaluation. Lessons from the evaluation field also point to factors that act at the individual and organisational levels to influence evaluation capacity. However, the evidence available to date is of insufficient scope or quality to guide ECB efforts in health promotion. In the next chapter, the rationale for this thesis is outlined, the subsequent aims and objectives of this thesis are given, and an overview of the chapters contained in this thesis are provided.

3 Thesis rationale and aims

3.1 Thesis rationale

Primary prevention and health promotion initiatives are essential to address the public health challenges that face Australia, however to ensure these efforts are effective, the evidence base for prevention strategies must grow. As outlined in Chapter One, evaluation of health promotion programs has a lot to offer policy and decision makers, program managers and practitioners. Evidence of program effectiveness and implementation from quality evaluation can guide resource allocation, policy direction, program selection and adaptation, and inform program learning and improvements. To improve the comprehensiveness and rigour of health promotion evaluation, there is a need to firstly understand what influences evaluation capacity in the health promotion field.

In Chapter Two, a review of the health promotion literature identified an emerging body of research that describes the barriers and facilitators to evaluation practice in health promotion settings. This provided insights into the range factors that appear to facilitate or hinder evaluation at the individual practitioner-, organisational- and prevention system-level. However in this review, only seven qualitative studies were identified, and each study focused on a discrete area of health promotion practice. To further understand evaluation capacity, the literature from the field of evaluation and ECB was also examined. The field of evaluation has progressed from the development of conceptual models for organisational evaluation capacity, and has begun to develop and validate instruments to measure aspects of organisational evaluation capacity. Despite some similarities between the barriers and facilitators identified in the health promotion literature and key aspects of evaluation capacity identified in the field of evaluation, there remains a number of limitations to applying this work directly to health promotion. Firstly, existing conceptual models and

measurement instruments focus on evaluation capacity within the organisation, leaving the system-level barriers and facilitators that appear to affect health promotion evaluation largely unconsidered. Further, there have been few attempts to validate evaluation capacity models and instruments. Of the three validation studies, two targeted evaluation professionals, and none have been validated within health promotion.[80,84,86] Finally, no studies that examined the association between health promotion evaluation capacity and evaluation quality were found.

The research in this thesis addresses a gap in the area of health promotion and primary prevention evaluation literature, contributing to knowledge on the determinants of evaluation capacity, and the associations with evaluation quality. A deeper understanding of components of health promotion evaluation capacity and how they interact with evaluation quality will provide guidance to policy makers, health promotion organisations and practitioners seeking to build evaluation capacity, and ultimately improve the effectiveness of prevention strategies.

3.2 Thesis aims and objectives

The aims of this thesis are to:

1. Determine the reported strengths and limitations of evaluation practice in the Australian health promotion field; and
2. Identify factors that need to be addressed to improve evaluation quality.

In order to address these aims, the specific research objectives are to:

1. Identify the influences upon evaluation practice as described by health promotion practitioners and managers in prevention agencies in Australia, and explore how these operate;
2. Develop and evaluate the psychometric properties of a survey instrument to measure evaluation capacity factors;

3. Evaluate a conceptual model that describes the interaction of the evaluation capacity factors;
4. Describe the levels of evaluation and data collection methods reported in Australian health promotion evaluation; and
5. Assess the quality of reported evaluations and identify the important organisational attributes and elements of evaluation capacity associated with evaluation quality.

3.3 Thesis overview

This thesis comprises nine chapters. In Chapter One, the background to this thesis was discussed. A review of the literature relating to evaluation practice in health promotion and lessons from the evaluation literature about the measurement of evaluation capacity was presented in Chapter Two. In this chapter (Chapter Three), the rationale for the thesis, and the research aims and objectives were listed. In Chapter Four, an overview of the mixed-methods research project, detail of participant recruitment and the data collection methods and analysis for each of the four sub-studies is outlined. Chapters Five to Eight present details of the methods and results for each phase of the research project. Chapter Five, which describes the Phase 1 interview study that examines the determinants of health promotion evaluation, contains two manuscripts published in *Health Education Research* and *Prevention Science*. Chapter Six contains a manuscript published in *Evaluation and Program Planning* which details Phase 2A development and validation of the EPAS. Chapter Seven includes a manuscript submitted to *Health Education Research* which describes the validation of a conceptual framework of health promotion evaluation capacity using a second sample of EPAS data from Phase 2B. Chapter Eight, which contains a manuscript submitted to the *Health Promotion Journal of Australia*, presents Phase 3 of the study involving the audit and appraisal of health promotion evaluation, and identifies the organisational attributes and evaluation capacity factors associated with evaluation quality. The final chapter (Chapter Nine) contains an integrated discussion of the research, including the key findings, and identifies future directions for ECB. Additionally, in Chapter Nine

recommendations for health promotion practitioners, organisations, policy makers and researchers seeking to improve evaluation capacity in health promotion are provided.

3.4 Summary

In this chapter, the rationale for this thesis, which is grounded in the need to improve the contribution of health promotion evaluation to the evidence base for prevention strategies, and address the limitations of the existing evaluation capacity literature, was outlined. The aims and objectives for this thesis were given and a synopsis of the chapters contained in this thesis was presented. In the following chapter (Chapter Four), an overview of the study design, participants and methods that were used in this thesis is given.

4 Study design, participants and methods overview

In this chapter, an overview of the methodology and study design for this mixed-methods research project, and outline of each of the four sub-studies included in this thesis, is provided. The process of recruiting government agencies and NGOs to the study is described, including ethical approvals obtained, and an overview of the sample of organisations that participated in the three main phases of data collection is given. To conclude this chapter, the data collection methods and analysis for each of the four sub-studies is summarised.

4.1 Study methodology and design

To meet the aims of this thesis, a mixed-methods approach was adopted. Mixed-methods research considers both the positivist perspective of quantitative research and constructivist perspective of qualitative methods, and is particularly useful in answering research questions that involve ‘real world’ circumstances.[93] Mixed-methods designs can help overcome the limitations of qualitative or quantitative methods alone,[94] allowing triangulation of designs and data collection methods that can strengthen the weight and completeness of findings.[93,94] A sequential quantitative dominant mixed-methods design was used.[95] This mixed-methods study comprised three main phases of data collection: 1) in-depth semi-structured interviews; 2) an online survey; and 3) audit and appraisal of prevention program evaluation reports, and involved four sub-studies.

The initial qualitative phase was conducted to explore the perspectives of health promotion practitioners and managers about the most important determinants of evaluation practice, and how these influences interact. The qualitative phase was also used to inform the development of a measurement instrument for assessing aspects of evaluation practice and

capacity within health promotion; this instrument was used in subsequent phases of the research.

The dominant quantitative component of the mixed-methods study was adopted to: 1) assess the psychometric properties of the newly developed instrument; 2) examine the relationships between the determinants of evaluation capacity and practice in Australian health promotion agencies; 3) describe evaluation quality and methods used by these agencies; and 4) identify the key factors associated with evaluation quality. Quantitative methods were chosen for their utility in capturing data from a large population, and for the potential to achieve a representative sample that can be generalised to the population of interest.[96] Further, statistical methods are necessary to assess the reliability and validity of measurement instruments.[97]

4.2 Ethics

Given that the research was based within organisations, and the multi-jurisdictional nature of the research, multiple Human Research Ethics Committee approvals were required.

Ethical committee approval to conduct this study was received from:

- Monash University Human Research Ethics Committee [#6728, 15th Sept, 2015];
- South Australian Department of Health and Ageing Human Research Ethics Committee [HREC/16/SAH/16, 16th May, 2016];
- Western Australia Country Human Research Ethics Committee [#2016/11, 24th May, 2016]; and
- Aboriginal Health and Medical Research Council ethics committee [#1156/16, 29th Mar, 2016].

4.3 Organisational identification and screening

4.3.1 Inclusion criteria

The primary participants for this study were Australian government agencies or NGOs in

New South Wales, South Australia, Victoria or Western Australia that conducted health promotion (primary prevention) initiatives. All levels (local, state and national) of government and NGOs (not-for-profit) were considered for inclusion given they are the primary organisations delivering health promotion initiatives in Australia. For-profit and private organisations were not considered for this study as they comprise a small minority of organisations undertaking health promotion activities in Australia. The four states were selected for two reasons. Firstly, each state has a well-established tradition of delivering health promotion activities, albeit via funding mechanisms and organisational structures unique to their respective jurisdictions. Secondly, limiting primary study recruitment to selected jurisdictions would allow for a separate sample of individual participants to be invited from the Australian Capital Territory, Northern Territory, Queensland and Tasmania to participate in the instrument development sub-study (for more detail, refer to Chapter Four). Additional criteria based on the size and experience of the organisation were established prior to commencement of screening. It was anticipated that organisations that met the criteria would have sufficient capacity to allow staff to participate in the interviews and surveys, and to have implemented health promotion strategies that were likely to have been evaluated.

The minimum criteria for health promotion organisational participants in the study were:

- An annual health promotion budget (including staff salaries) of at least \$300 000 as this indicates substantial resources, making the chances for evaluation more likely;
- At least 3 full-time equivalent (FTE) staff dedicated to health promotion activities; and
- A minimum 5 years of experience of delivering health promotion activities.

4.3.2 Identifying eligible health promotion organisations in Australia

As there is no single register of organisations conducting health promotion activities in Australia, potential organisations in each of the four selected states were mapped. Internet

searching, and telephone calls to experienced health promotion practitioners and researchers was used to compile a list of potential participants. Key health promotion network coordinators from within government departments and NGOs were contacted, and invited to share their contact lists if possible, or suggest alternative contact points. Commencing in October 2015, organisations were contacted to assess eligibility for the study, and potential interest in participating. In the same contact, advice was sought in regards to other organisations that may be eligible to participate. A database of each organisation's contact details, key contact persons, potential eligibility and referral networks was maintained. Saturation of organisations was reached once no new suggestions for that region were received.

Seventy-seven key health promotion contacts, from peak professional bodies, government agencies and NGOs, including (but not limited to) Primary Health Networks and Primary Care Partnerships (Victoria only), state health departments and local government, were approached for advice or network information. Based on key contact advice, over 500 organisations were screened. Twenty organisations were likely eligible, but declined involvement in the study at the screening stage primarily due to lack of interest (n=9,) or lack of time or capacity (n=10). A final sample of 154 organisations were deemed eligible to participate in the study. The number of eligible organisations in each state is shown in Table 2. Fewer eligible organisations than expected were identified in South Australia. Several key contacts based in South Australia indicated that health promotion functions and staff had been significantly reduced around the state since a review of health services deemed preventive services to have little demonstrable value.[98] A large number of NGOs were identified in Victoria, which reflects that state government funding for health promotion is allocated to non-government community health services that operate at a local or regional level.

Table 2. Summary of eligible and participating organisations

Organisation type	New South Wales	South Australia	Victoria	Western Australia
	Recruited (Eligible)	Recruited (Eligible)	Recruited (Eligible)	Recruited Eligible
Government	17 (22)	4 (8)	9 (15)	12 (19)
Non-government	13 (14)	9 (11)	39 (49)	13 (16)
Total	30 (36)	13 (19)	48 (64)	25 (35)

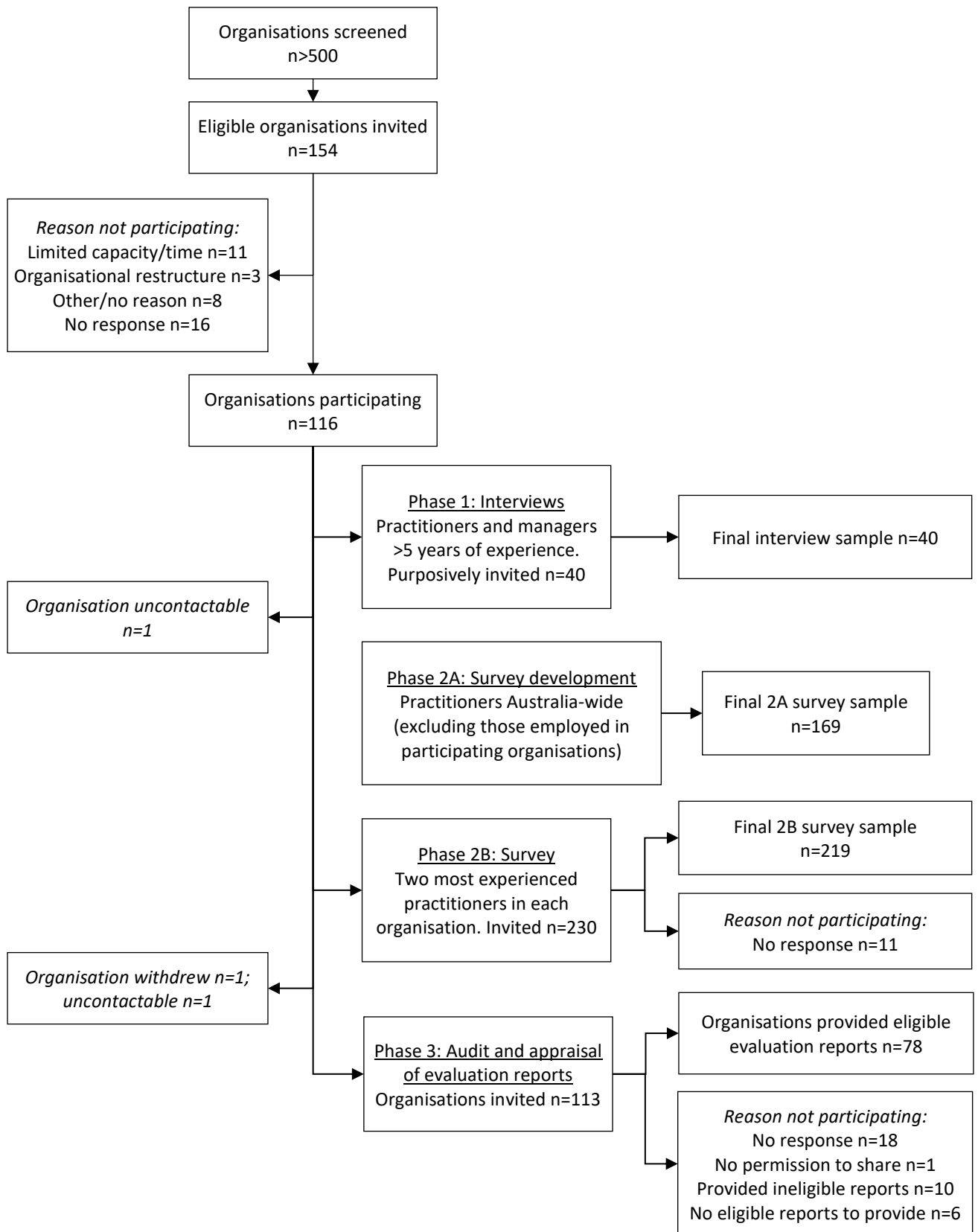
4.4 Organisational recruitment

Figure 1 shows the stages of recruitment and data collection for the study. Recruitment to the study began in January, 2016. The 154 eligible organisations were contacted via their nominated representative. The project information sheet was provided by email, and if the organisation agreed to participate, permission was returned via email or letter. In total, 116 (75%) organisations returned written permission to participate. Of those which declined, the two primary reasons were limited time or capacity to participate and uncertainty given organisational restructures. One organisation declined to participate as they felt their evaluation processes were not developed enough to be examined by an external organisation, while another did not have permission to share their evaluation reports, so declined participation in the whole project.

4.5 Participating organisations

A breakdown of the participating organisations, by type and state, is shown in Table 2. NGOs comprised 64% of the sample. The representativeness of the organisations within each state could not be ascertained, as the number of organisations conducting health promotion and disease prevention work has not been determined. Nevertheless, a broad range of organisations were recruited due to the comprehensive screening and sampling strategy.

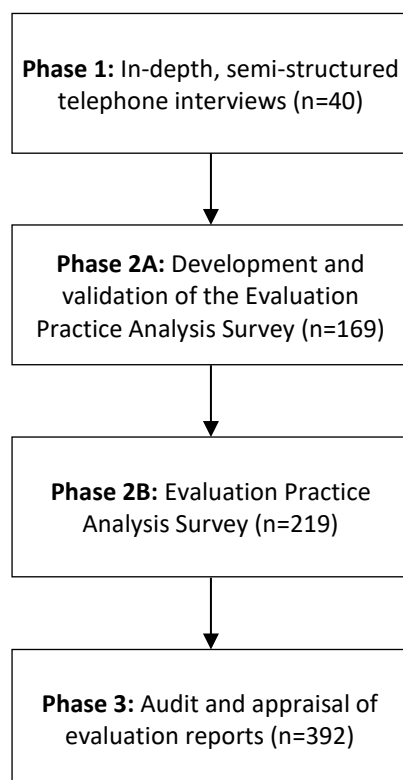
Figure 1. Stages of participant recruitment



4.6 Data collection and analysis overview

Data were collected in three main phases, comprising four sub-studies (Figure 2). The phases of data collection were: Phase 1 qualitative interviews with experienced health promotion practitioners and managers in participating organisations; Phase 2A validation stage of the EPAS, collected from health promotion practitioners from outside of participating organisations; Phase 2B the EPAS with respondents from organisations participating in the main study; and Phase 3 audit and appraisal of health promotion program evaluation reports provided by each participating organisation. An overview of each phase of the research project is shown in Figure 2.

Figure 2. Phases of data collection



4.6.1 Phase 1: Interviews

In-depth, semi-structured telephone interviews were conducted with 40 health promotion practitioners and managers between February and June 2016. Practitioners from participating organisations were purposively invited if they had least 5 years of experience in

health promotion practice. Five years' working in the health promotion was considered sufficient to have encountered a range of experiences in health promotion practice, including evaluation. All eligible participants who were invited agreed to participate. The sampling frame allowed for representatives from government agencies and NGOs based across the four states, a range of organisational sizes, and metropolitan and regional locations. Organisational characteristics were reported by the participant, and regional status confirmed using the Australian Statistical Geography Standard Remoteness Area classification where necessary. This phase of data collection aimed to identify the influences upon evaluation practice from the perspectives of health promotion practitioners. Thematic analysis was used to extract the most important practitioner-, organisational- and system-level determinants of evaluation practice in the field. The results of this study were used to develop the domains (i.e. the important determinants of evaluation practice) and items (i.e. key indicators for each domain) for the Phase 2A survey instrument.

4.6.2 Phase 2A: Survey validation

A sample of health promotion practitioners from outside of the participating organisations were recruited to complete the newly developed EPAS. The survey was developed in multiple stages, including: construction of a conceptual framework and survey items based on Phase 1 interviews and literature synthesis; review by experienced health promotion practitioners and researchers; cognitive interviews with health promotion practitioners (n=10); and pilot testing with a further six practitioners. The instrument was revised based on each stage of feedback to improve face validity, clarity of item wording and response options. The initial online survey comprised 158 items, grouped into 15 sections that described practitioner-, organisational- and system-level aspects of evaluation capacity. The EPAS was widely distributed online via key health promotion contacts, snowball sampling, and publically available contact details for health promotion organisations. A total of 169 respondents took part in the survey between November 2016 and February 2017. Principal components analysis (PCA) was used to determine scale structure, and logistic regression to examine the

predictive validity of selected scales. Findings from Phase 1 also informed the selection of variables in the logistic regression, and were used in the interpretation of the quantitative findings.

4.6.3 Phase 2B: Survey implementation

The validated EPAS was emailed to two nominated contacts in each of the participating organisations. Health promotion practitioners and managers took part in the online survey (n=219) between June and September, 2017. The data were first analysed using CFA to further assess the validity of scale structure and secondly, path analysis was used to assess the fit of a conceptual model that was developed and refined based on the findings from the interview study in Phase 1 and results of the predictive validity test in Phase 2A. The model described the relationship between seven determinants of evaluation capacity and a self-reported measure of impact evaluation.

The survey data were also used in the final study where the association between the determinants of evaluation practice and evaluation quality was explored by linking data from Phase 2B and Phase 3, as described below.

4.6.4 Phase 3: Documentary audit and appraisal

The final phase of data collection was an audit and quality appraisal of evaluation reports completed by participating organisations in a two year period. All evaluation reports completed between May 2015 and May 2017 were requested from participating organisations. Evaluation report collection took place between June and December, 2017. A total of 392 evaluation reports, from 78 of the participating organisations, were audited and appraised. The audit and appraisal tool was developed specifically for this study in several stages: review of existing evaluation quality guidelines, consensus-based item selection, piloting, and refinement. The tool collected program information, evaluation details, evaluation levels and data collection methods, and included a 22-item quality appraisal section. The audit and appraisal data were linked with Phase 2B EPAS data to identify

factors that related to evaluation methods and quality. Where two EPAS responses for an organisation were provided, the mean score for each EPAS scale was calculated. Specifically, organisational attributes, and measures of practitioner-, organisational-, and system-level evaluation capacity, were examined in relation to evaluation quality.

4.7 Summary

In this chapter an overview of the study design for this mixed-methods research project was provided, the screening and recruitment of organisational participants to the study detailed, and the characteristics of participating organisations described. The data collection methods and analysis used in each of the four sub-studies were also outlined. The following four chapters (Chapters Five to Eight) contain manuscripts in which the four sub-studies that comprise this thesis are described in detail.

5 The factors that influence health promotion evaluation: findings from a qualitative study

The literature review in Chapter One illustrated that, despite recognition of the importance of high quality and comprehensive evaluation evidence to inform health promotion practice, there is limited understanding of the most important determinants of evaluation in the health promotion field. Consequently, the qualitative study presented in this chapter aimed to identify the important practitioner-, organisational-, and system-level influences upon evaluation practice according to experienced health promotion practitioners and managers, and to examine how these factors act to influence evaluation practice. In-depth, semi-structured interviews (n=40) were conducted and the data analysed thematically. Additionally, in the context of the overarching mixed-methods study, this qualitative phase was used to inform the development of a survey instrument which is further discussed in Chapter Six.

This chapter contains two published papers which present the details of the study conduct and results. In the first paper:

Schwarzman J, Bauman A, Gabbe B, Rissel C, Shilton T, Smith BJ. Organizational determinants of evaluation practice in Australian prevention agencies. *Health Education Research*. 2018;33(3):243–55,

the findings that focus on the organisational-level determinants of health promotion are presented. These findings identify leadership, organisational culture, systems and structures, partnerships, resources, workforce development, and recruitment and skills mix as important factors that influence evaluation.

In the second paper:

Schwarzman J, Bauman A, Gabbe BJ, Rissel C, Shilton T, Smith BJ. The Funding, Administrative, and Policy Influences on the Evaluation of Primary Prevention Programs in Australia. *Prevention Science*. 2019. doi:10.1007/s11121-019-00997-4,

key themes emerging from the interview data that concern the determinants of health promotion evaluation in the wider prevention system are described. The contributions of funding, policy and administrative arrangements for health promotion programs to evaluation conditions in health promotion organisations are explored. In this paper, how organisations interact within the prevention system to modify conditions for evaluation was examined.

Organizational determinants of evaluation practice in Australian prevention agencies

J. Schwarzman^{1*}, A. Bauman², B. Gabbe¹, C. Rissel³, T. Shilton⁴ and B.J. Smith^{1,2}

¹School of Public Health and Preventive Medicine, Monash University, 553 St Kilda Road, Melbourne, VIC 3004, Australia,

²Prevention Research Collaboration, School of Public Health, L6 The Hub, Charles Perkins Centre, University of Sydney,

Sydney, NSW 2006, Australia, ³NSW Office of Preventive Health, Don Everett Building, Level 1, Liverpool Hospital,

Liverpool, NSW 2170, Australia and ⁴National Heart Foundation Western Australia, 334 Rokeby Road, Subiaco, WA 6008, Australia

*Correspondence to: J. Schwarzman. E-mail: joanna.schwarzman@monash.edu

Received on October 10, 2017; editorial decision on March 5, 2018; accepted on April 20, 2018

Abstract

Program evaluation is essential to inform decision making, contribute to the evidence base for strategies, and facilitate learning in health promotion and disease prevention organizations. Theoretical frameworks of organizational learning, and studies of evaluation capacity building describe the organization as central to evaluation capacity. Australian prevention organizations recognize limitations to current evaluation effectiveness and are seeking guidance to build evaluation capacity. This qualitative study identifies organizational facilitators and barriers to evaluation practice, and explores their interactions in Australian prevention organizations. We conducted semi-structured interviews with 40 experienced practitioners from government and non-government organizations. Using thematic analysis, we identified seven key themes that influence evaluation practice: leadership, organizational culture, organizational systems and structures, partnerships, resources, workforce development and training and recruitment and skills mix. We found organizational determinants of evaluation to have multi-level interactions. Leadership and organizational culture influenced organizational systems, resource allocation and support of staff. Partnerships were important to overcome resource deficits, and systems were critical to embed evaluation within the

organization. Organizational factors also influenced the opportunities for staff to develop skills and confidence. We argue that investment to improve these factors would allow organizations to address evaluation capacity at multiple levels, and ultimately facilitate effective evaluation practice.

Introduction

The growing burden of chronic disease, and persistence of social disparities in health status, requires continued development of evidence to underpin effective prevention strategies [1, 2]. Evidence of what works, for whom, and in which circumstances is essential to inform policies and programs that can address current population health priorities [3]. It has been widely argued that evidence generated through program evaluation can provide contextually relevant insights for decision makers and practitioners, enable accountability to funders and community stakeholders, facilitate organizational and program level learning and improve effectiveness [4–6].

Despite the known importance and demand for evaluation, there are significant challenges to undertaking evaluation and using the evidence generated [7–9]. In response to these, the need for workforce development and partnerships between research and practice organizations has been highlighted [3].

Strategies have been implemented to build capacity for evaluation practice and use to meet this demand, including training [3, 10] formation of networks and partnerships [11] and refinement of evaluation designs to improve the practice relevance of evidence generated [12]. The increased attention to evaluation capacity building is a promising sign, however there is limited evidence of what influences evaluation practice and use in prevention organizations.

Australia has a well-developed infrastructure for prevention policy and programs, encompassing a diverse range of government and non-government organizations (NGOs), yet there have been few studies of the organizational barriers and facilitators to evaluation practice in this context. Three studies used qualitative methods to explore these influences in specific health promotion settings. In the evaluation of peer-based programs, limited capacity and funding, beliefs about evaluation and a reliance on volunteers were identified as barriers [13]. In the South Australian community health setting, time, resources, culture of evaluation and limited use of findings hindered effective evaluation practice [7]. Interviews with practitioners and managers in metropolitan Melbourne prevention organizations identified funding, staff skills and access to appropriate tools for evaluation as influential within the organization [4]. Huckel Schneider *et al.* [14] interviewed policy makers and researchers in Australia about barriers and facilitators to evaluation, and at the organizational level identified limited time, inadequate funding, caution over negative outcomes and inadequate staff experience with evaluation as barriers, and a culture of sharing and learning as an important facilitator.

These findings align with theoretical models which highlight the importance of organizational factors as determinants of evaluation capacity [15, 16], and are consistent with what has been found in prevention agencies in other nations. A qualitative study of Dutch local government organizations implementing health programs described time, budgetary constraints, limited leadership and program experience, unclear role expectations and competing priorities as barriers to evaluation [6].

A number of authors have emphasized the importance of placing evaluation practice and use in the context of organizational learning capacity [17, 18]. From this perspective, the need to address organizational culture and systems, and to go beyond the traditional training and development approach, has been recognized as integral to evaluation capacity building. Preskill and Boyle's [17] multi-disciplinary model of evaluation capacity building embeds evaluation capacity within a framework of organizational learning, and describes elements of organizational culture, leadership, systems and structures and communication as central influences upon the evaluation capacity of organizations.

Despite the increasing demand for rigorous and practice-relevant evaluation, and the growing body of literature on evaluation methods and capacity building, there are still gaps in understanding of the determinants of evaluation practice in prevention organizations. In the Australian context, studies to date have been limited in scope and not explored the determinants of evaluation practice across government organizations, NGOs and multiple jurisdictions. Furthermore, the way that organizational characteristics, structures and processes influence evaluation practice has not been explored and these insights are vital for successful capacity building. Therefore, this study aimed to investigate the influences upon evaluation described by managers and practitioners in government and non-government prevention agencies in Australia, to explore how these operate, and identify implications for evaluation capacity building.

Materials and methods

This qualitative study, utilizing semi-structured, in-depth interviews, was undertaken in the context of a larger mixed-methods project, in which findings from this first phase will inform future phases of the research. Forty prevention practitioners and managers were identified from four states' of Australia. Sampling was purposive and stratified [19] including both government organizations and NGOs. Criteria were set to ensure a minimum level

of prevention experience (≥ 5 years) for individuals and minimum organization staffing levels (three full time equivalent prevention staff), budget (\$300 000 or greater) and an organizational history (≥ 5 years) in conducting prevention activities. The sample size was considered sufficient to reach saturation of themes within government and NGOs, and across jurisdictions.

Interview participants were recruited from New South Wales ($n = 10$), Victoria ($n = 11$), Western Australia ($n = 10$) and South Australia ($n = 9$) and represented a mix of government organizations ($n = 18$) and NGOs ($n = 22$). Government organizations included local and state government, while NGOs comprised community health services, Aboriginal community controlled and mainstream public health agencies. All eligible participants invited agreed to participate. Participants had a mean (range) of 15 (5–30) years of experience in the field of health promotion and primary prevention and a mean (range) of 9 (0.9–25) years within their organization. Most participants were employed as managers or directors ($n = 29$) at the time of interview, with the next largest groups being senior practitioners ($n = 4$) or team leaders ($n = 4$). Participants held a wide range of undergraduate qualifications, and 18 participants had post-graduate qualifications in either health promotion or public health.

A semi-structured interview guide was developed based on a review of evaluation practice and capacity building literature, with particular reference to the health promotion and disease prevention field, a previous pilot study describing determinants of health promotion evaluation practice in government organizations [4] and input from experienced health promotion researchers. The interview guide explored three levels of potential influence as identified in the literature: individual, organizational (i.e. primary delivery agency) and system level (e.g. policy, accountability requirements) factors. In this paper, we focus on methods and findings related to organizational level factors, which were discussed extensively by interviewees. The interview questions explored five topics: recent organizational evaluation practice, drivers of evaluation, organizational level facilitators and barriers to effective

evaluation and recommendations to improve evaluation capacity. Demographic questions were included to capture years of experience, role title, field of work, qualifications and training and organizational size and location. The final interview guide consisted of 10 open-ended questions and was piloted with 3 practitioners. The interviewer allowed the participant to guide the direction of the interview, while ensuring each topic was covered.

Interviews were conducted by author (JS) via telephone and were audio recorded. Participants were emailed an information sheet prior to the interview and consent was obtained verbally at commencement of the interview prior to recording. The interviewer has worked in the prevention sector and has qualitative research training. Prior to, and during the interview, the interviewer documented their potential biases and opinions about the subject, and maintained notes on their reactions and reflections after each interview. The interviews were conducted between February and June 2016 and averaged 42 min (range 28–62) in duration.

Ethics approval was received from Monash University HREC, Aboriginal Health and Medical Research Council Ethics, South Australian Department of Health and Ageing HREC.

Audio recordings were transcribed and provided to each participant for review and comment. Final transcripts were imported into NVivo 11 software for coding. Two researchers (JS and EM) commenced the first round of coding using the interview guide topics of individual, organizational and system-level influences as a guide and an iterative approach was taken to the development of the coding framework and new codes based on the data. JS and EM dual coded 10 transcripts, meeting regularly to discuss any discrepancies in coding until agreement was reached. The framework was discussed regularly with a senior researcher (BJS) for clarity of concept and refined and applied to the remaining data. Further rounds of coding identified processes and values and was analysed thematically by JS and EM together to identify patterns of barriers and enablers to evaluation practice, and links between themes [20].

Table 1. *Summary of key themes*

Key themes Factors that influence evaluation practice and use	Features and practices that facilitate evaluation	Barriers to evaluation
<p>Leadership Leadership through official role or position in organization, and ability to influence others from that position.</p>	<p>Champion evaluation. Valuing and understanding evaluation. Demonstrating benefits of evaluation. Supporting staff to develop evaluation skills. Embedding practices into culture, systems, structures. Facilitating partnerships, relationships. Valuing evaluation and evidence. Valuing health promotion. Learning about program effectiveness, improvement and sharing findings. Culture that is supportive of staff learning. Long term vision and approach to evaluation.</p>	<p>Lack of leadership, or frequently changing leadership. Lack of understanding or value of health promotion. Lack of value of evaluation.</p>
<p>Organizational culture Can take time to develop, and is dependent on leadership at some stage. Connects to the values of the organization, and diverse organizational cultures can influence evaluation practice in different ways.</p>	<p>Communication systems. Resource allocations systems. Clear reporting systems and frameworks against organizational plans (e.g. work plans, strategic plans). Dedicated evaluation role.</p>	<p>Burdensome systems. Lack of system, clarity for reporting, or agreement on indicators. Unclear expectations of responsibility for evaluation. Lack of expert or 'go to' role.</p>
<p>Organizational systems and structures Organizational processes, internal systems and structures of teams and workforce. Can be developed through leadership and interacts with organizational cultures. Often aligns with reporting and accountability requirements.</p>	<p>Shared and agreed indicators, goals and objectives at the project planning stage. Acknowledge respective strengths, mutual benefit for each partner. Long term relationships with evaluation advisors. Organizational partnerships with universities. Professional networks with similar organizations to share tools and other resources.</p>	<p>Partnerships that were felt to be unequal (e.g. being used for data provision). Partnerships that were more time consuming than beneficial.</p>
<p>Partnerships New partnerships formed for the purpose of evaluation, also the influences on evaluation within existing partnerships.</p>	<p>Planning and allocating budget to evaluation in advance (including roles, external evaluators etc.). Using students, volunteers or partners to assist in evaluation. Allocating time for evaluation planning and conduct, reflection and use of findings. Access to frameworks, indicators and data collection instruments for health promotion priority areas.</p>	<p>Competing with delivery of programs, or the delivery of health services. Insufficient health promotion budget overall. Workloads. Lack of access to appropriate data collection instruments. Unclear, or lack of agreement on appropriate frameworks for evaluation.</p>
<p>Resources Access to and use of resources required for evaluation including funding, time, appropriately skilled workforce, technology, evaluation frameworks, guides and instruments.</p>		

(continued)

Table I. Continued

Key themes Factors that influence evaluation practice and use	Features and practices that facilitate evaluation	Barriers to evaluation
Workforce development and training The role of the organization in developing the skills, knowledge, beliefs and attitudes of staff towards evaluation.	Fostering positive attitudes through a belief in evaluation benefits and the value of evidence-based decision making. Support to develop technical, interpersonal and other professional skills. Pathways to access evaluation skill development opportunities.	Culture of fear of negative findings and consequent budget cuts. Beliefs that all funding should go to delivering programs. Excessive demands on practitioners to be skilled in all areas of health promotion. Some backgrounds (e.g. sports science) had limited understanding of evaluation.
Recruitment and skills mix Recruiting staff to ensure appropriate evaluation skills and expertise.	Qualifications, training and experience contribute to skills for evaluation. Evaluation requirements embedded within job descriptions and aligned with program and work plans. Recruitment practices to ensure skills and approach to evaluations match organizational requirements.	New graduates had limited skills in planning and evaluation.

Results

Seven key themes were identified as important factors that influence evaluation practice and use in health promotion and disease prevention organizations (Table I). Within each theme, the way these factors act to facilitate or hinder effective evaluation practice and use are explained.

Leadership

Leadership to facilitate evaluation practice was described as having a range of attributes and was similar across government organizations and NGOs. The main pathways through which leadership was influential were championing evaluation and promoting the value of health promotion and evidence; making decisions about resource allocation; the ability and networks to initiate and sustain partnerships to support evaluation and, helping to embed these elements into organizational culture, systems and structures.

Leaders were largely described as those in senior roles in the organization who influenced practice through a 'top down' approach, but effective leadership for evaluation could also come from others, such as research and evaluation officers. Leaders that supported evaluation practice in the organization demonstrated an ability to effect change, had the expertise to argue the benefits of evaluation, to arrange skill development for staff and for themselves and successfully established and managed collaborations that supported evaluation.

'I think leadership and vision from a director is extremely important and I can think of two leaders . . . that have had vision, and they've had the skills to be able to sell that to [organization executive] and to be able . . . carry it out very well, and so I think they've got to have the interest, the vision and the leadership' (P13, Senior Practitioner, Government).

Some participants also described the type of leadership that had a negative effect on evaluation practice, including disengaged or passive leadership and organizational structures where leadership changed

often, limiting real influence and decision making power. A lack of understanding of health promotion practice was also detrimental to evaluation of prevention programs and use of evaluation findings.

Organizational culture

Organizational values were integral to creating a positive culture for evaluation. Valuing health promotion and disease prevention work appeared to be a necessary foundation for health promotion evaluation. Additionally, as expected, organizations that placed value on evaluation as a practice and using evidence to inform decision making and program planning also created a supportive organizational environment for evaluation.

‘...when I came into tobacco control I was struck very much by how much of a role that data and evaluation played in what was happening.’ (P25, Manager, Government).

A culture of evaluation was a clear facilitator, but not the only way to enable quality and effective practice. Participants, particularly from NGOs, described the culture of improvement, learning and sharing as crucial to undertaking and using evaluation in health promotion. Additionally, a culture of support was important to allow staff time to undertake evaluation, seek expertise and advice from evaluation specialists, engage with community members and program partners and to develop their own evaluation skills and capacity. Key outcomes of a facilitating culture were that evaluation practice was rigorous, meaningful, shared with program staff, celebrated their success and was embedded within systems and structures of the organization.

‘[Evaluation practice] is really embedded, it’s not something you turn attention to once a year’. (P22, Director, NGO).

Participants also spoke about a lack of understanding and limited value placed on prevention, often in favour of health service delivery, leading to a difficult organizational environment for evaluation. In some organizations, participants reported that a limited understanding or value of health

promotion practice and evaluation was associated with a demand for minimal quantitative reporting. Practitioners reported the tension this created when working in complex programs in particular, and the challenges of reporting evaluation findings that were meaningful for decision makers accustomed to dealing with patient numbers or purely quantitative data.

‘I think being embedded within a health service can be difficult... [The board activity report] does not include any data about health promotion because we can’t count it... So because we can’t count it, it’s not viewed as being as important’. (P6, Director, NGO).

Organizational systems and structures

Effective organizational processes directly facilitated stages of evaluation, such as planning evaluation or collecting data. More broadly, systems that contributed to effective evaluation were described as those that supported clear communication within the organization and with external stakeholders, facilitated reporting against work plans and had clear pathways and timelines for approvals and resource allocation. The role of systems and structures to encourage evaluation was more prominent in responses from government agencies compared to NGOs.

The introduction of systems to facilitate effective evaluation was often related to the presence of clear organizational structures including reporting lines and dedicated evaluation positions. Leadership, culture and availability of time were also discussed as important contributors in ensuring systems were implemented and maintained across the organization.

‘I do come from an evaluation background myself so I was very keen to make sure that we had a solid evaluation protocol and discipline when I first took the job and so I made sure that all of our major program work was independently evaluated’. (P12, Manager, NGO).

Participants that described the use of organizational plans to facilitate and embed evaluation practice in

the organization tended to be from government organizations or larger NGOs.

‘... it goes to the work plans that we have and whether they’re personal work plans, team work plans, annual plans, 5 year plans monitoring and evaluation should run through every layer of that ...’ (P22, Director, NGO).

These plans guided reporting and supported allocation of resources to evaluation roles, or practitioner time. Clear communication, opportunities to meet, discuss and reflect on evaluation based on the work plan was important in ensuring the document remained active and useful. The alignment of the plans added clarity across teams and the organization, promoting consistency and shared goals between teams and practitioners.

Also discussed was the concept of flexibility within plans to allow for unexpected opportunities and changes. Systems that were too rigid and burdensome could hinder evaluation through untimely approval processes, unwieldy paperwork, or missing opportunities. At the other end of the spectrum a lack of reporting systems around health promotion program work and evaluation appeared to contribute to ad hoc and piecemeal evaluation practice.

Partnerships

It was notable that NGO participants described more barriers and challenges to partnerships compared to government participants. Government participants described more benefits.

Effective partnerships for evaluation were formed on the basis of shared clear goals, and agreed indicators, as well as respect and acknowledgement of the respective strengths each partner could bring. Early planning for the evaluation between partners, and the time spent on negotiating the goals of the evaluation so that benefits were shared, was an important foundation.

‘... it was about having that steering group together at the beginning and actually setting out, what is it that we need to evaluate, because everyone’s putting in different elements ...’ (P9, Manager, Government).

Participants described a range of partnerships, from informal support, to multi-year funded collaborations. The importance of building an ongoing working relationship over time between organizations was critical for effective evaluation, including when working with paid consultants or universities. Mutual benefit could be achieved in a partnership through the exchange of skills and evaluation expertise for access to data, or assistance to engage with the community.

‘... [this partnership] was awesome because they have an evaluation team, they have a marketing team, so the fact that we can partner on our [program], and then say to them, “look we’ll run it we can get you the participation.” ... But can you assist us in kind with writing a media release, with writing us up an abstract, with helping us evaluate it?’ (P26, Manager, NGO).

In many situations where skills and relationships were currency in the partnership, there was an expectation of a degree of capacity building. Where this expectation was unmet, tension and disappointment could arise.

The complexity of partnership arrangements or a power imbalance affected agreement on evaluation priorities or methods, and one participant described the pressure to reduce the rigour of methods when working with partners unskilled in evaluation. Several participants describing challenges of partnership also highlighted that despite the difficulties, it was worthwhile to conduct evaluation in partnership.

Partnering with universities and academic researchers was a common theme when discussing evaluation practice. Some practitioners identified that improvement was needed in the academic and practitioner partnerships to spend time to build relationships, ensure collaborative work with shared benefits and develop a mutual respect for each other’s skills and knowledge.

‘... they have their researchers and they have their money and they do their great work and we come in as fringe dwellers if they need input from us but it would be lovely if not only were we there to either provide the

name or the brand or to comment but we could even be integrated just into some of their training ...' (P8 Executive Officer, NGO).

Despite identifying academic researchers as skilled in evaluation methods, participants also felt some researchers had limited understanding of their program context and ways of working. Several participants discussed cost as a barrier to partnering with universities, and some rural participants described distance as an obstacle to accessing academic support, although this could be overcome by having access to effective networks and lists of skilled evaluators available in that region.

Resources

'Resources' or 'capacity' were general terms used by participants, or were discussed in terms of funding, staff, expertise and time. Resources for health promotion and evaluation were reported to be limited within the sector broadly, in addition to resource allocation being limited within the organization. In general, NGO participants discussed resource limitations much more than government participants. Health promotion managers that had control of the budget were able to allocate funding to evaluation as they saw appropriate, while others may have had control and were looking for further guidance about the necessary resource allocation.

Several participants described compromise as a product of limited resources. This caused evaluation to become piecemeal, lower quality or be completely cut in favour of program delivery and practitioners to undertake evaluation tasks beyond their skills and experience. A key theme that arose from the interviews was that challenges associated with limited resources appeared to be heightened by unrealistic expectations about what could be achieved with the available resources.

'...if people expected to do the same thing, it's pretty frustrating when money is so tight and they're working their little butts off and they're not meeting the indicators and there seems to be just more blame than sort of

understanding ...' (P13, Senior Practitioner, Government).

Lack of time was considered a barrier to effective evaluation practice. Participants felt that there was not enough time to conduct rigorous evaluation in their day to day roles, as well as over the life of a funding or program cycle. Particularly time consuming aspects of evaluation practice described were preparing evaluation plans, seeking ethics approval, analysing data to the degree desired, working in partnership for evaluation and using evaluation findings to improve programs. Some participants did not feel that resources were a limitation to evaluation, with one describing a well-funded evaluation that still produced disappointing reports and others satisfied with the level of funding available, despite time being a challenge.

Some participants, particularly from government agencies, identified strategies to minimize the impact of limited time by planning ahead, especially for the time consuming elements of evaluation as identified above.

'Actually saying, "...when we plan out this particular project, you know, we need to build in some time and dollars as well to allow for, you know, measurement and evaluation."' (P3, Team Leader, Government).

A number of participants reported access to templates and frameworks for evaluation planning, data collection and reporting that they considered helpful for evaluation practice. The guidelines provided by funders and other partners were often welcomed, but some participants expressed need for more detail and structure, and some sought tools relevant to non-health disciplines, especially when working in cross-sectoral partnerships. Participants reported that updated and validated data collection instruments were required to meet current needs and population groups, as well as consistency between regions and organizations to enable comparison. Only a small number of participants reported being in a position to adapt existing instruments appropriately.

Workforce support and development

The organization was recognized as having a key role in the development of practitioner beliefs and attitudes towards evaluation, as well as evaluation knowledge and technical skills. Participants reported that staff attitudes and beliefs about evaluation were influenced by colleagues and managers who provided leadership and acted as role models in the use of evidence for planning and decision making. These leaders sought ways to embed evaluation into everyday practice, demonstrating to practitioners the benefits of evaluation.

Additionally, negative attitudes towards evaluation were overcome by leadership and an organizational culture that supported learning and the sharing of findings. A common attitudinal barrier to evaluation was described as a fear of negative evaluation findings leading to loss of resources or major changes to programs. However, several participants mentioned that they had seen this attitude decline due to positive experiences of evaluation within their organization.

‘...the people who think like that for whatever reason haven’t been exposed to the power of evaluation and once you are and once you can see what it can do for you there’s no point in having those attitudes anymore’. (P25, Manager, Government).

Many participants described how the evaluation knowledge and skills of staff were influenced by organizational factors, particularly to engage with partner organizations, internal evaluation teams, or professional networks that provided access to expertise and mentoring. Additionally, effective leadership was described as an important facilitator of staff access to structured evaluation training, or technical support to learn on the job, particularly for staff who demonstrated strengths or interest in professional development opportunities for evaluation.

A small number of participants identified some managers and practitioners who saw program delivery as the only aspect of a practitioner’s role as a frustrating limitation to evaluation practice and the development of evaluation capacity.

‘... [Programs with centralized evaluation] turns health promotion officers who work on it into implementers rather than project designers and evaluators. So it kind of limits their skill development’. (P10, Director, Government).

In addition a few participants, and notably those at an earlier stage in their career, were exasperated by the wide range of competencies a practitioner was expected to hold within their organization and felt evaluation was better suited to specialists.

Recruitment and skills mix

Recruitment practices and clear organizational structures ensured those who valued evaluation, had a willingness to learn and had qualifications and skills in evaluation were available to support evaluation practice and use. Health promotion managers, particularly in NGOs, used job descriptions and recruitment processes to ensure the expectations concerning evaluation required were made clear to prospective employees. This was challenging to some participants where recruiting to roles in which health promotion program delivery was the primary purpose of the role.

‘When you’re looking at recruitment you’ve got to take into consideration that those evaluation skills maybe needed to be developed...they’d have different skills or value for delivering the program. So it would be difficult to find people to have the coverage of skills required’. (P24, Manager, NGO).

‘... Your professional background... really influences how you go about evaluating and the approach that you probably tend to take...’ (P31, Director, NGO).

Participants described the wide range of professional backgrounds in health promotion teams as a potential strength for the team that brings diversity in opinion and skills. Notwithstanding this, a number reported that a health promotion background was associated with an awareness of the importance and frameworks used in evaluation, a shared language around

evaluation and often a positive attitude towards evaluation compared to those from other backgrounds. These qualities were also important for people in positions of management.

'... [in health promotion degrees] there is a real culture of evaluation, ... they've obviously had that hammered into them and that's part of becoming a health promotion professional, so I think there is an expectation and a culture within people trained in the field that evaluation is important'. (P34, Manager, NGO).

Some participants described only having resources to employ graduates, and expressed concern that junior practitioners were not equipped to undertake evaluation tasks required in their roles. Staff who held a masters degree, including public health, were seen to have well developed knowledge and skills in problem solving, evidence-based practice and evaluation and research. In particular, a background in research or evaluation was valuable to understand and overcome challenges of evaluation and champion evaluation within the team.

Discussion

This is the first multi-jurisdictional study to describe the important facilitators and barriers to evaluation practice in the Australian prevention context. Our findings contribute needed explanations of how these factors interact within government and NGO prevention organizations and highlight the essential factors that need to be addressed to improve evaluation practice and use. The central role of the organization in determining evaluation capacity is highlighted in our study and supported in studies outside the Australian prevention sector where authors have applied organizational learning theory to frame evaluation capacity, emphasizing the role of leadership, organizational culture, systems and structures and communication [17, 18].

The cumulative influence of the facilitators to evaluation practice could possibly be explained by increased opportunities for practitioners to take part in evaluation, affecting both their individual

capacities and the wider culture of their organization. Patton [21] describes the organizational learning and program benefits that can occur from process use, defined as 'changes in attitude, thinking and behavior that result from participating in an evaluation. Process use includes individual learnings from evaluation involvement as well as effects on program functioning and organizational culture' [21, p. 99].

It is interesting to note that practitioner comments about influences on evaluation practice included frequent references to evaluation use, particularly the demand for evaluation use within organizational program management and planning systems, or the demonstrated use of evaluation findings by champions and leaders. In their conceptual framework for evaluation capacity building, Cousins *et al.* [22] highlighted the interdependent relationship between capacity to use and do evaluation, and emphasizing the importance of organizational capacity to use evaluation within evaluation capacity building frameworks. While our study does not focus explicitly on identifying factors influencing evaluation use in prevention organizations, the demand for increased use of evaluation findings appears to be an important driver of evaluation capacity building in the Australian prevention sector. This demand arises from the use of evaluation for program improvement [23], to enhance evidence-informed policy and decision making in government [1] and to demonstrate program effectiveness [24].

Our findings revealed leadership within organizations as a pivotal determinant of evaluation by influencing organizational culture, systems, structures and support for staff to gain evaluation skills and experience. This concurs with previous findings about the relationships between leadership and evaluation practice in government organizations through managerial skills, communication and mentoring [25] and the value leaders placed on evidence in human service organizations [26]. We found leadership and culture appear to play a greater role in determining how competing priorities for evaluation are managed within NGOs, which may partially be explained by resource constraints in some NGOs and the expectation of practitioners to undertake

multiple roles. This relationship is supported by a study that found voluntary sector managers were able to use evaluation findings to a greater extent due to their often dual responsibilities of program and evaluation management [18].

Organizational culture was another crucial factor in our study, as it impacted other organizational determinants of evaluation. Organizational culture ‘reflects the traditions, values and basic assumptions shared by its members and that establish its behaviour norms’ [25, p. 301] and our findings reveal a relationship between supportive organizational cultures and resource allocation to develop staff skills and confidence, and embedding evaluation in organizational systems. Additionally, the range of different cultures that facilitate evaluation practice in prevention organizations demonstrated the opportunity for influence on multiple fronts. For example organizations could address organizational values concerning evidence-informed practice [26], learning and improvement [23, 25], support of staff, or a specific focus on a culture of evaluation itself [6, 17].

Organizational systems and structures enable evaluation to be embedded within an organization in the longer term [17]. Participants described systems to ensure expectations around data and evaluation were clearly communicated between levels of management, and systems to access technical support and capacity building for evaluation. Organizational systems to facilitate evaluation practice, such as formal requirements for data collection and internal plans and reporting mechanisms have been highlighted as important elsewhere [6, 15, 23, 26].

Participants from government organizations gave greater attention to the role of organizational systems compared to NGOs. This emphasis may be explained by the different drivers of evaluation in each organization type [15], for example, government agencies are likely to be driven by accountability and reporting. Cousins *et al.* [18] discusses the use of tools such as ‘results-based accountability’, and the use of government systems and procedures to manage budgets and ‘exert control over agencies’ as important contributors to the focus on

systems for evaluation in government [18]. Our findings suggest that differences in priorities concerning evaluation between organizational types affect the role that systems play, with participants from NGOs reporting organizational systems related to governance and review of programs as facilitators of evaluation, whereas government participants more often emphasized systems for program recording and monitoring.

Role definitions, skills, qualifications and attitudes of practitioners were identified to influence efforts in evaluation. Dedicated evaluation functions appear to be an obvious facilitator of evaluation practice [17], however if responsibilities are unclear and work is not undertaken collaboratively with practitioners and managers, evaluation may not occur [6]. Finding the balance between supports for ‘doing’ evaluation and supporting the wider prevention team in program planning and evaluation appears to be crucial to meeting evaluation needs across the organization. This is a challenge when practitioners do not have dedicated evaluation time and are encouraged to prioritize implementation [13]. Our study found that a dedicated evaluation position can have an influential role in championing improvements to evaluation systems, culture and supporting practitioners.

Partnerships, although often requiring an investment of time and funding, were found to be beneficial for accessing and sharing resources. In other studies of evaluation practice, academic partnerships were highly valued for expertise, skill development [4, 6, 23, 27, 28] and ensuring quality evaluation [3], however as we also identified, these were often described as a wish not widespread practice. In line with the partnership literature [29], effective partnerships for evaluation were also based on respect, agreement of partner contributions and clear mutual benefits [6]. The relationship between partnerships and resource sharing was a strong theme, and others have noted that benefits that result from formal partnerships include sharing evaluation insights and findings, extending networks and informal mentoring and support [4, 11, 17, 23, 30].

Resource limitations have been frequently described as barriers to effective evaluation in the

health promotion and prevention sector [4, 13, 27, 28, 31]. This study confirms that budgetary constraints for prevention programs are common barriers to evaluation, and that a lack of time can be a major hindrance even if the budget is sufficient [6, 26]. Our findings further reveal how limited resources interact with other factors within organizations, in particular how leadership and organizational systems can determine resource allocation and how appropriate expectations can ensure evaluation activities can be conducted with available resources. Jolley *et al.* [7] described the South Australian Community Health experience of mismatch between demand for complex evaluation, and the availability of resources to implement these. The limited understanding of evaluation methodologies and inappropriate allocation of budget, time and expertise was also apparent in our study. Our findings further highlight that lack of resources and support from organizational decision makers, and pressure to focus on program delivery, are significant barriers even when the importance of evaluation is understood [6, 13, 27].

This study is unique in that it examines the perspectives of both government and NGOs in the health promotion and disease prevention sector. A limitation of this study is that participants were recruited from organizations with a minimum of three health promotion practitioners, and health promotion and prevention budgets of at least \$300 000 per annum. In the Australian prevention context, there are many programs being delivered by smaller organizations than this, and it is likely that they face additional barriers to evaluation. Additionally, not all Australian jurisdictions and sectors were represented in this study, and unique challenges may exist for specialist prevention organizations, such as those working with culturally and linguistically diverse communities.

We have comprehensively described how seven important factors interact within prevention organizations to influence evaluation practice. Given the demand for improved evaluation practice, the findings from this study provide prevention organizations with a useful framework for considering evaluation capacity building strategies. The important

interactions between each factor should not be underestimated, with major gains in organizational evaluation capacity building possible when considering multi-level effects of leadership, organizational culture and systems that support evaluation practice. While government and NGOs described some differences, we found more similarities overall in the mechanisms of influence between the organization types. Further research should explore the role of factors acting outside the organizations, as well as addressing gaps in objective measures of evaluation capacity in the prevention sector.

Acknowledgement

The authors would like to thank Emily McCluskey for her valuable contribution to participant recruitment and data analysis in this study.

Funding

This work was supported by the Australian Research Council [grant number DP150103575].

Conflict of interest statement

None declared.

References

1. Edwards B, Stickney B, Milat A *et al.* Building research and evaluation capacity in population health: the NSW Health approach. *Health Promot J Austr* 2016; **27**: 264–7.
2. Li V, Carter SM, Rychetnik L. Evidence valued and used by health promotion practitioners. *Health Educ Res* 2015; **30**: 193–205.
3. Pettman TL, Armstrong R, Doyle J *et al.* Strengthening evaluation to capture the breadth of public health practice: ideal vs. real. *J Public Health (Oxf)* 2012; **34**: 151–5.
4. Francis LJ, Smith BJ. Toward best practice in evaluation: a study of Australian. *Health Promotion Agencies. Health Promot Pract* 2015; **16**: 715–23.
5. Nutbeam D, Bauman A. *Evaluation in a Nutshell*. Northe Ryde: McGraw-Hill, 2006.
6. van Koperen TM, Renders CM, Spierings EJM *et al.* Recommendations and improvements for the evaluation of integrated community-wide interventions approaches. *J Obes* 2016; **2016**: 1.

7. Jolley GM, Lawless AP, Baum FE *et al.* Building an evidence base for community health: a review of the quality of program evaluations. *Aust Health Rev* 2007; **31**: 603–10.
8. Lim MSC, Wright CJC, Carrotte ER *et al.* Reach, engagement, and effectiveness: a systematic review of evaluation methodologies used in health promotion via social networking sites. *Health Promot J Austr* 2016; **27**: 187–97.
9. Smith BJ, Rissel C, Shilton T *et al.* Advancing evaluation practice in health promotion. *Health Promot J Austr* 2016; **27**: 184–6.
10. Lloyd B, Rychetnik L, Maxwell M *et al.* Building capacity for evidence-based practice in the health promotion workforce: evaluation of a train-the-trainer initiative in NSW. *Health Promot J Austr* 2009; **20**: 151–4.
11. Lobo R, Crawford G, Hallett J *et al.* A research and evaluation capacity building model in Western Australia. *Health Promot Int* 2016; 1–11. 10.1093/heapro/daw088 [EPub ahead of print].
12. Wolfenden L, Williams CM, Wiggers J *et al.* Improving the translation of health promotion interventions using effectiveness-implementation hybrid designs in program evaluations. *Health Promot J Austr* 2016; **27**: 204–7.
13. Lobo L, McManus A, Brown G *et al.* Evaluating peer-based youth programs: barriers and enablers. *Eval J Australia* 2010; **10**: 36–43.
14. Huckel Schneider C, Milat AJ, Moore G. Barriers and facilitators to evaluation of health policies and programs: policy-maker and researcher perspectives. *Eval Program Plann* 2016; **58**: 208–15.
15. Nielsen SB, Lemire S, Skov M. Measuring evaluation capacity—results and implications of a Danish study. *Am J Eval* 2011; **32**: 324–44.
16. Taylor-Ritzler T, Suarez-Balcazar Y, Garcia-Iriarte E *et al.* Understanding and measuring evaluation capacity. *A Model and Instrument Validation Study*. *Am J Eval* 2013; **34**: 190–206.
17. Preskill H, Boyle S. A multidisciplinary model of evaluation capacity building. *Am J Eval* 2008; **29**: 443–59.
18. Cousins JB, Goh SC, Elliott C *et al.* Government and voluntary sector differences in organizational capacity to do and use evaluation. *Eval Program Plann* 2014; **44**: 1–13.
19. Patton MQ. *Qualitative Research and Evaluation Methods*. Thousand Oaks, California: Sage Publications, 2002.
20. Miles MB, Huberman AM, Saldanña J. *Qualitative Data Analysis: A Methods Sourcebook*, 3rd edn. Thousand Oaks, California: SAGE Publications, 2014.
21. Patton MQ. Process use as a usefulness. *New Directions Eval* 2007; **2007**: 99–112.
22. Cousins JB, Goh SC, Elliott CJ *et al.* Framing the capacity to do and use evaluation. *New Directions Eval* 2014; **2014**: 7–23.
23. Lobo R, Petrich M, Burns SK. Supporting health promotion practitioners to undertake evaluation for program development. *BMC Public Health* 2014; **14**: 1315.
24. Naccarella L, Pirkis J, Kohn F *et al.* Building evaluation capacity: definitional and practical implications from an Australian case study. *Eval Program Plann* 2007; **30**: 231–6.
25. Bourgeois I, Cousins JB. Understanding dimensions of organizational evaluation capacity. *Am J Eval* 2013; **34**: 299–319.
26. Carman JG, Fredericks KA. Evaluation capacity and nonprofit organizations: is the glass half-empty or half-full? *Am J Eval* 2010; **31**: 84–104.
27. Brug J, Tak NI, Te Velde SJ. Evaluation of nationwide health promotion campaigns in The Netherlands: an exploration of practices, wishes and opportunities. *Health Promot Int* 2011; **26**: 244–54.
28. Napp D, Gibbs D, Jolly D *et al.* Evaluation barriers and facilitators among community-based HIV prevention programs. *AIDS Educ Prev* 2002; **14**: 38–48.
29. Baldwin L, Abernethy P, Roberts L *et al.* Forming, managing and sustaining alliances for health promotion. *Health Promot J Austr* 2005; **16**: 138–43.
30. Higgins JW, Strange K, Scarr J *et al.* It's a feel. That's what a lot of our evidence would consist of: public health practitioners' perspectives on evidence. *Eval Health Prof* 2011; **34**: 278–96.
31. Latham JR, Magiorakos AP, Monnet DL *et al.* The role and utilisation of public health evaluations in Europe: a case study of national hand hygiene campaigns. *BMC Public Health* 2014; **14**: 131.



The Funding, Administrative, and Policy Influences on the Evaluation of Primary Prevention Programs in Australia

J. Schwarzman¹ · A. Bauman² · B. J. Gabbe^{1,3} · C. Rissel² · T. Shilton⁴ · B. J. Smith^{1,2}

© Society for Prevention Research 2019

Abstract

Evaluation of primary prevention and health promotion programs contributes necessary information to the evidence base for prevention programs. There is increasing demand for high-quality evaluation of program impact and effectiveness for use in public health decision making. Despite the demand for evidence and known benefits, evaluation of prevention programs can be challenging and organizations face barriers to conducting rigorous evaluation. Evaluation capacity building efforts are gaining attention in the prevention field; however, there is limited knowledge about how components of the health promotion and primary prevention system (e.g., funding, administrative arrangements, and the policy environment) may facilitate or hinder this work. We sought to identify the important influences on evaluation practice within the Australian primary prevention and health promotion system. We conducted in-depth semi-structured interviews with experienced practitioners and managers ($n = 40$) from government and non-government organizations, and used thematic analysis to identify the main factors that impact on prevention program evaluation. Firstly, accountability and reporting requirements impacted on evaluation, especially if expectations were poorly aligned between the funding body and prevention organization. Secondly, the funding and political context was found to directly and indirectly affect the resources available and evaluation approach. Finally, it was found that participants made use of various strategies to modify the prevention system for more favorable conditions for evaluation. We highlight the opportunities to address barriers to evaluation in the prevention system, and argue that through targeted investment, there is potential for widespread gain through improved evaluation capacity.

Keywords Evaluation capacity · Primary prevention · Health promotion · Evidence-based public health

Background

The need to generate quality evidence through evaluation of primary prevention and health promotion (from here, referred to as prevention) programs has been gaining attention

(Jolley et al. 2007; Smith et al. 2016). Evaluation evidence is particularly relevant to practitioners if it demonstrates external validity, and use in program improvement and organizational planning (Francis and Smith 2015; Lobo et al. 2014; van Koperen et al. 2016). There is also the expectation that rigorous, practice-based evidence would be used to inform public health policy (Oxman et al. 2010) and guide the design and implementation of prevention strategies (Pettman et al. 2012). In particular, evidence of effectiveness and impact is needed (Oxman et al. 2010) to build knowledge about implementation success and contextual influences (Datta and Petticrew 2013), inform the scale-up of prevention initiatives (Milat et al. 2013), and enable ongoing accountability to funders (Bourgeois and Cousins 2013; Brug et al. 2011). Yet, the international prevention literature describes several challenges to program evaluation due to the complex nature and cross-sectoral approaches used in primary prevention (Nebot 2006), pressure to use traditional biomedical research methods (Petticrew 2013), and the ongoing requirement to justify investments in prevention programs (South and Tilford 2000).

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11121-019-00997-4>) contains supplementary material, which is available to authorized users.

✉ J. Schwarzman
Joanna.schwarzman@monash.edu

¹ School of Public Health and Preventive Medicine, Monash University, 553 St Kilda Road, Melbourne, VIC 3004, Australia

² Prevention Research Collaboration, School of Public Health, The University of Sydney, Sydney, NSW 2006, Australia

³ Farr Institute, Swansea University Medical School, Swansea University, Singleton Park, Swansea, Wales SA2 8PP, UK

⁴ National Heart Foundation Western Australia, 334 Rokeby Road, Subiaco, WA 6008, Australia

To further understand the gap in the availability of rigorous and relevant evaluation for policy makers and practitioners, evidence exploring the factors underlying evaluation practice in prevention organizations is emerging. This research predominantly identifies factors that affect evaluation practice at the individual level, such as knowledge, skills, and confidence to conduct and use evaluation, and at the organizational level, including organizational culture, leadership, support, resourcing, and internal systems for reporting and communication (Brug et al. 2011; Francis and Smith 2015; Huckel Schneider et al. 2016; Lobo et al. 2010; Napp et al. 2002; Schwarzman et al. 2018; van Koperen et al. 2016). Evaluation capacity building models and theoretical frameworks also focus on organizational- and individual-level capacity for evaluation (Labin et al. 2012; Nielsen et al. 2011), and these have become the targets of evaluation capacity building efforts (Norton et al. 2016). Yet, despite the attention given to organizational and individual level factors, there is recognition that the broader policy environment and funding systems can pose challenges to, or facilitate evaluation of prevention initiatives (Brug et al. 2011; Francis and Smith 2015; Jolley et al. 2007; Lobo et al. 2010; Napp et al. 2002; van Koperen et al. 2016).

To define the possible scope of factors that may influence evaluation of prevention programs beyond the individual and organization, we consider two conceptualizations of systems from the Australian health promotion context. Elements that are encompassed within the system include what has been described as the infrastructure for health promotion by Australia's National Health and Medical Research Council, that is: "the systems for policy development, monitoring and surveillance, research and evaluation, workforce development, and program delivery that direct and support action to promote, protect and maintain the health of the population" (Nutbeam 1998). More recently, a study analyzing the state policy context for health promotion in Australia adapted the World Health Organization's five building blocks of a health system for the health promotion system in Australia and included leadership and governance, financing, workforce, health promotion practice, and information (Baugh Littlejohns et al. 2018). In light of these definitions we conceive the prevention system to comprise primarily the funding, administrative, and policy functions that engage with the agencies and organizations delivering prevention activities.

In Australia, prevention activities are undertaken by local, state, and federal governments, as well as non-government organizations (NGOs) including community health services, issue-specific prevention organizations (e.g., AIDS Councils, Cancer Councils), and Aboriginal community controlled organizations. Models of regional coordination and support for prevention activities have been established, with functions that range from the commissioning of programs, regional

coordination, partnership development, and workforce support. Policy direction for prevention and health spending in Australia has fluctuated over the past decades (Wutzke et al. 2017). From 2008 to 2014 the National Partnership Agreement on Preventive Health Australia provided resourcing and direction for multiple initiatives across Australia (Wutzke et al. 2016). Within the state context, dramatic shifts in the support for prevention activities have also taken place over the past 15 years (Baugh Littlejohns et al. 2018).

The nature of prevention programs, workforce, organizations, and funding arrangements likely present additional challenges for evaluation when compared to health care or social services settings. Health promotion and primary prevention programs seek to improve health and well-being through addressing the social determinants of health and reducing risk factors associated with disease and injury. Multiple strategies are used concurrently to improve health in diverse populations, in a range of contexts. The approaches are varied, from policy development, to environmental change, community mobilization, to social marketing, and health education. Typically, prevention agencies in Australia work in partnership with communities and other health and non-health organizations, and are responsible for the planning, delivery, and evaluation of prevention initiatives.

The challenges to evaluation practice and use in the prevention field are paralleled by those described in the research translation literature. While Type 2 translation has typically focused on the translation of medical and clinical evidence into practice (Milat and Li 2017), the recognized importance of the contextual influences upon the adoption and sustained implementation of evidence-based interventions has led to attention on policy, funding mechanisms, and infrastructure in translation frameworks for prevention initiatives (Spoth et al. 2013). The Translation Science to Population Impact (TSci Impact) framework highlights the multiple contexts acting at national, state, and local levels, and the infrastructure supports, such as practice-oriented research, research-practice partnerships, and financing structures that interact with each phase of translation from evidence of efficacy or effectiveness into widespread evidence-based interventions (Spoth et al. 2013). The attention to policy, context, and funding levers that can influence the generation and use of research evidence reinforces the necessity to consider Australian prevention system's influence on evaluation practice.

Organizational theory can also identify the potential facilitators for evaluation in the prevention setting, especially as the nature of the relationship between funder and delivery organization affects evaluation practices. Carman (2011) mapped the evaluation motivations and actions of not-for-profit organizations against four organizational theories, highlighting the crucial issues of trust, power, and conflicting administrative rules in the relationship between funder and the

organization. The study identified the potential to adapt contract and accountability arrangements to enhance evaluation practice and encourage use of findings. The authors recommended a shift in the nature of funder-organization relationships (Carman 2011). In government agencies, the widespread adoption of the New Public Management framework, which focuses on performance management, monitoring, and accountability, has driven the uptake of outcomes focused evaluation (Dahler-Larsen 2009).

To date, there has been limited exploration of system-level influences on evaluation practice, particularly in the Australian context. In one qualitative study where prevention practitioners within community health services were interviewed, program timelines, funding constraints, and reporting to funding bodies were identified as influential on evaluation practice (Francis and Smith 2015). A study which explored the perspectives of policy makers and evaluators working within government health departments found that political context, policy, and projects with short time frames, limited funding, and administrative arrangements presented challenges to initiating evaluation (Huckel Schneider et al. 2016). A more recent study examining two Australian national prevention policies found that these could promote evaluation of prevention programs by establishing a national prevention agenda, and providing infrastructure for evaluation (Wutzke et al. 2016).

Despite the growing body of literature on organizational evaluation capacity, and recognition by prevention practitioners and leaders that funding arrangements, political, and system factors influence evaluation practice, there has been no comprehensive exploration of the prevention system's role in facilitating or hindering evaluation practice. Furthermore, the perspectives of different organization types engaged in delivering prevention strategies, and in multiple jurisdictions, have not been explored. In order to address these gaps, and to guide evaluation capacity building efforts, this study aimed to (a) identify the system-level factors that influence evaluation practice from the perspective of experienced prevention practitioners and managers working in government and NGOs and (b) describe how these factors act to influence evaluation practice.

Methods

This study used a qualitative semi-structured interview method. This study was one component of a larger mixed-method research project, in which 116 NGOs and government agencies had agreed to participate. Organizations were eligible if they had at least a 5-year history of conducting prevention strategies, and met the minimum size criteria (prevention budget > \$300,000 per annum, > 3 full time equivalent prevention staff) to enable the conduct evaluation and establishment of

supportive processes and systems. Organizations were recruited from New South Wales, Victoria, South Australia, and Western Australia given their history of conducting primary prevention work, and likelihood of having organizations with sufficient experience and size. For this qualitative component of the research, prevention practitioners and managers with a minimum of 5 years of experience were invited from within recruited organizations to participate ($n = 40$). Participants each represented a distinct agency or organization, and were purposively recruited to allow for saturation of themes across jurisdictions, government agencies, and NGOs.

The semi-structured interview guide consisted of ten open-ended questions and additional demographic questions (Electronic supplementary material). The guide was developed based on a review of the literature, researcher input, and a pilot study of government organizations that explored determinants of health promotion evaluation practice (Francis and Smith 2015). The interview guide was piloted with three practitioners not involved in the main study. First author (JS) conducted the audio-recorded telephone interviews between February and June 2016. Invited participants were provided with an information sheet before scheduling the interview, and verbal consent was obtained prior to commencement of recording. JS recorded potential biases prior to and during the interview process using a diary, and also recorded reactions and reflections after each interview. The interviews averaged 42-min duration. Monash University Human Research Ethics Committee (HREC), Aboriginal Health and Medical Research Council Ethics, and South Australian Department of Health and Aging HREC granted ethical approval for this study.

Data Analysis

Transcriptions of the audio recordings were provided to participants for comment and approval. Approved transcripts were uploaded into NVivo 11 software for coding. The interview guide and literature review informed the initial coding framework. The first round of coding was completed by two researchers (JS and EM) after dual coding 10 transcripts. The researchers met regularly to refine the framework and discuss discrepancies until consensus was achieved. A third researcher (BJS) reviewed the coding framework regularly to ensure that clarity of concepts before the framework was applied to the remaining transcripts. Thematic analysis was undertaken to identify barriers and facilitators to evaluation practice within the prevention context, and actions taken by organizations to support evaluation.

Results

All 40 participants who were invited agreed to participate. Participants were recruited almost equally from the four

participating states (Table 1), and government agencies ($n = 18$) and NGOs ($n = 22$). Almost half of the participants worked in prevention teams of 3 to 7 staff (full-time equivalent) ($n = 17$). The sample had considerable experience and were highly qualified, with 78% having worked 10 or more years in the prevention field and 80% holding a postgraduate qualification in health promotion or public health. Participant demographics are described in Table 1. Three main themes emerged; the first two describe important system-level influences upon evaluation practice, and the third explores organization's actions to modify system-level influences (Fig. 1).

Evaluation for Accountability: Aligning Expectations

The most widely discussed system-level influence on evaluation practice was discrepant expectations between funding bodies and funded organizations about what constituted an appropriate minimum evaluation, leading to tensions around

Table 1 Participant demographics

	<i>n</i>
State	
New South Wales	10
South Australia	9
Victoria	11
Western Australia	10
Organizational characteristics	
Type	
Government	18
Non-government organization	22
Prevention staff (full-time equivalent)	
3 to 7	17
8 to 15	9
More than 15	9
Not specified	5
Individual characteristics	
Prevention experience	
5 to 9 years	9
10 to 19 years	20
20 or more years	11
Role	
Manager/Director/CEO	29
Team leader	4
Senior practitioner	4
Other	3
Qualifications	
Postgraduate—health promotion/public health	18
Postgraduate—other	10
Undergraduate	12

reporting, resourcing, timelines, evaluation level (e.g., process, impact, or outcome), and indicators. Many participants expressed frustration that funders provided insufficient clarity about their expectations, whereas some who had stronger evaluation capacity appreciated the flexibility. While government participants experienced similar challenges to NGOs in reaching agreement with the central funding agency about evaluation standards, it was NGO participants who discussed at length the challenges in aligning with the expectations of funders for evaluation.

Participants referred to four aspects of their funding agreements when describing influences on evaluation practice, namely, reporting requirements, timelines, evaluation level and indicators, and funding levels. While each factor could act alone to facilitate or hinder evaluation practice, what emerged as most critical was the alignment of expectations between the funding agency and delivery organization about what level of evaluation could realistically be achieved with the resources available.

Many participants described the importance of evaluation to enable accountability to their funding body, donors, or in the case of government agencies, tax payers. While accountability was a key driver of evaluation, a small number of participants speculated that evaluations sent to funding bodies were unlikely to be used for any purpose other than meeting minimum accountability criteria, and were therefore a waste of time. Where organizations were required to meet only the minimum requirements, participants appreciated the demand for regular reporting against key indicators, or a template based evaluation, to facilitate data collection from the beginning of a program. Others expressed frustration at reporting against outputs or indicators that added no value for organizational learning, program improvements or could not demonstrate impact of their programs, and found this to distract from efforts to conduct evaluation more useful to the organization.

“I think for the funders it's kind of almost, like ... it's a little bit ticking the box...; we did 15 projects and they were all evaluated.” (Participant 1, NGO)

Even with reporting requirements that facilitated evaluative activities, many participants described the challenge of meeting the expectations of funding agencies for evaluation with the available resources, particularly in the case of unclear or changing expectations.

“... the requirements of the regional office does not align with what the guidelines say. I think if I sat down with people from both of those situations and say do you want us to do formative evaluation? Oh yes of course. Do you want us to do a process evaluation? Oh yes. Impact? Yes if you can that would be good too.” (Participant 6, NGO)

Fig. 1 Key themes concerning influences on program evaluation

<p>Theme 1. Evaluation for Accountability: Aligning Expectations.</p> <p>Organizations first seek to meet accountability requirements, even if they don't think it's the most useful evaluation. Additionally, expectations of the minimum evaluation expected can be clearly communicated and aligned between the two organizations, or they can be vague and misaligned. Expectations at this level are generally expressed through administrative requirements (e.g. reporting or templates), funding allocations or direct communication between funders and the organization.</p>
<p>Theme 2. Funding and Policy Context: Prioritizing Prevention and Evaluation.</p> <p>Evaluation is perceived as embedded in prevention practice by many participants. Valuing and understanding population level prevention programs at the political and funding levels is a pre-condition to evaluation practice. Tensions arise when the demand for evidence of effectiveness is expected, yet resourcing, and program and evaluation design does not allow for appropriate evaluation.</p>
<p>Theme 3. Modifying the System: Creating a Favourable Environment for Evaluation.</p> <p>Prevention organizations undertake a range of intentional activities to improve system conditions, encourage use and create demand for evaluation and prevention activities. The focus of organizational strategies were:</p> <p>A) Addressing evaluation requirements and resourcing, through direct negotiation with funding bodies, participation in forums and networks to provide feedback to policy makers, or proactively seeking additional funding for evaluation; and</p> <p>B) Approaches to strategically engage in policy and political decision making, such as advocacy for prevention approaches, investing in, and using rigorous evaluation evidence in communication with key officials, or strategically modifying program and evaluation language for the political climate.</p>

Several participants expressed frustration because they were expected to demonstrate health or social impacts over a 12-month, or shorter, program cycle. Others articulated their concerns that without appropriate guidance, indicators for impact evaluation would be irrelevant or difficult to measure. Participants that worked with vulnerable populations, in complicated or complex program areas, felt that the guidance, technical expertise, and resourcing they needed to design and conduct evaluation was often underestimated by the funding body. A smaller number of participants reported autonomy to determine the level, design, and indicators for their evaluation that was appropriate and commensurate with program time frames. This freedom was appreciated by organizations with access to expertise and resources, particularly those with multi-year funding agreements or ongoing funding.

Several participants discussed the expectation that around 10% of the program budget should be allocated to evaluation. The percentage was taken as a rule of thumb by some funders or program managers; two participants indicated that the 10% was used when receiving a "big grant," or as part of a special grant program for "evidence into practice," to "see if the program would work." Even when the 10% allocation was an expectation, there was lack of clarity about how the funding

should be used and frustration arose when the amount was not sufficient to meet the minimum requirements set by the funding body.

Funding and Policy Context: Prioritizing Prevention and Evaluation

Beyond the expectations for evaluation as set between the delivery organization and the funding body, participants described the influence of the political context and prevention funding models. Several participants emphasized the overarching role the political context played in setting the policy and funding environment for prevention activities. Participants associated favorable policy conditions for prevention activities, including adequate resourcing, with the likelihood of better evaluation. A small number of participants expressed satisfaction with supportive policy and funding conditions for prevention. The alternative experience, expressed by a number of participants, was of a political context that favored health care delivery, addressed individual risk factors, and used narrow lifestyle approaches to prevention. It was reported that policy makers, who lacked understanding of the necessary time or resources for every stage of prevention

program planning, delivery, and evaluation, often acted as an indirect barrier to evaluation.

Interviews took place 2 years after funding cuts to national- and state-level prevention initiatives. Examples cited by a number of participants were the cuts to the South Australian health promotion funding, and the disbandment of the National Partnership Agreement on Preventive Health after a change of Federal Government. Participants referred to the sudden reduction in funding in the wake of the National Partnership Agreement on Preventive Health announcements as the dissolution of the “value” of prevention and commitment to a national approach. With the political focus and resources shifting away from prevention, participants reported heightened pressure to focus remaining funds on program delivery. This was particularly evident in organizations which also delivered health services and those with limited evaluation technical expertise. At the same time, participants expressed frustration that the imperative to demonstrate program effectiveness was greater than ever. Several NGOs and some government participants felt particularly vulnerable as funding bodies and government departments implemented competitive tender models for program funding.

“... that constant need or imperative to defend and justify the work, either the work that’s being done, having to defend that work or the necessity or the need for that work. Also then trying to build up the case or the rationale for the work that needs to be done that keeps getting lost in the acute focus or the treatment focus that’s not looking upstream.” (Participant 27, Government)

Modifying the System: Creating a More Favorable Environment for Evaluation

Many participants described the policy, funding, and administrative influences on evaluation as modifiable, and outlined the strategies they and their organizations engaged in to improve reporting and evaluation requirements and resourcing. Additionally, participants engaged strategically with the prevention system at a policy and political level to promote evidence based decision making, and used evidence to increase support for prevention.

Participants recognized that in many instances unfavorable conditions for evaluation as set out in funding agreements or reporting templates could be changed. A small number of participants outlined their success in negotiating with their funder to better align expectations about evaluation with organizational need and capacity. This was particularly useful for NGOs who had multiple funding sources and sought to minimize reporting duplication. While some were not successful in negotiations, several participants recommended direct communication with the funder to ensure that expectations for

evaluation were achievable and the information generated satisfied the minimum requirements for all parties involved. Funding agreements with poorly conceived evaluation requirements tended to impact adversely on participants from smaller NGOs working with underserved populations, or addressing complex areas of health and social need. These participants were less likely to enter into negotiations about reporting or evaluation requirements, with factors such as resources, expertise, inflexible funding contracts, or the short-term nature of prevention funding appearing to affect the confidence of staff within their organizations to negotiate.

As an alternative to direct negotiation with funding bodies, both government and NGO participants identified prevention practitioner networks and inter-agency committees as an important avenue to influence evaluation standards and expectations. Participants were able to provide feedback to policy makers and funding bodies about the practical requirements for evaluation, and facilitate understanding of prevention program evaluation priorities.

“So we are often asked to sit on advisory committees, provide feedback to their frameworks and plans and stuff like that so input into the evaluation design.”
(Participant 24, NGO)

Several participants discussed how involvement in networks supported them to negotiate with funding bodies, or design their own evaluation when guidance from their funding body or organization was lacking. One research and evaluation network in the sexual health and blood-borne virus field offered members access to varying levels of evaluation support, and created opportunities to share evaluation practice insights and findings among peers. In a state government setting, participants described their ability to influence state-wide direction for prevention program evaluation and provide feedback to central decision makers through participation in a regular forum.

While accessing evaluation funding was beyond the reach of many organizations, a small number proactively sought grants to comprehensively evaluate their programs. Some well-established NGOs with diverse funding sources were in a strong position to undertake evaluation beyond the requirements of their funder, as well as strategically engage partners and access funding from external sources. When resources were available, participants described the need to invest in evaluation capacity building, generate evidence as a strategy to protect from future funding cuts, and enhance competitiveness for program funding.

“... you’re spending all this time to set up these good evaluation models, it’s worthwhile because you know you’re going to be able to do it long enough to make a difference. ... So having those good structures in place

helped to I guess sustain you when things do change.”
(Participant 16, Government)

Although a small number of participants were resigned to the ad hoc nature of prevention funding, and focused their efforts on improving administrative requirements and resourcing for evaluation, a number of participants sought to bring about system-level changes, including investment in longer program funding cycles more in line with population health and social change.

Participants described using evaluation to advocate for specific health issues and greater investment in prevention more generally. The use of evaluation for advocacy required publication of evaluation results, conducting economic evaluation and demonstrating the longer-term benefits of disease or injury prevention compared to treatment approaches.

“...having a fundamental underpinning that we can and have made a difference and we’ve got runs on the board, means we should walk into those conversations on obesity or on anything else, with some confidence to say, we actually do good things. We’ve got evidence that we do good things. We know how to do things well, so we deserve your support.” (Participant 15, NGO)

Achieving the level of rigor in evaluation necessary for advocacy was facilitated by strong partnerships between government, NGOs, and research organizations; the capacity of organizations to commission such work; and access to additional research or evaluation grants. Larger NGOs, in particular, were able to use evaluation evidence effectively to generate political support and resource allocation to prevention initiatives and promote the use of evidence in decision making.

Both government and NGO participants described the importance of communicating evidence of successful programs proactively to decision makers and elected officials. Participants reported that taking a strategic approach to the timing and style of evaluation was necessary. In some situations, participants indicated that a high-quality evaluation may not be as useful to engage politicians and decision makers as a lower quality report that demonstrated program effectiveness for popular strategies. To maximize funding opportunities and avoid program cuts participants considered the political climate and sometimes disguised aspects of programs to suit the dominant prevention paradigm. For example, some participants converted evaluation findings into metrics such as reductions in hospital usage and economic costs. One participant described changes to the language of their activities to align with the prevailing preferences regarding prevention methods.

If we say we’re a service, so we’re a health service, then we line up with the language and the jargon of clinical

services planning. They’re very happy when the system’s rhetoric links with partnering with community, or partnering with community organizations. Then suddenly, we’ll be allowed to say, oh look, that’s what we do. (Participant 33, Government)

A number of participants recognized that it was necessary to distil evaluation into short statements, focus on program effectiveness, and use quantitative evidence to facilitate use. Some government participants described the development of centralized data collection and monitoring systems in order to have information readily available to meet the demand from elected officials for evidence.

Discussion

To date, research concerning evaluation capacity in the health and human services sectors has focused largely on organizational- and practitioner-level factors. In this national study we have identified key aspects of the prevention system that have an impact on evaluation practice. We found that prevention evaluation can be successfully resourced and conducted when there is commitment to prevention programs, evaluation, and evidence use at the political and administrative levels.

Our findings indicate that emphasis on evidence-based (or evidence-informed) public health and research translation at the policy level can serve as an important driver of evaluation practice and capacity in the prevention system. In their framework for evidence informed policy and practice, Bowen and Zwi (2005) propose a range of necessary capacities to implement and adopt evidence-based policy, such as understanding of political arguments, decision-making processes, advocacy, economics of health policy and programs, and priority issues in government. In a qualitative study exploring the use of evidence for decision making in Australian public health insurance agencies, external elements such as political and ministerial factors, stakeholder influences, and media played a key role in shaping participants’ use of evidence (Zardo et al. 2014). Similarly, in research translation, there is recognition of the dynamic relationship between research and practice infrastructure, and multiple contexts impacting on each stage of translation from pre-adoption, to intervention sustainability and population health impact (Spath et al. 2013).

The common challenges identified in our study and literature from the research translation and evidence-based public health fields point to opportunities for widespread gains to evaluation practice and use. However, despite recognition of the value of evidence for health policy making gains, our study has revealed that prevention organizations and practitioners face additional challenges in promoting

the value and effectiveness of primary prevention within the broader policymaking environment. This in turn affects priorities for program administration, expectations, and resourcing for program evaluation. Frustration at a lack of political commitment and the tension between prevention and health service delivery has been described before. The pressure of constant comparison to health services has been described by prevention practitioners in the United Kingdom (South and Tilford 2000), and in Australia, practitioners and leaders expressed frustration at the narrow approach to prevention where environmental-level initiatives had greater potential for population health gains (Huckel Schneider et al. 2016). We found that this tension was recognized by organizations who used evaluation to demonstrate the value of prevention in economic and health service terms in order to remain competitive for limited health funding. For NGOs in particular, these actions demonstrate how organizations survive in a “resource-dependent” relationship where the driver for evaluation is primarily to access and maintain funding (Carman 2011).

Our study highlighted the detrimental impact of misaligned expectations for evaluation between funding bodies and delivery organizations, particularly in the context of limited time and resources. The programmatic and short-term nature of funding for prevention activities have also been described as detrimental to evaluating population-level prevention initiatives (Brug et al. 2011; Huckel Schneider et al. 2016; Napp et al. 2002), particularly the demonstration of effectiveness (Lobo et al. 2014), use of rigorous methods, and follow-up (Pettman et al. 2012). A recent review of prevention funding within Australia identified the particular challenges to delivering and reporting on strategies when receiving time-limited project funding compared to more flexible funding models (Jackson and Shiell 2017). In the case of a poor alignment between the funding model and program design, the authors described the efforts of funded organizations to “subvert” reporting requirements. Agency theory provides an explanation for these actions: the funder is to be served by the delivery organization, and in this relationship, conflict exists between the interests and preferences of each party, resulting in evaluation for the sole purpose of meeting contract requirements (Carman 2011). We also found that practitioners sought to disguise aspects of the program and evaluations for tactical purposes, particularly in a climate of prevention funding cuts, although our finding that larger NGOs and some government agencies appeared somewhat protected from the pressure to report only positive evaluation findings is an important new contribution. This may be explained by an organizations’ ability to access diverse funding sources, or realization of the potential for quality evaluation to lead to successful funding applications (Carman 2011; Francis and Smith 2015). This success may result from leadership for evaluation within the organization (Schwarzman et al. 2018).

We identified accountability as a key driver of evaluation in both government and NGOs; however, our findings appear to indicate that there is more at stake for NGOs and some sections of government that must constantly justify the effectiveness of prevention activities. Program evaluation is often driven by accountability in both government and NGOs (Bourgeois and Cousins 2013; Brug et al. 2011; Carman and Fredericks 2009) rather than program improvement (Lobo et al. 2014) or demonstration of impact (Pettman et al. 2012). Requirements for evaluation that focus on meeting accountability requirements through outputs, or indicators perceived to have little relevance to the program, have been found to hinder program evaluation (Carman and Fredericks 2009; Francis and Smith 2015; Lobo et al. 2010, 2014; Napp et al. 2002). Conversely, clear and well-structured reporting requirements can act as a facilitator to evaluation and embedding data collection systems within the organization (Carman and Fredericks 2009; Francis and Smith 2015). Encouraging organizations to firstly focus on using reporting data for program improvement and planning may promote the value of evaluation beyond an accountability requirement (Carman 2011).

Our study identified strategies used by organizations to modify evaluation requirements. Communication between funder and delivery organization is essential to establish an agreed evaluation purpose, design, and resourcing (Napp et al. 2002; Pettman et al. 2012). The theoretical model of a “stewardship” relationship assumes that the funder and organization have high levels of trust, based around “partnership” (as opposed to dependency), and that mutually agreed goals could generate evaluation that drives quality improvement and is shared widely (Carman 2011). We found wide variation in organizations’ attempts to align their own evaluation requirements with funding body expectations. Given the range of evaluation designs, time, and expertise required, funding bodies need to take into account the capacity of the prevention organization to contribute to evaluation, and identify ways to support or facilitate access to expertise. In particular, allocation of evaluation resources at the policy and program planning stages is necessary. In the Netherlands, participants argued that resources for evaluation could be readily allocated if they were requested at the planning stage (van Koperen et al. 2016). Lobo et al. (2014) also encouraged funders to engage in dialog with prevention organizations, and highlighted a framework for negotiating details of evaluation, particularly for program improvement (Liket et al. 2014).

In addition to facilitating direct negotiation between funding bodies and prevention organizations, our findings suggest that practitioner and organizational networks can encourage practitioners to engage in dialog with their funding bodies, and take up evaluation capacity building strategies. One such network is the Sexual Health and Blood-borne Virus Applied Research and Evaluation Network (SiREN) in

Western Australia, an initiative that aims to build evaluation capacity by facilitating collaboration between evaluators and practitioners, and supporting networks of practitioners, researchers, and policy makers (Lobo et al. 2016). Research-practice partnerships, networks, and funding models to incentivize collaborative relationships that embed the expectation of evaluation and practice relevant research have been a key focus of Type 2 translation models and recommendations (Spoth et al. 2013). Communities of Practice are also recognized as a strategy to promote translation of evidence-based interventions into practice. Communities of Practice facilitate learning through networks of practitioners based on common professional interests and practice needs. Improvements to practice are thought to occur through informal audit and feedback, relationship-based education, and opportunities for mentoring (Brehaut and Eva 2012). While our study participants found value in network contacts for evaluation, few structured examples were described. The establishment and maintenance of practitioner networks for evaluation requires investment in expertise, administration, and the support of practitioner's organizations to allow time to participate (Lobo et al. 2016). There is a role for leading prevention agencies and funding administrators to initiate the partnerships and assistance for organizations who have less confidence in evaluation (Lobo et al. 2010, 2016).

Beyond SiREN, there are few published examples of system-wide initiatives to improve evaluation capacity. The New South Wales Government has implemented a strategy to address evaluation capacity using a multi-level approach. The strategy uses a government-wide evaluation policy and guidelines, including assessment of evaluation capacity needs, communities of practice, and evaluation technical support (Edwards et al. 2016). To build evaluation capacity across the prevention system, there have been calls to explicitly allocate budget for evaluation in every funding agreement, initiate and strengthen research-practice partnerships, and ensure that evaluation methods and program objectives are achievable (Lobo et al. 2014; Pettman et al. 2012). Further, funding should focus on translation research and evaluation at the key stages addressing program replication and dissemination to ensure widespread implementation of effective programs (Rychetnik et al. 2012).

Implications

The evidence base for prevention remains limited, especially given the complexity of prevention programs and by nature, the lack of standardization between them. Hence, consistent approaches to evaluation, expectations for evaluation, and methods are warranted. Our findings highlight several system-level challenges to the evaluation of prevention programs that can be addressed through advocacy, investment, improved communication, and collaborative

efforts between practitioners, policy makers, funding bodies, and researchers.

Specifically, to address challenges that arise from administrative and reporting requirements (Fig. 1, Theme 1), we recommend *strategies that encourage funding bodies and organizations to engage in dialog*, in order to align evaluation requirements that enable accountability, but also contribute to program and organizational learning. To promote the value of prevention work, and encourage appropriate approaches to generating prevention evidence that can inform policy and funding (Fig. 1, Theme 2), we recommend *policy makers, prevention program, research, and evaluation experts come together* to better understand potential health promotion and social outcomes, and advocate for improved program and evaluation funding models. Further, there is a need to *support rigor and independence* of evaluation and disseminate findings widely. Finally, efforts to address the challenges of funding, administrative procedures, and policy influences must occur alongside sustained efforts to build evaluation capacity in prevention organizations and the workforce. Investment in infrastructure that directly supports prevention organizations and workforce to advance evaluation capacity is needed. Two examples of investment for support include *incentives for leading prevention organizations and university evaluators* to support those less capable with evaluation practice, facilitate partnerships, and assist in negotiations for evaluation requirements and resources, and *investment in inter-agency Communities of Practice* based on shared evaluation interests (e.g., obesity) or geography (particularly in regional areas) for all levels of practitioners and managers. We also recommend *researchers engage with practitioners, organizations, and funding bodies* to support quality prevention evaluation, and evaluate the efficacy and effectiveness of evaluation capacity building strategies.

Limitations

This study is unique in that it examines system-level issues across multiple states of Australia and from the perspectives of both government and NGOs in the prevention sector. A limitation of this study is that not all Australian jurisdictions and sectors were represented, and unique challenges may exist within prevention systems of other states or territories. Additionally, participants for this study were recruited from organizations with minimum staffing levels of three prevention practitioners, and budgets for prevention work of at least \$300,000 per annum. In the Australian prevention context there are many programs being delivered by organizations of smaller size than this, and it is likely that they may experience the prevention system differently to larger organizations. Furthermore, while we sought to understand the perspectives of prevention practitioners and managers, the experience of those in policy,

funding administration, and evaluation roles has not been explored, and may add depth to the understanding of the prevention system influence on evaluation.

Conclusion

This study contributes empirical evidence of the role of the primary prevention system in facilitating or hindering evaluation practice through funding and reporting arrangements, political context, and priorities, as well as opportunities for organizations to influence the prevention system for more favorable evaluation conditions. To improve the environment for prevention program evaluation, and ultimately improve prevention program effectiveness, the prevention field needs to look beyond each organization. Policy makers, funding bodies, and organizations seeking to address barriers to program evaluation should look for opportunities to collaborate to leverage wide-scale improvements to evaluation practice and use.

Acknowledgements We would like to thank Emily McCluskey for her valuable contribution to participant recruitment and data analysis in this study.

Funding This work was supported by the Australian Research Council (ARC) [DP150103575]. Prof. Gabbe was supported by ARC Future Fellowship [FT170100048] during the preparation of this manuscript.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from all individual participants included in the study.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Baugh Littlejohns, L., Baum, F., Lawless, A., & Freeman, T. (2018). Disappearing health system building blocks in the health promotion policy context in South Australia (2003–2013). *Critical Public Health*, 1–13. <https://doi.org/10.1080/09581596.2017.1418501>.
- Bourgeois, I., & Cousins, J. B. (2013). Understanding dimensions of organizational evaluation capacity. *American Journal of Evaluation*, 34, 299–319. <https://doi.org/10.1177/1098214013477235>.
- Bowen, S., & Zwi, A. B. (2005). Pathways to “evidence-informed” policy and practice: a framework for action. *PLoS Medicine*, 2, e166. <https://doi.org/10.1371/journal.pmed.0020166>.
- Brehaut, J. C., & Eva, K. W. (2012). Building theories of knowledge translation interventions: Use the entire menu of constructs. *Implementation Science*, 7, 114. <https://doi.org/10.1186/1748-5908-7-114>.
- Brug, J., Tak, N. I., & Te Velde, S. J. (2011). Evaluation of nationwide health promotion campaigns in the Netherlands: an exploration of practices, wishes and opportunities. *Health Promotion International*, 26, 244–254. <https://doi.org/10.1093/heapro/daq058>.
- Carman, J. G. (2011). Understanding evaluation in nonprofit organizations. *Public Performance & Management Review*, 34, 350–377. <https://doi.org/10.2753/PMR1530-9576340302>.
- Carman, J. G., & Fredericks, K. A. (2009). Evaluation capacity and nonprofit organizations: Is the glass half-empty or half-full? *American Journal of Evaluation*, 31, 84–104. <https://doi.org/10.1177/1098214009352361>.
- Dahler-Larsen. (2009). Evaluation and public management. In E. Ferlie, L. E. Lynn, & C. Pollitt (Eds.), *The Oxford handbook of public management*. Oxford: Oxford University Press.
- Datta, J., & Petticrew, M. (2013). Challenges to evaluating complex interventions: a content analysis of published papers. *BMC Public Health*, 13, 568. <https://doi.org/10.1186/1471-2458-13-568>.
- Edwards, B., Stickney, B., Milat, A., Campbell, D., & Thackway, S. (2016). Building research and evaluation capacity in population health: the NSW health approach. *Health Promotion Journal of Australia*, 27, 264–267. <https://doi.org/10.1071/HE16045>.
- Francis, L. J., & Smith, B. J. (2015). Toward best practice in evaluation: a study of Australian health promotion agencies. *Health Promotion Practice*, 16, 715–723. <https://doi.org/10.1177/1524839915572574>.
- Huckel Schneider, C., Milat, A. J., & Moore, G. (2016). Barriers and facilitators to evaluation of health policies and programs: Policymaker and researcher perspectives. *Evaluation and Program Planning*, 58, 208–215. <https://doi.org/10.1016/j.evalprogplan.2016.06.011>.
- Jackson, H., & Shiell, A. (2017). *Preventive health: How much does Australia spend and is it enough?* Canberra: Foundation for Alcohol Research and Education.
- Jolley, G. M., Lawless, A. P., Baum, F. E., Hurley, C. J., & Fry, D. (2007). Building an evidence base for community health: a review of the quality of program evaluations. *Australian Health Review*, 31, 603–610.
- Labin, S. N., Duffy, J. L., Meyers, D. C., Wandersman, A., & Lesesne, C. A. (2012). A research synthesis of the evaluation capacity building literature. *American Journal of Evaluation*, 33, 307–338. <https://doi.org/10.1177/1098214011434608>.
- Liket, K. C., Rey-Garcia, M., & Maas, K. E. H. (2014). Why aren't evaluations working and what to do about it. *American Journal of Evaluation*, 35, 171–188. <https://doi.org/10.1177/1098214013517736>.
- Lobo, R., McManus, A., Brown, G., Hildebrand, J., & Maycock, B. (2010). Evaluating peer-based youth programs: Barriers and enablers. *Evaluation Journal of Australasia*, 10, 36–43.
- Lobo, R., Petrich, M., & Burns, S. K. (2014). Supporting health promotion practitioners to undertake evaluation for program development. *BMC Public Health*, 14. <https://doi.org/10.1186/1471-2458-14-1315>.
- Lobo, R., Crawford, G., Hallett, J., Laing, S., Mak, D. B., Jancey, J., ..., Doherty, M. (2016). A research and evaluation capacity building model in Western Australia. *Health Promotion International* <https://doi.org/10.1093/heapro/daw088>.
- Milat, A., & Li, B. (2017). Narrative review of frameworks for translating research evidence into policy and practice. *Public Health Research & Practice*, 27, e2711704. <https://doi.org/10.17061/phrp2711704>.
- Milat, A. J., King, L., Bauman, A. E., & Redman, S. (2013). The concept of scalability: Increasing the scale and potential adoption of health promotion interventions into policy and practice. *Health Promotion International*, 28, 285–298.

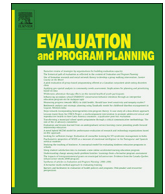
- Napp, D., Gibbs, D., Jolly, D., Westover, B., & Uhl, G. (2002). Evaluation barriers and facilitators among community-based HIV prevention programs. *AIDS Education and Prevention, 14*, 38–48.
- Nebot, M. (2006). Health promotion evaluation and the principle of prevention. *Journal of Epidemiology and Community Health, 60*, 5–6.
- Nielsen, S. B., Lemire, S., & Skov, M. (2011). Measuring evaluation capacity—Results and implications of a Danish study. *American Journal of Evaluation, 32*, 324–344. <https://doi.org/10.1177/1098214010396075>.
- Norton, S., Milat, A., Edwards, B., & Giffin, M. (2016). Narrative review of strategies by organizations for building evaluation capacity. *Evaluation and Program Planning, 58*, 1–19. <https://doi.org/10.1016/j.evalprogplan.2016.04.004>.
- Nutbeam, D. (1998). Promoting the health of Australians-how strong is our infrastructure support? *Australian and New Zealand Journal of Public Health, 22*, 301–302.
- Oxman, A. D., Bjørndal, A., Becerra-Posada, F., Gibson, M., Block, M. A. G., Haines, A., ... Wibulpolprasert, S. (2010). A framework for mandatory impact evaluation to ensure well informed public policy decisions. *The Lancet, 375*, 427–431. [https://doi.org/10.1016/S0140-6736\(09\)61251-4](https://doi.org/10.1016/S0140-6736(09)61251-4).
- Petticrew, M. (2013). Public health evaluation: Epistemological challenges to evidence production and use. *Evidence & Policy: A Journal of Research, Debate and Practice, 9*, 87–95. <https://doi.org/10.1332/174426413X663742>.
- Pettman, T. L., Armstrong, R., Doyle, J., Burford, B., Anderson, L. M., Hillgrove, T., ... Waters, E. (2012). Strengthening evaluation to capture the breadth of public health practice: Ideal vs. real. *Journal of Public Health (Oxford, England), 34*, 151–155. <https://doi.org/10.1093/pubmed/fds014>.
- Rychetnik, L., Bauman, A., Laws, R., King, L., Rissel, C., Nutbeam, D., ... Caterson, I. (2012). Translating research for evidence-based public health: Key concepts and future directions. *Journal of Epidemiology and Community Health, 66*, 1187–1192. <https://doi.org/10.1136/jech-2011-200038>.
- Schwarzman, J., Bauman, A., Gabbe, B., Rissel, C., Shilton, T., & Smith, B. J. (2018). Organizational determinants of evaluation practice in Australian prevention agencies. *Health Education Research*, Published online ahead of print, 1–13. <https://doi.org/10.1093/her/cyy015>.
- Smith, B. J., Rissel, C., Shilton, T., & Bauman, A. (2016). Advancing evaluation practice in health promotion. *Health Promotion Journal of Australia, 27*, 184–186. https://doi.org/10.1071/HEv27n3_ED2.
- South, J., & Tilford, S. (2000). Perceptions of research and evaluation in health promotion practice and influences on activity. *Health Education Research, 15*, 729–741. <https://doi.org/10.1093/her/15.6.729>.
- Spoth, R., Rohrbach, L., Greenberg, M., Leaf, P., Brown, C., Fagan, A., ... Hawkins, J. (2013). Addressing Core challenges for the next generation of type 2 translation research and systems: the translation science to population impact (TSci impact) framework. *Prevention Science, 14*, 319–351. <https://doi.org/10.1007/s11121-012-0362-6>.
- van Koperen, T. M., Renders, C. M., Spierings, E. J. M., Hendriks, A.-M., Westerman, M. J., Seidell, J. C., & Schuit, A. J. (2016). Recommendations and improvements for the evaluation of integrated community-wide interventions approaches. *Journal of Obesity, 2016*, 2385698–2385698. <https://doi.org/10.1155/2016/2385698>.
- Wutzke, S., Morrice, E., Benton, M., & Wilson, A. (2016). What will it take to improve prevention of chronic diseases in Australia? A case study of two national approaches. *Australian Health Review, 41*, 176–181. <https://doi.org/10.1071/AH16002>.
- Wutzke, S., Morrice, E., Benton, M., Milat, A., Russell, L., & Wilson, A. (2017). Australia's National Partnership Agreement on preventive health: Critical reflections from states and territories. *Health Promotion Journal of Australia. https://doi.org/10.1002/hpja.9*
- Zardo, P., Collie, A., & Livingstone, C. (2014). External factors affecting decision-making and use of evidence in an Australian public health policy environment. *Social Science and Medicine, 108*, 120–127. <https://doi.org/10.1016/j.socscimed.2014.02.046>.

6 The development and validation of the Evaluation Practice Analysis Survey

In Chapter Two, the need for a valid and reliable instrument to measure health promotion evaluation capacity was identified, given existing instruments had limitations in their assessment of psychometric properties, and had not been validated in health promotion organisations. In Chapter Five, a qualitative study that identified important practitioner-, organisational-, and system-level determinants of health promotion evaluation from the perspectives of 40 experienced practitioners and managers, was described. Practitioners identified leadership, organisational culture, systems and structures, partnership, resource allocation and workforce characteristics as influences upon evaluation within their organisations. Further, practitioners described how funding for health promotion and evaluation, political and administrative factors affected evaluation. The findings from Chapter Five, together with evaluation capacity and practice literature, formed the basis of the EPAS structure and items. To develop and assess the psychometric properties of the EPAS, a separate sample of health promotion practitioners were recruited from outside participating organisations. In this chapter, which comprises the published manuscript:

Schwarzman J, Bauman AE, Gabbe BJ, Rissel C, Shilton T, Smith BJ. Understanding the factors that influence health promotion evaluation: the development and validation of the Evaluation Practice Analysis Survey. *Evaluation and Program Planning*. 2019;74:76-83.

the stages of instrument development, refinement, and assessment of scale structure using PCA, are described. Further, the internal reliability of scales are assessed, and the predictive validity of selected resource, funding, leadership, culture and organisational system variables in relation to evaluation practice, is evaluated using logistic regression.



Understanding the factors that influence health promotion evaluation: The development and validation of the evaluation practice analysis survey

Joanna Schwarzman^{a,*}, Adrian Bauman^b, Belinda J. Gabbe^{a,c}, Chris Rissel^b, Trevor Shilton^d, Ben J. Smith^{a,b}

^a School of Public Health and Preventive Medicine, Monash University, 553 St Kilda Road, Melbourne, VIC, 3004, Australia

^b Prevention Research Collaboration, School of Public Health, The University of Sydney, Sydney, NSW 2006, Australia

^c Health Data Research UK, Swansea University Medical School, Swansea University, Singleton Park, Swansea, SA2 8PP, Wales, UK

^d National Heart Foundation of Australia, 334 Rokeby Road, Subiaco, WA 6008, Australia

ARTICLE INFO

Keywords:

Health promotion
Evaluation practice
Evaluation capacity building
Measurement
Reliability
Validity

ABSTRACT

The demand for improved quality of health promotion evaluation and greater capacity to undertake evaluation is growing, yet evidence of the challenges and facilitators to evaluation practice within the health promotion field is lacking. A limited number of evaluation capacity measurement instruments have been validated in government or non-government organisations (NGO), however there is no instrument designed for health promotion organisations. This study aimed to develop and validate an Evaluation Practice Analysis Survey (EPAS) to examine evaluation practices in health promotion organisations. Qualitative interviews, existing frameworks and instruments informed the survey development. Health promotion practitioners from government agencies and NGOs completed the survey ($n = 169$). Principal components analysis was used to determine scale structure and Cronbach's α used to estimate internal reliability. Logistic regression was conducted to assess predictive validity of selected EPAS scale. The final survey instrument included 25 scales (125 items). The EPAS demonstrated good internal reliability ($\alpha > 0.7$) for 23 scales. Dedicated resources and time for evaluation, leadership, organisational culture and internal support for evaluation showed promising predictive validity. The EPAS can be used to describe elements of evaluation capacity at the individual, organisational and system levels and to guide initiatives to improve evaluation practice in health promotion organisations.

1. Introduction

There is a misalignment between the demand for rigorous evaluation of health promotion programs, and the documented quality of evaluation practice in this field (Chambers, Murphy, & Kolbe, 2015; Lim, Wright, Carrotte, & Pedrana, 2016; Oxman et al., 2010; Pettman et al., 2012; Smith, Rissel, Shilton, & Bauman, 2016). Health promotion programs typically aim to address socio-ecological determinants of health using multiple and varied strategies; from environmental change, to social marketing or education. In many cases health promotion agencies will work with diverse populations, on a variety of health and social issues, in a range of contexts. In Australia, these agencies are commonly responsible for planning, delivery (frequently through partner organisations) and evaluation of programs. Consequently, program planning, implementation and evaluation are identified core competencies for health promotion practitioners (International Union for Health Promotion & Education, 2016).

Practitioners and organisations often recognise the importance of evaluation, and initiatives to build evaluation capacity in health promotion organisations have been reported (Edwards, Stickney, Milat, Campbell, & Thackway, 2016; Lindeman, Bettin, Beach, Adames, & Johnson, 2018; Lobo et al., 2016; Nichols et al., 2018; Valenti, Campetti, Schoenborn, Quinlan, & Dash, 2017). The interest in evaluation capacity within the field is encouraging, however there is a need for robust evidence to guide these efforts.

As the evaluation and ECB field progresses from theoretical frameworks to empirically supported models of ECB, commentators are calling for the development and validation of measurement instruments that can guide the planning of comprehensive ECB strategies and enable their evaluation (Labin, Duffy, Meyers, Wandersman, & Lesesne, 2012; Norton, Milat, Edwards, & Giffin, 2016). To date a limited number of survey instruments have been developed to assess evaluation capacity and practice within organisations. In Canada two instruments, based on a framework of capacity to undertake and use evaluation, were

* Corresponding author.

E-mail address: joanna.schwarzman@monash.edu (J. Schwarzman).

<https://doi.org/10.1016/j.evalprogplan.2019.03.002>

Received 31 August 2018; Received in revised form 22 January 2019; Accepted 18 March 2019

Available online 22 March 2019

0149-7189/© 2019 Elsevier Ltd. All rights reserved.

developed to assess evaluation capacity. The Evaluation Capacity in Organisations Questionnaire (ECOQ) was developed for use by internal evaluators within Canadian government and NGOs (Cousins et al., 2008). Nine scales measured capacity for organisational learning, support structures, capacity to do evaluation, ECB activities, use of findings and process use, and stakeholder participation. Eight of the scales demonstrated good internal reliability ($\alpha > 0.8$) (Cousins et al., 2008) and there was found to be good fit of the nine scales to the conceptual framework (Gagnon, Aubry, Cousins, Goh, & Elliott, 2018). Another measure, the Organisational Evaluation Capacity Self-Assessment Instrument, explored six domains of evaluation, including human and organisational resources, evaluation planning and activities, evaluation literacy, organisational decision making and learning benefits (Bourgeois, Simmons, Hotte, & Osseni, 2016). The instrument demonstrated good face validity as assessed by expert review and pilot testing, and was implemented with predominantly government agencies.

The Evaluation Capacity Index (ECI) was designed to map local government evaluation capacity in Denmark across evaluation demand factors (objectives, structure and processes), and supply factors (technology, human capacity) (Nielsen, Lemire, & Skov, 2011). Construct validity of the ECI was assessed using confirmatory factor analysis and found to support the theoretical model. Taylor-Ritzler, Suarez-Balcazar, Garcia-Iriarte, Henry, and Balcazar, (2013) developed the Evaluation Capacity Assessment Index (ECAI) which also demonstrated good internal reliability of all ten scales ($\alpha > 0.8$). The study found construct validity of the ECAI through assessment of the relationship between awareness, motivation, competence constructs and a second order latent variable of individual factors, and between leadership, learning climate and resources and a latent variable of organisational factors (Taylor-Ritzler et al., 2013). The measurement properties of additional evaluation practice instruments have not yet been assessed (Carman & Fredericks, 2009; Carman, 2007; Carman, 2008; Fierro & Christie, 2017).

In the health promotion field efforts to conduct and assess the impact of ECB strategies are emerging. In HIV prevention organisations a 7-item knowledge and confidence scale was adapted from existing instruments (Taylor-Ritzler et al., 2013) completed by participants of an ECB initiative ($n = 33$) (Lindeman et al., 2018). The authors reported a high internal consistency of the adapted scale in both the 'retrospective' ($\alpha = 0.88$) and the 'current' ($\alpha = 0.93$) application of the survey. Another case study used documentary review and qualitative interviews to report the impact of, and challenges to implementing an evaluation technical assistance and coaching initiative in substance abuse prevention organisations (Valenti et al., 2017). A further study, in an Aboriginal Community Controlled primary health care organisation, evaluated the effect of organisational systems changes, a workshop series and mentoring using pre-post workshops surveys to assess changes in skills and confidence among health promotion practitioners (Nichols et al., 2018). While these studies demonstrate motivation to conduct and evaluate initiatives to improve evaluation capacity in the health promotion field, they were conducted with small samples, and gave limited attention to the psychometric properties of survey instruments used. These studies also highlight the need for an instrument that can be used to describe and measure aspects of evaluation capacity beyond individual skills and knowledge.

Most studies of evaluation practices in the health promotion field have used qualitative approaches to explore barriers and facilitators to evaluation in homogenous organisation types (Francis & Smith, 2015; Jolley, Lawless, Baum, Hurley, & Fry, 2007; Napp, Gibbs, Jolly, Westover, & Uhl, 2002; van Koperen et al., 2016). We recently conducted a qualitative study that explored the perceptions of 40 managers, from a range of disease and injury fields, jurisdictions and organisation types, of the key factors affecting health promotion evaluation. This revealed the powerful influence of organisational factors (e.g. leadership, culture, resource allocation, support) upon evaluation practice and use, together with staff-level (e.g. skills, attitudes)

and system level (e.g. funding body requirements, political factors) determinants (Schwarzman, et al., 2018). Further studies in health promotion organisations identified human resources, time and the availability of evaluation tools or templates as necessary to evaluation (Francis & Smith, 2015; van Koperen et al., 2016).

In the wider field of research concerning evaluation in human services and other organisations, several studies and frameworks position the organisation and organisational learning as central to evaluation capacity (Bourgeois & Cousins, 2013; Cousins et al., 2008; Nielsen et al., 2011; Preskill & Boyle, 2008). Resources, leadership, organisational culture and systems or processes that support evaluation within the organisation have been consistently described as pivotal to evaluation practice, use and/or capacity in a range of organisational settings (Schwarzman, et al., 2018; Bourgeois & Cousins, 2013; Carman, 2007; Cousins et al., 2008; Fierro & Christie, 2017; Labin, 2014; Taylor-Ritzler et al., 2013). Internal and external support for evaluation has been reported to have a mediating effect on evaluation capacity building (ECB) (Labin, 2014), particularly support and expertise accessed through an internal evaluation team, external partnerships, or engagement with university or consultant evaluators (Schwarzman, et al., 2018).

The organisation has been shown to play an important role in developing staff level capacities such as evaluation knowledge, skills, attitudes and beliefs, through leadership and provision of a supportive environment (Schwarzman, et al., 2018; Labin et al., 2012; Labin, 2014; Taylor-Ritzler et al., 2013). These individual factors have been identified as important barriers to, and facilitators of, evaluation in health promotion organisations (Francis & Smith, 2015; Huckel Schneider, Milat, & Moore, 2016; Lobo, McManus, Brown, Hildebrand, & Maycock, 2010). The relationship between individual and organisational level evaluation capacity has been postulated in theoretical frameworks (Preskill & Boyle, 2008), and described in a range of models underpinned by empirical research (Labin et al., 2012; Labin, 2014; Taylor-Ritzler et al., 2013).

Evaluation capacity has also been portrayed as the relationship between evaluation supply and demand in an organisation (Nielsen et al., 2011). Several qualitative studies exploring barriers and facilitators to evaluation practice have identified health policy, funding and administrative requirements, and expectations that can act as powerful drivers or hindrances to evaluation in health promotion agencies (Francis & Smith, 2015; Lobo et al., 2010; Lobo, Petrich, & Burns, 2014; Napp et al., 2002).

The findings of research to date show that influences on health promotion evaluation practice act at the individual-, organisational- and wider prevention system- levels. Outside the health promotion field several attempts to measure evaluation practice and capacity have been made, however these instruments focus predominantly on a select number of organisational and individual factors. Given the distinctive administrative, organisational and practice characteristics of health promotion, there is a need for purposefully developed instruments for measuring evaluation capacity and practice in this field. Additionally, the validity and reliability of existing instruments have not been assessed in health promotion agencies, where evaluation is often the responsibility of program staff, as opposed to specialist evaluators surveyed in previous studies (Cousins et al., 2008; Nielsen et al., 2011). In light of these limitations, the purpose of this study was to develop and evaluate the psychometric properties of a survey instrument to measure the factors that influence health promotion and disease prevention evaluation practice.

2. Methods

2.1. Survey development

The Evaluation Practice Analysis Survey (EPAS) was developed in multiple stages to enhance face and content validity (Fig. 1). The

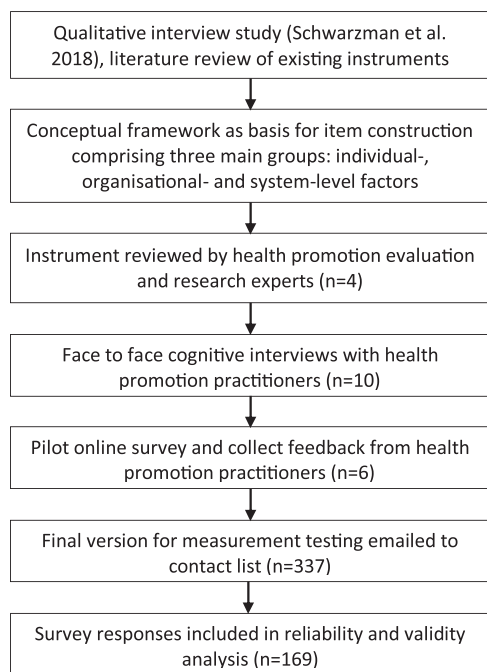


Fig. 1. Evaluation Practice Analysis Survey: Stages of development.

Individual level	Organisational level	System level
Skills for evaluation	Organisational culture	Funding for evaluation
Attitudes and beliefs	Leadership for evaluation	Reporting requirements
	Systems and structures	Funding body priorities and requirements
	Evaluation team or role	Political and prevention context
	Support for evaluation	
	Partnerships	
	External evaluators	
	University partners	
	Resources	

Fig. 2. Proposed framework of the factors that influence evaluation practice in health promotion.

conceptual framework that informed the content and structure of the survey was based on a synthesis of the literature presented above. Particular weight was given to evidence generated from within the health promotion field. Factors that influence evaluation practice in health promotion were grouped into three domains (Fig. 2). Survey items were created for each major domain (individual, organisation and system), and grouped into sub-headings. Items from existing evaluation practice and capacity instruments were reviewed for suitability to the health promotion context.

A panel of four experts in health promotion and disease prevention evaluation and research reviewed the items for face validity, clarity of wording, and provided feedback on suggested Likert scales. The revised instrument was used in face-to-face cognitive interviews with 10 health promotion practitioners in Victoria, Australia. Further revisions were made for clarity of definitions, item wording and response options. The survey was transferred into an online platform and sent to 10 eligible practitioners outside of Victoria who agreed to provide feedback after completion of the survey. Six participants provided feedback and minor revisions were made to instructions to participants and response options.

The preliminary survey included 158 items exploring factors influencing evaluation practice in 15 groups, with all but two sections being rated on 6 point Likert scales from strongly agree to strongly disagree (Table 1). Minor adaptations were made to 17 items from the ECAI ‘Awareness’ and ‘Competence’ scales (Taylor-Ritzler et al., 2013) and included in the preliminary survey. A “don’t know” response option was

Table 1
Items and response scales in preliminary Evaluation Practice Analysis Survey.

Initial proposed scales	N items	Response options
Self-reported evaluation practice	6	Continuous scale 0–100 ^a
<i>Individual level factors</i>		
Skills – evaluation steps	13 ^b	6 point beginner-expert scale
Skills – levels of evaluation	5	6 point beginner-expert scale
<i>Organisational level factors</i>		
Attitudes and beliefs	13 ^c	6 point agreement scale
Organisational culture	12	6 point agreement scale
Leadership	9	6 point agreement scale
Systems and structures	10	6 point agreement scale
Evaluation team or role [#]	7	6 point agreement scale
Support	9	6 point agreement scale
Partnerships	8	6 point agreement scale
External evaluators and University partners [#]	7	6 point agreement scale
Resources	20	6 point agreement scale [*]
<i>System level factors</i>		
Funding for evaluation	10	6 point agreement scale [*]
Reporting requirements [#]	10	6 point agreement scale [*]
Funding body priorities and requirements	11	6 point agreement scale [*]
Political and prevention context	14	6 point agreement scale [*]

^a Presented on sliding scale with increments of 5, with allowance for free entry of number.0–100.

^b 6 items from the Evaluation Capacity Assessment Instrument (Taylor-Ritzler et al., 2013).

^c 11 items from the Evaluation Capacity Assessment Instrument (Taylor-Ritzler et al., 2013).

* Included “don’t know” response option.

[#] Screening question to skip item set if not applicable e.g. Does your organisation have a designated evaluation team or role?.

added to 59 items based on feedback during expert review and cognitive interviews (Table 1). Respondent and organisational characteristics and evaluation practice measures were also collected.

2.2. Participant recruitment

Ethics approval was received from the Monash University Human Research Ethics Committee. Practitioner and health promotion organisation email addresses were compiled through internet searching, health promotion practitioner networks and the 2015 Australian Health Promotion Association (AHPA) Conference delegate list. The AHPA also forwarded an invitation to members in the Australian Capital Territory, Northern Territory, Queensland and Tasmania. In total, an invitation was sent by email to 337 practitioners and organisational administrative contacts. To encourage snowball recruitment, invited participants were encouraged to forward the email to eligible colleagues. This survey development study was undertaken within the context of a larger research project in which health promotion organisations had previously been recruited from New South Wales, South Australia, Victoria and Western Australia. The organisations recruited to the larger research project were considered ineligible to participate in this current study as they would be invited to complete the revised and validated EPAS at a later date. To avoid contamination of the participant groups, respondents to the initial invitation were screened based on their state and if necessary, their organisation. All participants were provided with an information sheet and commencement of the online survey was accepted as consent. Participants were offered an AUD\$ 20 gift voucher on completion of the survey.

2.3. Preliminary analysis

Survey data were imported into IBM SPSS 23 for analysis. All system level scales, and five items from the preliminary ‘Organisational

resources' scale had "don't know" options. When "don't know" was present for 25% or fewer items within a scale this was replaced by the mean of scores from the other items in the scale to retain cases for analysis (Tabachnick & Fidell, 2013). All other "don't know" and not applicable responses were managed as pairwise missing values in principal components analysis (PCA).

Spearman's correlation coefficients were generated for each preliminary scale to examine correlations between scale items and to guide removal of items with very low or very high correlations (Field, 2013). Where items demonstrated poor correlation ($R < 0.25$ for more than 50% of the scale items), these were deleted from the preliminary item pool. Preliminary scales were checked for multicollinearity, with a matrix determinant of < 0.000001 indicating multicollinearity (Field, 2013). Scales that were found not to meet the multicollinearity criteria were reviewed for highly correlating pairs and if required, an item was deleted prior to analysis.

2.4. Principal components analysis and scale internal reliability

We selected PCA as a technique that allowed us to understand the structure of each aspect of evaluation practice, extract the important components and eliminate redundant items using principal components analysis (PCA) (Field, 2013). This approach to assessing scale structure was used as the survey items were developed on the basis theoretical frameworks and empirical evidence that identify levels and dimensions of evaluation capacity in the health promotion field. As this survey measured a wide range of influences upon evaluation practice using a large number of items relative to the sample size, each preliminary construct was analysed using separate PCA. Prior to PCA it was determined whether each scale met the required assumptions; Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) (> 0.5), Bartlett's test of sphericity ($p < 0.001$), and absence of multicollinearity based on the determinant matrix. PCA was conducted with oblique (oblimin in SPSS) rotation as the components were assumed to be correlated (Field, 2013) and initial PCA set the minimum eigenvalue of 1.0 for component extraction. A minimum component loading of 0.5 was selected as this threshold is statistically significant in samples of over 100 participants (Stevens, 2002, in Field, 2013). A higher loading of 0.7 was adopted for the 'Evaluation role' scale which had a substantially lower sample size ($n = 49$) due to a prior screening question (Stevens, 2002, in Field, 2013). Extracted components with fewer than three items were not considered suitable for final scale output. Component extraction criteria for PCA were used to confirm the final scales, including Scree plots, eigenvalues, total variance accounted for, as well as appropriate theoretical fit of items within a scale (Field, 2013).

For each final scale, the internal reliability was determined using Cronbach's α . A threshold of $\alpha > 0.6$ was considered acceptable for this study, particularly as the number of items in a scale influence α and many final EPAS scales consisted of fewer than seven items (Streiner & Norman, 2008).

2.5. Predictive validity

Predictive validity of the EPAS was assessed using logistic regression. The dependent variable of self-reported evaluation practice was developed using survey items that asked participants to estimate the percentage of projects evaluated at formative, process, impact and outcome levels in the past two years. Principal components analysis was undertaken using these four items. All required assumptions for PCA were met and a single evaluation practice scale was extracted with item loadings > 0.7 and $\alpha = 0.79$.

The 'Self-reported evaluation practice' scale was then dichotomised at a threshold of at least 60% of programs evaluated (average of formative, process, impact and outcome evaluation). This cut point closely aligned with the highest tertile of scale distribution ($T2 = 61.2$, range 0–100) (Table 3). Sensitivity analysis was undertaken with cut points

for the evaluation practice scale set at higher and lower levels (50%, 70%), and this had negligible impact upon the predictive validity findings.

Based on prior qualitative studies (Schwarzman, et al., 2018), nine EPAS scales that were considered likely to influence evaluation practice were selected as independent variables for analysis. Each of the nine EPAS scales was categorised based on tertiles into high, medium or low EPAS scale scores, and was entered into a logistic regression model with self-reported evaluation practice as the outcome to establish the predictive capacity of the scales. The tertile distribution of each scale is shown in Table 3. Variables associated with both self-reported evaluation practice and each EPAS variable at the $p < 0.1$ level were considered potential confounders, and included selectively in the respective logistic regression models. Odds ratios (OR) and 95% confidence intervals (CI) were reported.

3. Results

3.1. Participant characteristics

The final sample included the responses of 169 participants, of which 159 were complete for all items. A response rate could not be calculated due to potential overlap of participants from each recruitment channel and the use of central organisational email contacts. Participant organisation and individual characteristics are shown in Table 2.

3.2. Scale structure and internal reliability

Sixteen items were deleted from ten preliminary scales prior to PCA based on the correlation analysis. A further 17 items with low loadings were excluded in the PCA process. Eight of the preliminary scales retained all items and were unchanged after PCA.

The PCA resulted in 25 EPAS scales, and retained 125 items

Table 2
Organisational, program and participant characteristics.

	Non-Government		Government		Total	
	n	(%)	n	(%)	n	(%)
Total	83	(49)	86	(51)	169	(100)
<i>Location</i>						
Metropolitan area	34	(41)	48	(56)	82	(48)
Non-metropolitan area	49	(59)	38	(44)	87	(51)
<i>Annual health promotion budget (Australian Dollars)</i>						
Less than \$499 999	49	(59)	42	(49)	91	(54)
\$500 000 or more	18	(22)	28	(33)	46	(27)
Don't know	16	(19)	16	(19)	32	(19)
<i>Staffing levels for health promotion (full time equivalent [FTE])</i>						
Less than 3 FTE	40	(48)	37	(43)	77	(46)
3 or more FTE	43	(52)	49	(57)	92	(54)
<i>Organisation's experience in health promotion (years)</i>						
10 years or less	27	(33)	15	(17)	42	(25)
More than 10 years	55	(66)	60	(70)	115	(68)
Don't know	1	(1)	11	(13)	12	(7)
<i>Participant characteristics</i>						
<i>Current role n = 159</i>						
Manager	33	(40)	15	(17)	48	(28)
Non-manager	45	(54)	66	(77)	111	(66)
<i>Experience in health promotion n = 169</i>						
Less than 10 years	45	(54)	47	(55)	92	(54)
10 or more years	38	(46)	39	(45)	77	(46)
<i>Qualifications n = 159</i>						
Undergraduate or post graduate public health or health promotion	48	(62)	57	(70)	105	(66)
Other qualifications	25	(32)	22	(27)	47	(30)
None	5	(6)	2	(2)	7	(4)

Table 3
Descriptive and internal reliability statistics for Evaluation Practice Analysis Survey scales.

Scale	Sample (n)	Items (n)	Scale range	Cronbach's α	Mean	(SD)	Ter1, Ter2
<i>Individual level</i>							
Attitudes and beliefs – program level	169	5	5–30	0.83	27.8	(2.5)	27, 30
Attitudes and beliefs – wider benefits	169	5	5–30	0.83	25.9	(3.3)	24, 28
Skills – levels of evaluation	169	5	5–30	0.90	18.2	(5.4)	16, 21
Skills – evaluation tasks	169	12	12–72	0.95	51.3	(12.4)	47, 57.3
<i>Organisational level</i>							
Leadership	168	9	9–54	0.96	36.5	(9.8)	34, 41
Organisational culture for the right environment	168	6	6–36	0.87	25.3	(5.5)	24, 27
Organisational culture for systems	168	6	6–36	0.91	24.5	(6.7)	21, 28
Systems and structures	167	7	7–42	0.85	26.2	(6.9)	23, 30
Internal support	164	4	4–24	0.88	15.4	(4.2)	14, 17
External support	164	3	3–18	0.77	11.4	(3.2)	10, 13
Tools	161	5	5–30	0.86	19.5	(4.8)	18, 22
Time	161	4	4–24	0.87	15.3	(3.8)	14, 18
Resources dedicated to evaluation	144	4	4–24	0.85	12.9	(4.8)	10, 15
Evaluation role	49	4	4–24	0.86	18.8	(3.7)	18, 21.3
Partnerships	162	6	6 – 36	0.86	27.3	(4.8)	26, 30
University evaluators	94	3	3 – 18	0.75	12.6	(2.7)	11, 14
External evaluators	95	3	3–18	0.73	14.4	(2.5)	14, 15
<i>System level</i>							
Adequate funding for evaluation	141	5	5–30	0.94	16.7	(6.5)	15, 20
Clear and realistic expectations	96	4	4–24	0.78	15.8	(4.0)	15, 17.7
Control and flexibility	128	3	3–18	0.66	11.5	(2.9)	10, 13
Competing demands	109	5	5–30	0.77	15.2	(4.7)	12, 17
Organisational influence	96	4	4–24	0.74	13.9	(4.0)	12.3, 15
Learning and sharing	112	4	4–24	0.76	14.5	(3.7)	13, 16
Political decision making	116	5	5–30	0.63	24.0	(3.2)	23, 25
Reporting facilitates evaluation	94	4	4–24	0.78	27.8	(5.2)	26, 30
<i>Evaluation practice</i>							
Self-reported evaluation practice	169	4	0–100	0.79	46.6	(27.7)	25, 61.2

SD = standard deviation, Ter1 = 1st tertile, Ter2 = 2nd tertile.

addressing individual (four scales), organisational (13 scales) and system level factors (eight scales) (Table 3). While the majority of scales were found to be unidimensional, five preliminary scales were found to be comprised of two dimensions ('Organisational culture', 'Support', 'Attitudes' and 'External evaluation and university partnerships') or three dimensions ('Resources'). The item list for each final scale is provided in Appendix A in Supplementary material. Most EPAS scales demonstrated good to excellent internal reliability, with $\alpha > 0.7$ for 23 of the 25 scales. The remaining two scales ('Control and flexibility' and 'Political decision making') showing acceptable internal reliability ($\alpha > 0.6$). The properties of the dependent variable 'Self-reported evaluation practice' are also shown in Table 3.

3.3. Predictive validity of key EPAS variables on evaluation practice

Chi-square analysis identified that annual health promotion budget and/or the respondent's role within their organisation were potential confounders of the relationship between four of the EPAS scales and the dependent evaluation practice variable. These potential confounders were included as covariates in logistic regression analysis for the four identified EPAS scales (Table 4). The remaining five EPAS scales were analysed using univariate logistic regression. The logistic regression results found seven of the nine of selected EPAS scales to be significantly associated with evaluation practice (Table 4).

High levels of resources for evaluation, including adequate funding, dedicated resources and time for evaluation were associated with between three and fivefold greater odds of evaluating of more than 60% of projects (Table 4). Reporting high levels of leadership, organisational culture to create a supportive environment, organisational culture that facilitates systems and internal support were also associated with at least two times greater odds of evaluating at least 60% of projects. Variables 'Clear and realistic expectations', and 'Systems and structures' were not found to be significantly associated with evaluation practice at the $p < 0.05$ level.

Table 4
Odds ratios for selected EPAS scales to predict program evaluation conducted for more than 60% of projects.

EPAS scale		OR (95% CI)	p value
Adequate funding for evaluation	Low	Ref	
	Medium	1.43 (0.52–3.90)	0.49
	High	5.19 (2.10–12.80)	0.00
Clear and realistic expectations	Low	Ref	
	Medium	1.80 (0.64–5.09)	0.27
	High	2.77 (0.98–7.85)	0.06
Resources dedicated to evaluation ^a	Low	Ref	
	Medium	4.73 (1.58–14.19)	0.01
	High	4.76 (1.62–14.00)	0.01
Time	Low	Ref	
	Medium	2.26 (0.98–5.22)	0.06
	High	3.14 (1.39–7.12)	0.01
Leadership	Low	Ref	
	Medium	2.12 (0.94–4.80)	0.07
	High	2.69 (1.19–6.10)	0.02
Organisational culture for the right environment	Low	Ref	
	Medium	0.81 (0.36–1.82)	0.61
	High	2.26 (1.03–4.93)	0.04
Organisational culture for systems ^b	Low	Ref	
	Medium	1.92 (0.80–4.60)	0.14
	High	3.75 (1.61–8.71)	0.00
Internal support ^b	Low	Ref	
	Medium	1.66 (0.71–3.89)	0.24
	High	2.89 (1.22–6.83)	0.02
Systems and structures ^{a,b}	Low	Ref	
	Medium	1.51 (0.55–4.13)	0.42
	High	1.57 (0.62–4.03)	0.34

OR = odds ratio, CI = confidence interval, Ref = Reference category.

^a Adjusted for *Annual health promotion budget*.

^b Adjusted for *Current role*.

4. Discussion

The availability of reliable and valid measures to describe evaluation practice is important for the initiation, design and evaluation of ECB strategies (Labin et al., 2012; Norton et al., 2016). This current study describes the development of the first instrument designed to measure evaluation practice in health promotion agencies. In particular, the EPAS uniquely contributes a wider range of reliable and valid scales that measure the individual, organisational and system-level dimensions of health promotion evaluation capacity than previous instruments. Principal components analysis was used to identify the dimensionality of the EPAS scales and reduce the number of survey items, resulting in an instrument that demonstrated good to excellent internal reliability for all but two of the 25 scales. Further, seven of the nine key EPAS scales showed promising predictive validity of evaluation practice in health promotion. This study is the first to our knowledge to assess the predictive validity of evaluation capacity measures upon reported evaluation practice.

To progress understanding of the influences on evaluation practice, this study builds on qualitative research exploring evaluation capacity in health promotion, theoretical frameworks of evaluation capacity and previous attempts at evaluation capacity measurement. We developed several scales that are consistent with instruments and frameworks published to date, especially within the individual and organisational domains. To our knowledge the ECAI is the only other instrument to assess the measurement properties of individual level factors influencing evaluation practice, including awareness, motivation and competence, and we adapted items from the ‘Awareness’ and ‘Competence’ scales for the EPAS (Taylor-Ritzler et al., 2013). In contrast with Taylor-Ritzler’s study, we found that the individual level factors of attitudes and beliefs about evaluation had distinct dimensions, each demonstrating high levels of internal reliability ($\alpha > 0.8$). The first dimension described perceived benefits at the program level, the second addressed perceived organisational and system wide benefits of evaluation.

Organisational level factors contributed the highest number of scales to the EPAS, and the majority of these had high levels of internal reliability (10 of 13 scales $\alpha > 0.8$). This was unsurprising given the range of existing instruments and theories that position the organisation as central to evaluation capacity. There are some similarities between the EPAS and the three organisational constructs measured in the ECAI, namely ‘Resources’, ‘Leadership’ and ‘Learning climate’ (Taylor-Ritzler et al., 2013). While we also developed scales for resources, PCA identified three resource sub-scales that delineated the elements of time, evaluation tools, and dedicated resources for evaluation within an organisation. Another distinction between the ECAI and the EPAS is the nature of the leadership constructs. The ECAI ‘Leadership’ scale (5 items, $\alpha = 0.82$) addresses the evaluation leadership characteristics of program managers only (Taylor-Ritzler et al., 2013), whereas the EPAS ‘Leadership’ scale encompasses more diverse sources of evaluation leadership. This interpretation of leadership reflects a breadth of examples of leadership for evaluation (e.g., champion evaluation, research and evaluation support role, facilitate partnerships) as identified in qualitative research in health promotion organisations (Schwarzman, et al., 2018).

The identification of suitable groupings of system level items for PCA was more challenging compared to organisational and individual factors. While most system level scales still demonstrated good internal reliability, the scales for ‘Control and flexibility’ and ‘Political decision making’ reached only acceptable levels of internal reliability ($\alpha > 0.6$). Beyond identification of funding body requirements as influential on health promotion evaluation (Francis & Smith, 2015; Lobo et al., 2010), evidence of how the health promotion system influences evaluation has not been extensively explored. Therefore, the constructs relating to system level influences on evaluation practice may evolve with further qualitative and quantitative exploration of the health promotion system.

In our analysis of the predictive validity of selected EPAS scales we found that ‘Dedicated resources’ and ‘Time’ for evaluation were positively associated with conducting program evaluation. Resources are frequently described as essential to evaluation practice, particularly for health promotion practitioners who have heavy workloads and are expected to be specialists across many areas of program management, including evaluation (Francis & Smith, 2015; Jolley et al., 2007; Lobo et al., 2010; Schwarzman, et al., 2018).

Leadership and organisational culture were also associated with evaluation practice in this study. Qualitative research has identified the pivotal role of leadership in creating a culture of learning, establishing and maintaining systems for communication and advocating for resource allocation to evaluation in health promotion (Schwarzman, et al., 2018). Two dimensions of organisational culture for which we developed reliable and valid scales, namely ‘Organisational culture to create the right environment’ and ‘Organisational culture for systems and structures’, have been found to be important in previous studies concerning evaluation capacity. An organisational culture that is supportive of learning, innovation and evaluation has been described as a key facilitator of ECB (Preskill & Boyle, 2008; Taylor-Ritzler et al., 2013), as has an organisational culture that develops systems and processes for shared learning and reporting to actively encourage evaluation (Bourgeois & Cousins, 2013). Embedding evaluation firmly within organisational culture and systems should also be an important goal of ECB, especially to promote use of evaluation (Cousins, Goh, Elliott, Aubry, & Gilbert, 2014).

To date ECB initiatives have tended to focus on training or technical support for the development of knowledge and skills, and the provision of resources in the form of evaluation guides and tools (Cousins et al., 2014; Norton et al., 2016). The EPAS shows promising utility for planning more comprehensive ECB initiatives in health promotion organisations that address skills, organisational resources, time and tools, together with organisational culture and support, and leadership for evaluation. Additionally, the EPAS may be useful in furthering research and evaluation of the impact of ECB strategies, which has been identified as an important next step in ECB research (Norton et al., 2016).

We also acknowledge limitations to this study. The sample size for this study may be considered low, given PCA is generally a larger sample technique (Tabachnick & Fidell, 2013). However in samples of less than 300 participants a combination of loadings, communalities, items, and factors should be considered when interpreting the components extracted (Field, 2013; Guadagnoli & Velicer, 1988; MacCallum, Widaman, Preacher, & Hong, 2001; Tabachnick & Fidell, 2013). In our study, the scale ‘Political decision making’ should be interpreted cautiously as it did not meet the additional criteria to justify a sample of less than 300 based on item communalities of 0.5 (Guadagnoli & Velicer, 1988) and at least four items with loadings > 0.6 (MacCallum, Widaman, Zhang, & Hong, 1999), despite meeting the KMO measure of sampling adequacy. Due to the sample size of the study several of the logistic regression models returned wide confidence intervals in the predictive validity analysis. We considered reducing categories within each independent variable to increase cell size, however decided that the tertile categories used more of the information available with only a modest loss of precision. Despite sample size limitations, we believe we were able to capture a representative sample of eligible health promotion practitioners through a comprehensive sampling strategy and follow-up emails and phone calls.

The use of self-report to measure evaluation practice for the predictive validity analysis may also be considered a limitation, however, given the lack of alternatives, we considered it an adequate proxy for this study and the scale to measure this dependent variable showed good internal reliability. Establishing valid and robust measures of evaluation practice would be beneficial for further validation of the EPAS. Additionally, while we explored the relationship of selected constructs with evaluation practice, there is also a hypothesised relationship between the EPAS scales that would benefit from further

testing to help understand how each factor interacts to influence evaluation practice.

5. Conclusions and lessons learned

The EPAS, based on theoretical and empirical foundations, is a comprehensive measurement tool examining individual, organisational and system level aspects of health promotion evaluation capacity. The EPAS demonstrated good internal reliability and validity to predict evaluation practice. It also shows promising utility as a tool for planning more ECB initiatives in health promotion that include skill building, development of organisational resources, tools, organisational culture and leadership development in evaluation. There is potential for practitioners and organisations to use the EPAS to deepen understanding of the influences on evaluation practice and ultimately develop and evaluate effective strategies to guide ECB in health promotion organisations.

Funding

This work was supported by the Australian Research Council (ARC) [DP150103575]. Prof. Gabbe was supported by an ARC Future Fellowship [FT170100048] during the preparation of this manuscript.

Declarations of interest

None.

Acknowledgement

Thank you to Emily McCluskey for assistance with development of the instrument and participant recruitment, and Elizabeth Vella for assistance with participant recruitment.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.evalprogplan.2019.03.002>.

References

- Bourgeois, I., & Cousins, J. B. (2013). Understanding dimensions of organizational evaluation capacity. *The American Journal of Evaluation*, 34(3), 299–319. <https://doi.org/10.1177/1098214013477235>.
- Bourgeois, I., Simmons, L., Hotte, N., & Osseni, R. (2016). Measuring evaluation capacity in Ontario public health units. *The Canadian Journal of Program Evaluation*, 31(2), 1.
- Carman, J. G. (2007). Evaluation practice among community-based organizations—Research into the reality. *The American Journal of Evaluation*, 28(1), 60–75. <https://doi.org/10.1177/1098214006296245>.
- Carman, J. G. (2008). Nonprofits, funders, and evaluation: Accountability in action. *The American Review of Public Administration*, 39(4), 374–390. <https://doi.org/10.1177/0275074008320190>.
- Carman, J. G., & Fredericks, K. A. (2009). Evaluation capacity and nonprofit organizations: Is the glass half-empty or half-full? *The American Journal of Evaluation*, 31(1), 84–104. <https://doi.org/10.1177/1098214009352361>.
- Chambers, A. H., Murphy, K., & Kolbe, A. (2015). Designs and methods used in published Australian health promotion evaluations 1992–2011. *Australian and New Zealand Journal of Public Health*, 39(3), 222–226. <https://doi.org/10.1111/1753-6405.12359>.
- Cousins, J. B., Elliott, C., Amo, C., Bourgeois, I., Chouinard, J., Goh, S. C., et al. (2008). Organizational capacity to do and use evaluation: Results of a pan-Canadian survey of evaluators. *The Canadian Journal of Program Evaluation*, 23(3), 1–35.
- Cousins, J. B., Goh, S. C., Elliott, C., Aubry, T., & Gilbert, N. (2014). Government and voluntary sector differences in organizational capacity to do and use evaluation. *Evaluation and Program Planning*, 44, 1–13.
- Edwards, B., Stickney, B., Milat, A., Campbell, D., & Thackway, S. (2016). Building research and evaluation capacity in population health: The NSW health approach. *Health Promotion Journal of Australia*, 27(3), 264–267. <https://doi.org/10.1071/HE16045>.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics: (and sex and drugs and rock 'n' roll)* (4th ed.). Los Angeles: Sage.
- Francis, L. J., & Smith, B. J. (2015). Toward best practice in evaluation: A study of Australian health promotion agencies. *Health Promotion Practice*, 16(5), 715–723.
- Fierro, L. A., & Christie, C. A. (2017). Evaluator and program manager perceptions of evaluation capacity and evaluation practice. *The American Journal of Evaluation*, 38(3), 376–392.
- Gagnon, F., Aubry, T., Cousins, J. B., Goh, S. C., & Elliott, C. (2018). Validation of the evaluation capacity in organizations questionnaire. *Evaluation and Program Planning*, 68, 166–175. <https://doi.org/10.1016/j.evalprogplan.2018.01.002>.
- Guadagnoli, E., & Velicer, W. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin*, 103(2), 265. <https://doi.org/10.1037/0033-2909.103.2.265>.
- Huckel Schneider, C., Milat, A. J., & Moore, G. (2016). Barriers and facilitators to evaluation of health policies and programs: Policymaker and researcher perspectives. *Evaluation and Program Planning*, 58, 208–215. <https://doi.org/10.1016/j.evalprogplan.2016.06.011>.
- International Union for Health Promotion and Education (2016). *Core competencies and professional standards for health promotion*. Accessed online 17th January, 2018 https://www.iuhpe.org/images/JC-Accreditation/Core_Competencies_Standards_linkE.pdf.
- Jolley, G. M., Lawless, A. P., Baum, F. E., Hurley, C. J., & Fry, D. (2007). Building an evidence base for community health: A review of the quality of program evaluations. *Australian Health Review*, 31(4), 603–610.
- Labin, S. N. (2014). Developing common measures in evaluation capacity building: An iterative science and practice process. *The American Journal of Evaluation*, 35(1), 107–115. <https://doi.org/10.1177/1098214013499965>.
- Labin, S. N., Duffy, J. L., Meyers, D. C., Wandersman, A., & Lesesne, C. A. (2012). A research synthesis of the evaluation capacity building literature. *The American Journal of Evaluation*, 33(3), 307–338. <https://doi.org/10.1177/1098214011434608>.
- Lim, M. S. C., Wright, C. J. C., Carrotte, E. R., & Pedrana, A. E. (2016). Reach, engagement, and effectiveness: A systematic review of evaluation methodologies used in health promotion via social networking sites. *Health Promotion Journal of Australia*, 27(3), 187–197. <https://doi.org/10.1071/HE16057>.
- Lindeman, P. T., Bettin, E., Beach, L. B., Adames, C. N., Johnson, A. K., et al. (2018). Evaluation capacity building—Results and reflections across two years of a multisite empowerment evaluation in an HIV prevention context. *Evaluation and Program Planning*, 71, 83–88. <https://doi.org/10.1016/j.evalprogplan.2018.09.001>.
- Lobo, R., Crawford, G., Hallett, J., Laing, S., Mak, D. B., Jancey, J., et al. (2016). A research and evaluation capacity building model in Western Australia. *Health Promotion International. Online Publication*, 1–11. <https://doi.org/10.1093/heapro/daw088>.
- Lobo, R., McManus, A., Brown, G., Hildebrand, J., & Maycock, B. (2010). Evaluating peer-based youth programs: Barriers and enablers. *Evaluation Journal of Australasia*, 10(1), 36–43.
- Lobo, R., Petrich, M., & Burns, S. K. (2014). Supporting health promotion practitioners to undertake evaluation for program development. *BMC Public Health*, 15(1315), 1–8. <https://doi.org/10.1186/1471-2458-14-1315>.
- MacCallum, R., Widaman, K., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods [PscyARTICLES]*, 4(1), 84. <https://doi.org/10.1037/1082-989X.4.1.84>.
- MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, 36(4), 611–637. https://doi.org/10.1207/S15327906MBR3604_06.
- Napp, D., Gibbs, D., Jolly, D., Westover, B., & Uhl, G. (2002). Evaluation barriers and facilitators among community-based HIV prevention programs. *AIDS Education and Prevention*, 14(3 Suppl A), 38–48.
- Nichols, N., McFarlane, K., Gibson, P., Millard, F., Packer, A., & McDonald, M. (2018). Skills, systems and supports: An Aboriginal Community Controlled Health Service (Apunipima) approach to building health promotion evaluation capacity of staff. *Health Promotion Journal of Australia*, 29(1), 18–22. <https://doi.org/10.1002/hpja.18>.
- Nielsen, S. B., Lemire, S., & Skov, M. (2011). Measuring evaluation capacity—Results and implications of a Danish study. *The American Journal of Evaluation*, 32(3), 324–344. <https://doi.org/10.1177/1098214010396075>.
- Norton, S., Milat, A., Edwards, B., & Giffin, M. (2016). Narrative review of strategies by organizations for building evaluation capacity. *Evaluation and Program Planning*, 58, 1–19. <https://doi.org/10.1016/j.evalprogplan.2016.04.004>.
- Oxman, A. D., Bjørndal, A., Becerra-Posada, F., Gibson, M., Block, M. A. G., Haines, A., et al. (2010). A framework for mandatory impact evaluation to ensure well informed public policy decisions. *The Lancet*, 375(9712), 427–431. [https://doi.org/10.1016/S0140-6736\(09\)61251-4](https://doi.org/10.1016/S0140-6736(09)61251-4).
- Pettman, T. L., Armstrong, R., Doyle, J., Burford, B., Anderson, L. M., Hillgrove, T., et al. (2012). Strengthening evaluation to capture the breadth of public health practice: Ideal vs. real. *Journal of Public Health (Oxford, England)*, 34(1), 151–155. <https://doi.org/10.1093/pubmed/fds014>.
- Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *The American Journal of Evaluation*, 29(4), 443–459. <https://doi.org/10.1177/1098214008324182>.
- Schwarzman, J., Bauman, A., Gabbe, B., Rissel, C., Shilton, T., & Smith, B. J. (2018). Organizational determinants of evaluation practice in Australian prevention agencies. *Health Education Research*, 33(3), 243–255. <https://doi.org/10.1093/her/cyy015>.
- Smith, B. J., Rissel, C., Shilton, T., & Bauman, A. (2016). Advancing evaluation practice in health promotion. *Health Promotion Journal of Australia*, 27(3), 184–186. https://doi.org/10.1071/HEv27n3_ED2.
- Streiner, D. L., & Norman, G. R. (2008). *Health measurement scales a practical guide to their development and use* (4th ed.). New York: Oxford University Press.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston: Pearson Education.
- Taylor-Ritzler, T., Suarez-Balcazar, Y., Garcia-Iriarte, E., Henry, D. B., & Balcazar, F. E.

- (2013). Understanding and measuring evaluation capacity: A model and instrument validation study. *The American Journal of Evaluation*, 34(2), 190–206. <https://doi.org/10.1177/1098214012471421>.
- Valenti, M., Campetti, R., Schoenborn, N., Quinlan, K., & Dash, K. (2017). Building evaluation capacity of local substance abuse prevention programs serving LGBTQ populations. *Evaluation and Program Planning*, 63, 101.
- van Koperen, T. M., Renders, C. M., Spierings, E. J. M., Hendriks, A.-M., Westerman, M. J., Seidell, J. C., et al. (2016). Recommendations and improvements for the evaluation of integrated community-wide interventions approaches. *Journal of Obesity*, 2016, 1–13. <https://doi.org/10.1155/2016/2385698>.
- Joanna Schwarzman** is a final year PhD candidate at the School of Public Health and Preventive Medicine at Monash University, and holds a MPH from Flinders University and Bachelor Physiotherapy from University of Melbourne. Joanna's career interests focus on the use of evidence in health promotion and disease prevention and building workforce and organisational capacity for evaluation in order to more effectively address population health challenges.
- Adrian Bauman** is Professor of Public Health and co-directs the WHO Collaborating Centre on Physical Activity, Nutrition and Obesity. He has a medical degree, MPH and PhD from Sydney University. His research is in chronic disease prevention and public health research translation and in the evaluation methods for complex public health programs. His expertise includes physical activity research, epidemiological studies, and the evaluation of population interventions and behaviour change programs. He is well published in research, and was on the 2015, 2016 and 2017 Thompson-Reuter (Clarivate) lists of the most cited 1% of researchers in any discipline.
- Professor Belinda Gabbe** is the Head of the Prehospital, Emergency and Trauma Research Unit in the Department of Epidemiology and Preventive Medicine at Monash University. She is an Australian Research Council Future Fellow, and holds a Professorial appointment in the Health Data Research UK Institute at Swansea University (UK). Prof Gabbe holds an Honours degree in Physiotherapy, a PhD in Epidemiology and a Master of Biostatistics.
- Chris Rissel** is Director of the NSW Office of Preventive Health and Professor of Public Health with the School of Public Health, The University of Sydney. He has a BSc (with Honours) in Psychology from UNSW, a Master of Public Health degree from Sydney University, and a PhD in Epidemiology (with a focus on community health promotion) from the University of Minnesota (USA). He works at the interface between research, policy and program implementation and is interested in the role of evidence in advocacy and how to best scale-up interventions for maximum health outcomes.
- Trevor Shilton** is National Director, Active Living with the National Heart Foundation of Australia. He is Vice President with the International Union for Health Promotion and Education (IUHPE) and has played a prominent role in developing national and global health promotion competencies and in advancing the profession of health promotion. He was the Founding National President of the Australian Health Promotion Association (AHPA). He has a Masters Degree in Health Promotion and a Bachelor of Education. His research, policy and practice interests are in chronic disease prevention, health promotion, advocacy and physical activity.
- Ben Smith** is an Associate Professor in the School of Public Health at the University of Sydney, and holds a BSW (Hons) degree (University of New South Wales), along with an MPH and PhD from the University of Sydney. Prior to his academic career he worked in community development and program management roles addressing health issues in high needs urban populations. His research interests concern chronic disease prevention, physical activity promotion, and capacity building for health promotion. During his career he has worked with a wide range of government and non-government agencies in the design and conduct of project evaluations.

Chapter Seven

7 How do the determinants of health promotion evaluation interact? Validation of a conceptual framework

Using the EPAS developed and refined in Chapter Six, the study presented in this chapter builds an understanding of how the determinants of health promotion evaluation capacity act to influence evaluation practice. This study sought to address a gap in the literature for validated frameworks of evaluation capacity, and contribute to health promotion ECB initiatives. Specifically, in this study, the construct validity of EPAS variables that measure practitioner-, organisational-, and system-level determinants of evaluation were assessed, and a conceptual framework of health promotion evaluation capacity tested. Participants for the study were recruited from within participating organisations, and 219 responses from a possible 230 were received. Using SEM, a good fit for a model of health promotion evaluation capacity was found. The model confirmed the pivotal role of organisational capacity factors, including leadership, culture, systems, support and resources, as determinants of evaluation practice. Further, the model demonstrated the relationship between the funding and political environment, reporting requirements and reported impact evaluation in the health promotion field. This chapter includes a manuscript under review by *Health Education Research*.

Schwarzman J, Bauman A, Gabbe B, Rissel C, Shilton T, Smith BJ. How practitioner, organisational and system-level factors act to influence health promotion evaluation practice: validation of a conceptual framework. *Health Education Research* (under review).

How practitioner, organisational and system-level factors act to influence health promotion evaluation capacity: validation of a conceptual framework

Joanna Schwarzman¹, Adrian Bauman², Belinda J. Gabbe^{1,3}, Chris Rissel², Trevor Shilton^{4,5} and Ben J. Smith^{1,2}

¹ School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia;

² Prevention Research Collaboration, Sydney School of Public Health, The University of Sydney, Sydney, Australia;

³ Health Data Research UK, Swansea University Medical School, Swansea University, Singleton Park, Swansea, Wales, United Kingdom;

⁴ School of Public Health, Curtin University, Bentley, Australia;

⁵ National Heart Foundation Western Australia, Subiaco, Australia.

7.1 Abstract

The need to improve the practice and quality of evaluation in the health promotion and disease prevention field is widely recognised. In order to plan, implement and evaluate health promotion evaluation capacity building efforts, there is a need to better understand the practitioner, organisational and system-level determinants of evaluation capacity and practice. This study aimed to assess the validity Evaluation Practice Analysis Survey (EPAS) constructs using confirmatory factor analysis and validate a conceptual framework of health promotion evaluation capacity using path analysis. Experienced Australian health promotion practitioners completed the survey (n=219). Twenty-one of the original 23 EPAS scales were assessed as reliable and valid. The final model was found to have good fit ($\chi^2_{14}=18.72$, $p=0.18$, root mean square error of approximation=0.04, 90% CI 0.00-0.82, Comparative Fit Index=1.00, standardised root mean square residual=0.04). This model supports the role of the organisation in facilitating evaluation practice through leadership, culture, systems, support and resources. It builds on existing frameworks from other fields to incorporate political, funding and administrative factors. This study provides an evidence-based model of evaluation capacity that organisations, funders and policy makers can use to plan and implement more effective evaluation capacity building strategies within organisations and the wider prevention field.

7.2 Introduction

Evaluating health promotion programs has many well-known benefits for practitioners, organisations and the public health field,[31,44,55,99] but there are many challenges to undertaking rigorous evaluation.[28,60,62,100] Motivation to improve evaluation capacity is gaining momentum[28] and to overcome the challenges to evaluation and guide evaluation capacity building (ECB) initiatives there is an increasing body of literature addressing the facilitators and barriers to evaluation[44,59,72,74] and emerging literature describing ECB initiatives in public health organisations.[101-103]

Qualitative studies in the health promotion and public health field from Australia,[44,60,104] United Kingdom,[51] USA,[71] and the Netherlands[59,75] have documented the barriers and facilitators of evaluation in a range of government agencies and non-government organisations (NGOs). Similarly, the determinants of evaluation practice and capacity have been explored in human service organisations and government agencies, highlighting some parallels with the health promotion field.[68,80,87,89] The wider evaluation field has used quantitative methods to measure and validate the determinants of evaluation capacity within organisations.[68] We identified three models of organisational evaluation capacity validated using statistical techniques in different settings: Canadian evaluators in government and NGOs,[82] Danish local government evaluators,[86] and USA non-profit organisations.[84]

Despite progress in organisational evaluation capacity models for the evaluation field and human services sector, direct application to the health promotion field is limited. The nature of health promotion programs, funding mechanisms and workforce pose unique challenges and opportunities for evaluation. Health promotion programs frequently address complex problems (e.g. obesity, alcohol use, mental health), that necessitate focus on the social determinants of health and behavioural risk factors. Effective health promotion requires concurrent use of multiple strategies that address individual-level through to environmental and policy determinants. Further, many strategies are delivered in partnership, addressing multiple determinants of health across diverse populations. Health promotion practitioners are commonly operating in small teams, and often within agencies primarily oriented to health service delivery. This has the potential to compel health promotion to constantly justify the approach to programs and evaluation within a medical model of health.[51,104] Evaluation is a core health promotion competency[45] and is often expected to be completed alongside program delivery responsibilities. Given the limitations of existing models of evaluation capacity to adequately capture the perspectives and range of determinants of evaluation applicable to the prevention field, we set out to develop and test a conceptual

framework of health promotion evaluation capacity.

7.2.1 The factors that influence evaluation in health promotion – a conceptual framework

The proposed framework is based on the current literature, interviews with 40 experienced health promotion practitioners,[104,105] and input from health promotion experts in the research team. In this section we describe the important determinants of evaluation acting at the system, organisational and practitioner level.

7.2.2 System level factors

While rarely reported in the broader evaluation capacity literature, political, policy, funding and administrative arrangements have been identified as important determinants of evaluation capacity in health promotion.[74,105] The push towards evidence-informed public health and policy,[15,52] where evaluation evidence makes a significant contribution,[16] may also be an important driver of evaluation. Overarching political priorities affect funding and administrative arrangements for health promotion programs and evaluation.[22,26,106] Program funding agreements can encourage organisations to prioritise evaluation activities solely for accountability requirements[76,91,105,107] and neglect evaluation for program improvement or to demonstrate longer term impact.[16,44,56]

Practitioners and managers who use and value evaluation may seek evaluation requirements from funding agencies that align with existing organisational capacity.[56,105] Practitioners can also work strategically with politicians and policy makers to influence the funding and policy context for prevention activities and evaluation.[52,105,108,109]

7.2.3 Organisational level factors

Most evaluation capacity literature situates organisational factors as central to evaluation capacity,[80,83] with ECB typically focused at the organisational level.[68,83,84]

Leadership, organisational culture, systems and processes, and evaluation resources are

consistently described features of evaluation capacity in government and NGOs.[76,80,84,87,89,104]

Promoting the value of evidence-informed planning, an environment that supports staff and encourages learning,[56,76] and mandating evaluation[75,83] can encourage organisational cultures supportive of evaluation. Leadership also plays a pivotal role to influence the organisational and administrative environment,[83,104,105,110] for evaluation. The effects of organisational culture and leadership include engagement with partners for resource sharing and support for staff undertaking evaluation;[16,44,56,59,71,75,83,104] fostering positive attitudes and beliefs among staff about the benefits of evaluation;[68,87,104] and establishing organisational systems and structures.[76,104]

The processes, systems and structures within an organisation are critical to all stages of evaluation, and developing these is often the goal of ECB.[81,87] Embedding evaluation into work plans, organisational systems, tools and processes to access support can facilitate evaluation practice, use,[56,75,83,86,91] and the allocation of resources to evaluation.[104]

A lack of time, funding and human resources have been frequently described as organisational barriers to health promotion evaluation.[44,59,60,71,72,104] In organisations that are under pressure to focus on program delivery activities, allocation of resources to evaluation can be particularly challenging,[60,72] even when funding is adequate.[75,91] Organisations can also determine opportunities for training, or mentoring.[104] For example, engaging with academic evaluators and seeking support from partner agencies and health promotion networks, can influence evaluation practice and skill development.[87,103,104,111]

7.2.4 Practitioner level factors

A lack of evaluation skills or knowledge, and negative attitudes or beliefs about evaluation have been frequently raised as barriers to health promotion evaluation.[44,72,74] Most ECB initiatives target practitioner characteristics through training or technical

support,[65,84,101] although the role of the organisation in shaping or mediating individual characteristics is well documented.[68,83,84,87]

Health promotion practitioners come from diverse professional backgrounds and those with health promotion qualifications often study evaluation in their degrees.[44,104] Despite varied baseline knowledge, practitioners who have beliefs about the benefits, and a positive attitude towards evaluation are more likely to seek out opportunities to improve their skills and understanding.[104]

7.2.5 This study

The interactions between the determinants of evaluation practice described above suggest both opportunity and potential complexity in addressing evaluation capacity in health promotion organisations. In order to effectively plan ECB initiatives in this field, a sound conceptual framework is needed that encompasses the practitioner, organisational and system-level influences upon evaluation. Logic models are frequently used in health promotion practice to inform program planning.[31] A validated conceptual model of the how the determinants of evaluation capacity act to influence evaluation practice could guide the planning of ECB initiatives in the field.

A valid and reliable instrument to determine priorities and measure impacts of ECB initiatives will also support efforts of the health promotion field.[65,68] We recently developed and assessed the psychometric properties of the Evaluation Practice Analysis Survey (EPAS), an instrument that measures aspects of health promotion evaluation capacity[110] however we did not assess the construct validity of the determinants of evaluation acting at the practitioner, organisational and system-level factors, nor did we assess the relationships between these.

In this present study, we chose to use structural equation modelling (SEM) to address both the need to develop an evidence-based framework of health promotion evaluation capacity, and to further assess the validity and reliability of the EPAS instrument. Structural equation

modelling is well suited to the assessment of measurement properties (i.e. validate scale structure), and the examination of explanatory models where there are a number of relationships between observed and latent variables.[112] The aims of this study are twofold, namely: to assess the construct validity of the EPAS scales, and second order latent variables that measure practitioner, organisational and system-level factors; and to test a conceptual model of the factors that influence evaluation capacity in health promotion, describing and quantifying the direct and mediating pathways of association.

7.3 Methods

7.3.1 Participants and procedures

Ethics approval was received from the Monash University Human Research Ethics Committee (HREC), Aboriginal Health and Medical Research Council Ethics Committee, Western Australia Country Health HREC and the South Australia Department of Health and Aging HREC (under the National Mutual Acceptance Scheme).

Data were collected between June and September 2017 from Australian health promotion practitioners using the EPAS.[110] This study was conducted within the context of a larger mixed-methods research project in which 116 government and non-government health promotion and prevention agencies were recruited from four states of Australia (New South Wales, South Australia, Victoria and Western Australia). Organisations participating in the study met minimum size and experience criteria at the time of recruitment (annual health promotion budget >AUD\$300 000, health promotion staff >3 full time equivalent, at least 5 years' experience delivering health promotion). The two most experienced practitioners within each participating organisation were invited by email to complete the online survey. All participants were provided with an information sheet, and commencement of the survey implied consent. Participants were offered an AUD\$20 gift voucher on completion of the survey.

7.3.2 Survey instrument

The EPAS is a survey instrument that measures evaluation capacity across practitioner, organisational and system-level factors within health promotion organisations. The instrument was developed using in-depth qualitative interviews and review of the literature, including existing evaluation capacity scales. A study assessing scale structure, validity and reliability of the instrument in a sample of 169 health promotion practitioners identified 25 unidimensional scales measuring determinants of evaluation capacity, and one scale measuring the extent of evaluation practice. Of the total 26 scales in the EPAS, 24 demonstrated good to excellent internal reliability ($\alpha > 0.7$). Several scales were found to have good predictive validity for evaluation practice.[110] The final EPAS used in this study comprised 131 items in 26 scales, with three to 12 items per scale. The full EPAS has been published elsewhere.[110]

7.3.3 Preliminary analysis

Survey data were imported into Stata 14.0 for analysis. Response options of “don’t know” and “not applicable” were treated as missing in the analysis. Three groups of questions were affected by a prior screening question, and therefore had considerably smaller numbers of responses (‘External evaluators’, ‘Internal evaluation role’, and ‘University partnerships’). Due to small sample sizes, these scales were considered unsuitable for confirmatory factor analysis (CFA). The distribution of each item was assessed, and Skewness > 3 and Kurtosis > 10 was determined to be indicative of non-normal distribution.[112] Preliminary scales were assessed for item-total correlations and internal reliability. Single items were removed from two scales based on low item-total correlations ($r < 0.025$) and low matrix correlations ($r < 0.025$ for $> 50\%$ of pairs). Scales were considered to have good internal reliability if Cronbach’s $\alpha > 0.7$.

7.3.4 Statistical analysis

Stage 1. Confirmatory factor analysis of survey scale variables

CFA was undertaken for each of the remaining 23 EPAS scales (117 items). Each scale was assessed for model identification and degrees of freedom (df), [113,114] ensuring the model was not under-identified ($df < 0$). Models were estimated using maximum likelihood estimation (MLE). Final factor loadings (standardised beta coefficients) were considered to be excellent when > 0.71 , very good when > 0.62 , good when > 0.55 and fair when > 0.45 . [115] The loadings were assessed for their suitability to construct an 'average' scale variable (i.e. equal loadings were preferred).

The goodness of fit criteria were selected based on Kline [113]: χ^2 , root mean square error of approximation (RMSEA) [90%CI], comparative fit index (CFI), and standardised root mean square residual (SRMR). Acceptable fit was considered to be χ^2 ($p > 0.05$); RMSEA < 0.08 (90% CI lower < 0.05 , upper < 1.00); CFI > 0.95 ; and SRMR < 0.08 . Local fit was assessed based on standardised residuals (> 2.58 considered poor fit) and modification indices for each model. Models that demonstrated less than acceptable fit overall and locally were modified based primarily on theoretical appropriateness and support of standard residual or modification indices. [113,114] Adding covariance between selected related error terms [113] was the primary modification made for all models.

Given the relatively new field of research, and limited assessment of scale psychometric properties to date, we considered it appropriate to undertake up to two rounds of modification if necessary. The decision to retain or reject each scale for the next round of latent variable construction was recorded based on the combination of overall goodness of fit, local fit, and reasonably equal factor loadings.

Stage 2. Practitioner, organisational and system-level latent variables

The second stage of analysis involved the construction of second order latent variables that measure major factors that influence evaluation capacity. Four separate models examining

seven latent variables were assessed using CFA. Three of the four models represented practitioner, organisational and system-level factors, the fourth model represented organisational leadership.

Due to the limited sample size, and complexity of the theoretical model, scale variables from stage 1 were 'parcelled' by calculating scale mean scores for use in stage 2 of analysis. Mean scores from participants that had >50% of data missing within a single scale were treated as missing for CFA of the respective model.

Stage 3. Path analysis

In the final stage, path analysis was conducted and the direct and indirect associations between the practitioner, organisational and system-level factors and evaluation practice were examined. The baseline model included the seven factors identified in stage 2 (factor scores), and a measure of evaluation practice. Pearson's correlations between each of the variables in the model were assessed.

We used model fit and modification criteria described in stage 1 and took an exploratory approach to analysis. In addition to a strong theoretical justification, we considered overall goodness of fit, significance (or not) of parameter pathways and modification indices to inform any changes to the model.[113] In order to validate the model fit, we repeated the final path analysis on the split halves of the sample[113,114] where the two participants from each organisation were assigned to separate groups. For the final selected model, we calculated the unstandardised and standardised direct, indirect and total effects.

7.4 Results

7.4.1 Sample and preliminary data analysis

In total, 219 (95%) of the 230 invited practitioners participated in the survey. Participant characteristics are described in Table 3. Of the 126 items in the EPAS scales, 93 items had 5% or less missing data; a further 23 items had 15% or less missing data. Full Information

Maximum Likelihood was used in the estimation of models where there was on average >5% missing data within a scale. The internal consistency of the preliminary scales was found to be high, with Cronbach's alpha between 0.74 – 0.95. See Appendix C-1 for scale sample size, distribution and internal consistency data.

Table 3. Survey respondent characteristics

	Non-Government		Government		Total	
	n	%	n	%	n	%
Total	135		84		219	
Organisation						
<i>State</i>						
NSW	27	20.0	33	39.3	60	27.4
SA	16	11.9	8	9.5	24	11.0
Vic	67	49.6	23	27.4	90	41.1
WA	25	18.5	20	23.8	45	20.5
<i>Location</i>						
Metropolitan area	112	83	50	59.5	162	74
Non-metropolitan area	23	17.1	34	40.4	57	26
<i>Annual health promotion budget</i>						
Less than \$500 000	41	30.3	24	28.6	65	29.6
\$500 000 or more	74	54.8	46	54.8	120	54.8
Don't know	20	14.8	14	16.7	34	15.5
<i>Staffing levels for health promotion</i>						
Less than 8 FTE	73	54.1	39	46.4	112	51.1
8 or more FTE	62	45.9	45	53.6	107	48.9
<i>Organisation's experience in health promotion</i>						
10 years or less	17	12.5	17	20.3	34	15.6
More than 10 years	116	85.9	66	78.6	182	83.1
Don't know	2	1.5	1	1.2	3	1.4
<i>Region covered in programs</i>						
LGA or smaller	18	13.3	25	29.8	43	19.6
Regional (multiple LGA)	45	33.3	36	42.9	81	37.0
State-wide	64	47.4	23	27.4	87	39.7
National	8	5.9			8	3.7
Participant						
<i>Current role</i>						
Senior Manager, Director	27	20.3	10	12.0	37	17.1
Manager, Team Leader	49	36.8	33	39.8	82	38.0
Non-manager	48	36.1	29	34.9	77	35.6
Other (including R&E role)	9	6.8	11	13.3	20	9.3
<i>Experience in health promotion n=216</i>						
Less than 10 years	55	41.4	25	30.1	80	37.0
10 or more years	78	58.6	58	69.9	136	63.0
<i>Qualifications n=216</i>						
Undergraduate or post graduate public health or health promotion	77	57.9	49	59	126	58.3
Other qualifications	48	36.1	34	41	82	38
None	8	6.0			8	3.7

\$ = Australian Dollars; FTE = full time equivalent; LGA = Local Government Area; R&E = Research and Evaluation.

7.4.2 EPAS scale variables

All 23 initial CFA models were estimated. Four scales fit the data without modification. Seventeen CFA models were modified and subsequently evaluated as having acceptable fit. These 21 scales were retained for use in the second factor analysis stage. The 'External support' scale was not retained because related constructs had already been excluded from analysis due to low sample sizes, and the initial CFA model was just-identified ($df=0$). The second scale excluded from analysis was 'Evaluation practice' as no acceptable fit was found and suggested modifications were not theoretically acceptable. Factor loadings identified the item 'Reported impact evaluation' to load at 0.89, whereas the remaining evaluation practice item loadings were between 0.38 and 0.63. While conducting evaluation at all levels is important, evidence of program effectiveness (e.g. impact evaluation) is frequently prioritised,[16,116] therefore, given the strength of factor loading and theoretical importance the item 'Reported impact evaluation' was retained as the dependent variable. The scale properties, modification details and fit statistics can be found in Appendix C-1. Scale structure and factor loadings for each first order scales variables are reported in Appendix C-2.

7.4.3 Second order evaluation capacity variables

We identified seven latent variables for inclusion in the model, and the dependent variable of 'Reported impact evaluation'. Table 4 shows the variable structure, sample size, factor loadings and results of CFA. Two items on the 'Organisational leadership' scale were found to load at 0.42 each. Despite these loadings the items were considered theoretically important to retain in the model. Table 5 shows the correlations between the eight variables included in the path analysis.

Table 4. Confirmatory factor analysis of practitioner, organisational and system-level latent evaluation capacity variables.

Latent variables	Scales	Factor loadings	N	DF	Chi2	P value	RMSEA (90%CI)	CFI	SRMR
Practitioner level									
Attitude and beliefs	Program level	0.75	219	1	0.05	0.83	0.00 (0.00-0.11)	1.00	0.00
	Organisational and wider	0.83							
Skills and knowledge	Evaluation tasks	0.91							
	Evaluation levels	0.93							
Organisational level									
Systems and support	Culture for systems and structures	0.81	217	8	10.42	0.24	0.04 (0.00-0.09)	1.00	0.02
	Internal support	0.86							
	Systems and structures	0.90							
	Evaluation tools	0.66							
Evaluation resources	Dedicated resources	0.92							
	Time	0.68							
Organisational Leadership	Leadership	0.84	216	2	4.99	0.08	0.08 (0.00-0.18)	0.99	0.03
	Culture for the right environment	0.89							
	Organisational influence	0.42							
	Partnerships	0.42							
System level									
Funding body expectations	Clear and realistic expectations	0.95	195	13	21.51	0.06	0.06 (0.00-0.10)	0.98	0.05
	Reporting requirements	0.76							
	Promote learning and sharing	0.59							
Political and funding environment	Adequate funding for evaluation	0.83							
	Competing demands	0.63							
	Political decision making	0.58							
	Control and flexibility	0.59							

SD=Standard deviation, df=degrees of freedom, RMSEA=root mean square error of approximation, CI=confidence interval, CFI=comparative fit index, SRMR=standardised root mean square residual.

Table 5. Correlations between latent evaluation capacity variables and reported impact evaluation

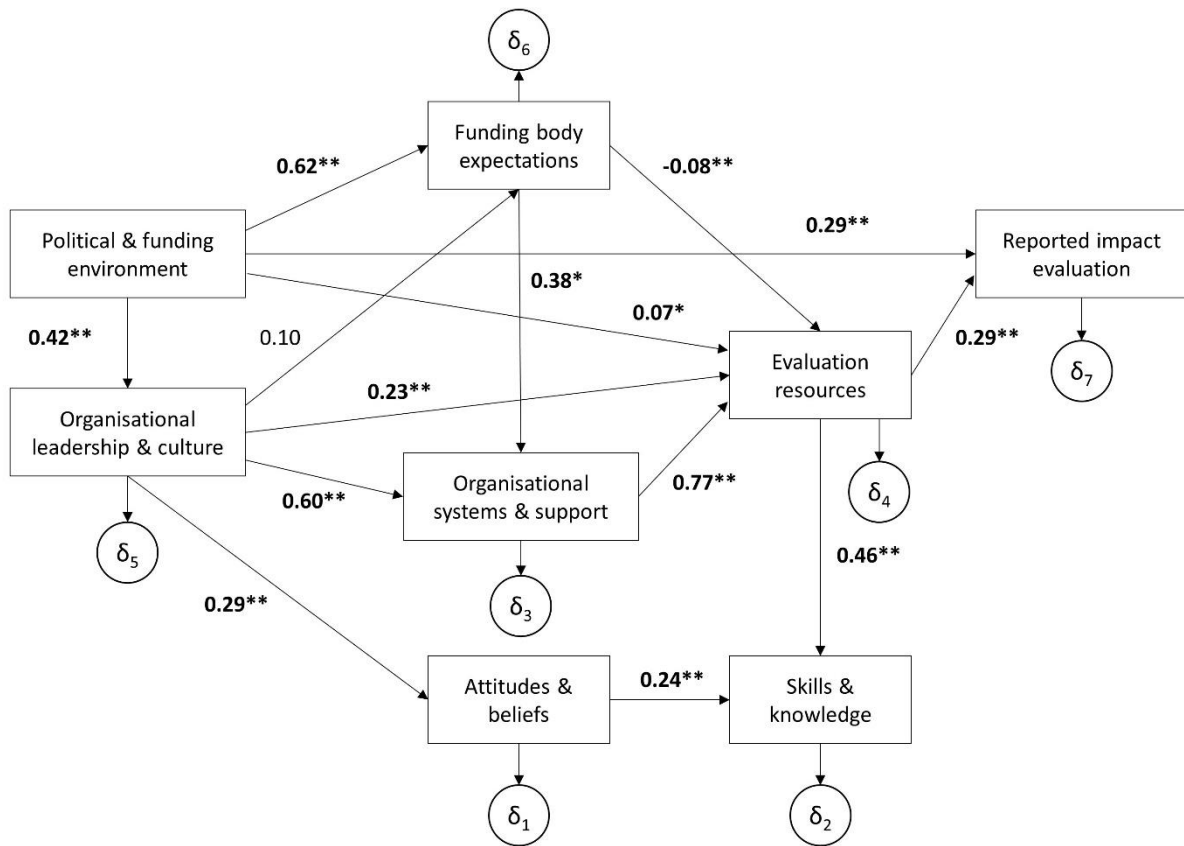
Latent variables	1	2	3	4	5	6	7	8
1. Reported impact evaluation	1							
2. Attitudes and beliefs	0.26	1						
3. Skills and knowledge	0.24	0.35	1					
4. Evaluation resources	0.43	0.40	0.43	1				
5. Organisational systems and support	0.42	0.41	0.40	0.93	1			
6. Organisational leadership	0.39	0.37	0.30	0.80	0.74	1		
7. Political and funding environment	0.43	0.33	0.20	0.49	0.49	0.42	1	
8. Funding body expectations	0.31	0.25	0.21	0.51	0.60	0.37	0.67	1
	1	2	3	4	5	6	7	8

NB. All correlations significant at $p < 0.01$

7.4.4 Path analysis for evaluation capacity and practice

The baseline model of the evaluation capacity was estimated using MLE ($n=216$), and this did not fit the data well. Model modifications were made in two stages. Full details of baseline model fit and modifications are available in Appendix C-3. The final model was found to have good fit on all criteria: $\chi^2_{14}=18.72$, $p=0.18$; RMSEA=0.04, (90% CI 0.00-0.82); CFI=1.00; and SRMR=0.04. Split-sample confirmation of the model for both Group A ($n=107$) and Group B ($n=109$) demonstrate good model fit on all criteria. The standardised direct effects are shown on the final model in Figure 3, and Appendix C-4 shows the direct, indirect and total effects.

Figure 3. Path model relationship between components of evaluation capacity and reported impact evaluation



Solid lines = Standardised direct effects, **p<0.01, *p<0.05; δ error of endogenous variables.

7.5 Discussion

There is a strong imperative to improve evaluation capacity in the health promotion field.[28] Rigorous evaluation contributes essential evidence for prevention strategies and informs policymaking. ECB initiatives are being implemented[101,103,117,118] however, there is limited evidence concerning how these should be designed and delivered for the health promotion field. In addition to confirming the reliability and validity of the EPAS instrument, this study identified second order factors that are relevant to capacity and revealed pathways between these and evaluation practice.

Previous investigation of evaluation capacity models, from outside the health promotion field, have focused heavily on organisational variables[82,86] and their interactions with practitioner capacity.[84] One study found that organisational factors, comprising

leadership, learning climate and resources, fully mediated the relationship of staff motivation, awareness and competence and outcomes of evaluation mainstreaming and use.[84] The present study also did not find a significant direct pathway between practitioner factors and reported impact evaluation, but found that organisational leadership, culture and resources appeared to influence attitudes, beliefs, skills and knowledge. The lack of a direct relationship between practitioner level factors and evaluation outcomes in these two models highlights the powerful role of the organisation in facilitating or hindering evaluation regardless, it seems, of staff motivation or competency. While we do not dismiss the need for training or mentoring, as has been common in ECB initiatives to date,[68] this study provides evidence that ECB initiatives should prioritise strategies to build capacity at an organisational level such as leadership to facilitate evaluation, developing a culture of evaluation and learning, embedding evaluation in organisational systems and providing support for evaluation.

The final explanatory model developed in this study showed that leadership, culture, systems, support and resources interact closely with each other to inform evaluation practice (Figure 3). Previous research concerning the influence of organisational factors upon evaluation capacity[86] has found relationships between evaluation demand (e.g. structures, processes, evaluation objectives) and supply (e.g. technical resources, human capital) within organisations. The role of structures within organisations that drive evaluation practice, and in our model, organisational systems that facilitate resource allocation for evaluation, highlight the importance of demand-side influences upon of evaluation, both within organisations and in the wider prevention system. In our study, funding body requirements appeared to contribute to the demand for evaluation within the organisation, but had a weak relationship to the resources allocated to this. The combination of limited funding and the requirement to undertake evaluation has been previously documented in health promotion organisations.[44,104] Yet accountability obligations remain a strong driver of evaluation as organisations often conduct evaluation to “satisfy funders” and less frequently for program

quality improvement, or organisational decision making.[107]

Our study found a direct relationship between the political and funding environment, and reported impact evaluation. Evaluation is often motivated by political and funding factors, such as the need to demonstrate program effectiveness,[103] seek or maintain funding[107] and underpin advocacy.[105] The role of the political environment has often been overlooked in studies of evidence use in public health and health promotion,[17,108] and has not been included in any models of evaluation capacity to our knowledge. Given our findings highlight the potential of political, funding and administrative systems to influence organisational capacity and evaluation practice, the scope of ECB frameworks and strategies should be expanded beyond the organisation. While it is arguable that public health agencies may be more susceptible to political, funding and administrative challenges than those involved in direct service provision, these findings have relevance to the wider human services field which operates under government policy requirements[74] and within not-for-profit funding agreements.[91,107]

Further research in the health promotion field should seek to replicate these investigations in other jurisdictions, firstly to evaluate the relationships as described in our model, and secondly to explore whether political and administrative factors have similar relationships with evaluation practice, given that the nature and influence of governance[108] and funding structures[92] may vary. Additionally, there is scope for researchers and practitioners seeking to understand and evaluate the effects of ECB strategies to use the EPAS, a valid and reliable survey instrument, in their studies.

We acknowledge the limitations in the scope of this evaluation capacity model and statistical approach. Specifically, the role of external evaluation relationships have not being taken into account in this analysis. Partnerships with universities have been reported as playing an important mediating role in the development of evaluation capacity.[16,44,59,71,72,75] It should be noted that impact evaluation was selected as the primary dependent variable. It is

possible that the final model parameters and fit may have shifted towards evaluation capacity relationships that favours the demonstration of program impact. While this is a priority within the health promotion field,[103] formative and process evaluation also contribute valuable information for shaping effective programs.[56] Despite the relatively small sample of 219 participants for SEM, this study is within the range of acceptable sample sizes according to recommended rule of thumb calculations.[113] The final model had 21 parameters and sample of 216 therefore exceeding the ratio of 10:1 for cases to parameters that is considered necessary for reliable estimation. The use of second order latent variables in path analysis was a pragmatic choice based on sample size. A better method would be to include both latent and observed variables in SEM, however our chosen approach was considered acceptable given that this is, we believe, the first explanatory model of the determinants of evaluation practice in the health promotion field.

In summary, we have developed and validated a model of evaluation capacity, specifically for the health promotion field. Importantly, our study identifies how health promotion organisations mediate the relationship between the wider health promotion and disease prevention system, and evaluation practice through leadership and organisational culture that values evaluation, and using organisational systems and support for evaluation to influence resource allocation and negotiate evaluation requirements. The model draws attention to the pivotal role that the organisation plays in generating demand for evaluation through leadership, culture, systems, support and resource allocation, as well as mediating the effects of political, funding and administrative factors. Critically, this model contributes a validated conceptual framework that health promotion organisations, and the wider health promotion field, can use as a logic model to guide planning and design of ECB initiatives.

7.6 Acknowledgements

The authors would like to acknowledge the contributions of Keren Shanley who recruited participants for the survey, and are grateful to Tilahun Haregu who provided guidance and advice for the statistical analysis. This work was supported by the Australian Research

Council (ARC) [DP150103575]. Prof. Gabbe was supported by an ARC Future Fellowship [FT170100048] during the preparation of this manuscript.

Chapter Eight

8 An audit and quality appraisal of health promotion evaluation

In the studies described in Chapters Five, Six and Seven, qualitative and quantitative methods were used to address gaps in the literature and understanding of the factors that determine health promotion evaluation practice and capacity, and how they interact. Despite these contributions to the literature, a gap in knowledge remains about the relationship between evaluation capacity and evaluation quality. As outlined in Chapter One, poor availability and quality of health promotion program evidence remains a challenge facing policy and program decision makers. It is necessary to understand what influences evaluation quality in health promotion organisations, in order to improve the usefulness of health promotion evaluation. Therefore, the aims of the final study in this thesis were to examine the evaluation methods and quality in Australia and the factors associated with this.

This chapter comprises the manuscript under review:

Schwarzman J, Nau T, Bauman A, Gabbe B, Rissel C, Shilton T, Smith BJ. An assessment of program evaluation methods and quality in Australian prevention agencies. *Health Promotion Journal of Australia* (under review).

In this paper, the content of the collected evaluation reports (n=392) focusing on the levels of evaluation and data collection methods used is summarised, and key organisational attributes associated with a range of evaluation characteristics examined. Using unique organisational identifiers, the data from the EPAS is linked with the evaluation appraisal scores to investigate the key associations between organisational attributes, evaluation capacity and evaluation quality.

An assessment of program evaluation methods and quality in Australian prevention agencies

Joanna Schwarzman¹, Tracy Nau^{1,2}, Adrian Bauman², Belinda J. Gabbe^{1,3}, Chris Rissel²,
Trevor Shilton^{4,5} and Ben J. Smith^{1,2}

¹ School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia;

² Prevention Research Collaboration, Sydney School of Public Health, The University of Sydney, Sydney, Australia;

³ Health Data Research UK, Swansea University Medical School, Swansea University, Singleton Park, Swansea, Wales, United Kingdom;

⁴ School of Public Health, Curtin University, Bentley, Australia;

⁵ National Heart Foundation Western Australia, Subiaco, Australia.

8.1 Abstract

Issue addressed: The evidence base for prevention strategies is limited, with the evidence generated through program evaluation by health promotion and disease prevention agencies lacking rigour. Despite the need to improve the quality of evaluation, there is limited evidence of what influences evaluation quality in the prevention field. This study aimed to examine evaluation methods and quality in Australian health promotion agencies, and the factors associated with this.

Methods: Data were collected using the Evaluation Practice Analysis Survey and an audit and appraisal of evaluation reports. Descriptive analysis was used to examine evaluation characteristics and multivariable regression was used to explore the association between evaluation and organisational attributes and evaluation quality.

Results: In total, 392 evaluation reports were reviewed from 78 government and non-government agencies. Evaluation of delivery or reach (process evaluation) was conducted most frequently (86% or 88% respectively), followed by impact evaluation (69%). Overall evaluation quality was low (median 24.5%). In multivariable regression analysis, only two factors were associated with evaluation quality: health promotion budget (ratio of geometric means 1.53 [95% CI 1.02-2.29]); and, conducting state-wide or national prevention programs (1.38 [95% CI 1.05-1.82]).

Conclusions: The findings show the most potential to improve evaluation quality in smaller organisations that deliver health promotion at a local or regional scale.

So what? By improving the rigour of existing evaluation, there is opportunity to build the evidence base for prevention strategies, and ultimately improve population health outcomes.

8.2 Introduction

In 2016, chronic diseases accounted for 89% of deaths in Australia,[3] with a substantial proportion of morbidity and mortality being preventable.[5] While there is an urgent need

for effective prevention strategies,[35] the evidence-base to select health promotion and disease prevention interventions is limited.[16,17] The nature of health promotion strategies brings inherent challenges to evidence generation, particularly as programs are usually adapted to each unique context and population. High quality evaluation is an essential contributor to the evidence-base needed to inform decisions about policy and resource allocation, the identification of strategies for wide-scale implementation, and learning regarding factors that facilitate successful implementation.[15,16,18]

There are well-established frameworks and tools to guide evaluation of health promotion.[31,32,46] These provide guidance at the various stages of program development and implementation. They emphasise the importance of clear program logic, clarification of evaluation purposes and questions, and selection of data collection methods and evaluation designs to generate useful findings.

Despite this guidance, the quality of evaluation in health promotion has been reported to be low. A study conducted in five metropolitan South Australian community health services used a 16-item tool to appraise the degree of reporting on program planning, logic models and evaluation methodology in 93 evaluations.[60] The reported evaluations scored poorly on items relating to measurement of long-term outcomes, and assessment of sustainability and transferability. Data were most often collected using only one method (typically participant feedback sheets), and sampling and data analysis methods were usually poorly described.[60] Another study compared evaluation documentation from six United Kingdom cities, funded to implement 'Walking Cities' activities, to the mandated Standard Evaluation Framework.[43] The authors concluded that many of the 52 aspects of evaluation and program reporting that were important to quality and consistency were not conducted or not reported, and the overall quality of the reports was low.[43]

Other reviews have examined the evaluation designs and data collection methods used as indicators of the quality of health promotion evaluation. Optimal designs were infrequently

used in physical activity mass media campaigns,[61] or in social media based health promotion.[62] In just over a third of published evaluations of Australian health promotion initiatives, data were collected at only one time point, and almost half of the studies used only one method of data collection.[63] Another study of unpublished health promotion evaluation reports from organisations in metropolitan Melbourne found that the level of evaluation conducted aligned with funding body requirements for process or impact evaluation, and that impact evaluations were conducted on a small scale and the data collection methods were rarely reported.[44]

While these studies have found consistently low quality evaluation, this provides limited information on what needs to be addressed to improve evaluation for decision making. Evaluation practice and quality may be hindered by practitioner skill gaps, lack of time, insufficient resources, a lack of organisational leadership, inadequate support and systems, and unclear or impractical requirements from funding bodies.[44,105] Within the wider funding and policy environment for prevention, a lack of commitment to evaluation can heighten pressures to focus attention away from evaluation.[51,74,105]

In light of the imperative to build the evidence base for health promotion strategies, this study examines reported evaluation methods and quality in Australia, and the factors associated with this. Specifically, this study investigates: the levels of evaluation and data collection methods reported in evaluations undertaken by Australian government and non-government health promotion agencies; the quality of reported evaluations; and organisational attributes and elements of evaluation capacity associated with evaluation quality.

8.3 Methods

8.3.1 Evaluation in health promotion study

The evaluation in health promotion study was conducted in Australia between 2015 and 2018. Government and non-government health promotion and disease prevention agencies

were recruited from New South Wales, South Australia, Victoria and Western Australia (n=116) to take part in up to three phases of data collection: 1) qualitative interviews; 2) the Evaluation Practice Analysis Survey (EPAS); and 3) review and appraisal of evaluation reports. Phases 1 and 2 have been described elsewhere.[104,105,110,119] Inclusion criteria were that organisations had an annual health promotion budget >\$300 000, >3 full time equivalent health promotion staff, and at least 5 years of experience delivering primary prevention strategies. The present study used data collected in phases 2 and 3, which took place between June 2017 and December 2017. The EPAS was emailed for online completion by the two most experienced health promotion practitioners or managers in each organisation. The EPAS is a recently developed survey instrument which measures 25 aspects of evaluation capacity in health promotion organisations; survey development and measurement properties are described elsewhere.[110,119]

Ethics approval for this study was received from Monash University Human Research Ethics Committee (HREC), the Aboriginal Health and Medical Research Council Ethics Committee, Western Australia Country Health HREC and the South Australian Department of Health and Aging HREC (under the National Mutual Acceptance Scheme). Additional site-specific research governance approvals were obtained where requested.

8.3.2 Evaluation report appraisal

Evaluation report collection

Participating organisations were asked to provide all primary prevention program evaluation reports (including full reports, brief reports, templates for funding bodies, evaluation summaries), completed between May 2015 and May 2017. Key organisational contacts were approached by email. Reports were excluded based on four criteria: program duration <12 months; primary purpose to generate knowledge, as opposed to improving health (e.g., research led initiatives, not program evaluation); needs assessment only (work not leading to

a clear intervention); or, quality improvement activities.

Audit and appraisal tool

The evaluation audit and appraisal tool comprised five sections. The first four sections collected descriptive information about the organisation (non-government or government, state), prevention project (health issue addressed, geographical coverage, target group characteristics, duration, budget and strategies used), evaluation arrangements (type of report, evaluators, time-point of evaluation, methodology) and evaluation methods (level of evaluation, design, data collection techniques). The final section comprised a 22-item evaluation quality appraisal tool (Appendix D-1).

The evaluation quality appraisal tool was developed for this study in several stages, including a literature review, consensus based item-selection, piloting and refinement, and assessment of inter-rater agreement. The preliminary tool was developed based on items from three existing evaluation quality frameworks.[120-122] Proposed items were reviewed by a panel of health promotion evaluation experts for relevance, importance and potential applicability to a range of health promotion program evaluations. The items selected for inclusion addressed: information about program design and implementation; statement of evaluation purpose; evaluation design; sampling methods; ethical approval; data collection tools; analysis methods; acknowledgement of limitations; and interpretation of evaluation results.

This tool was piloted on eight purposively selected projects, chosen for a range of project sizes, report formats, organisation types and locations. Three researchers each completed the pilot audits, met to resolve discrepancies and the tool was amended. Two researchers (JS and TN) then each applied the tool to a further 20 evaluation reports. The inter-rater agreement for the 22 quality appraisal items in the 20 reports audited was moderate (Cohen's kappa = 0.69, 95% CI 0.54-0.74). Minor adjustments were subsequently made to improve inter-rater reliability of the tool for appraisal of the remaining reports.

Appraisals were undertaken independently by JS or TN. A total evaluation quality score was

derived by adding the ratings given against each criteria. This was standardised into a percentage, taking into account variations in the maximum possible quality score.

8.3.3 Statistical analysis

Data were imported into Stata 14.0 for analysis. All organisational characteristics and evaluation capacity measures were collected using the online EPAS survey. The eight organisational characteristic variables were the organisation type, budget, staffing levels, location, program scale (e.g. local, regional, state-wide or national), the organisation's length of experience conducting health promotion programs, the availability of an internal evaluation support function within the organisation, and whether the health promotion organisations partnered with universities for evaluation. Any discrepancies between survey participant responses were resolved using publicly available information, or selecting the most conservative estimate of the two responses. Seven measures of evaluation capacity (attitudes and beliefs; skills and knowledge; organisational leadership and culture; organisational systems and support; organisational resources for evaluation; funding body expectations; political and funding environment) were created by summing EPAS sub-scale scores. Respondent scores were averaged to create an organisational level score where two responses were provided. Tertile cut points of evaluation capacity scores were used to create categorical evaluation capacity variables. The survey data were linked to the audit and appraisal data using unique organisational identifiers.

Descriptive analysis examined the characteristics of the evaluation report type and methodology. Additionally, the use of each of the four levels of evaluation, as described by Bauman and Nutbeam[31] (formative, process, impact, and outcome evaluation) were included in the descriptive analysis. Frequencies and percentages describing the data collection methods used at each level of evaluation were calculated. Non-parametric and chi-square statistics were used to assess associations between the attributes of participating organisations and the characteristics and levels of evaluation.

The dependent variable for regression analysis was the evaluation quality score, calculated based on the 22-item appraisal tool. The distribution of the evaluation quality score was non-parametric and was therefore log transformed to achieve normality.

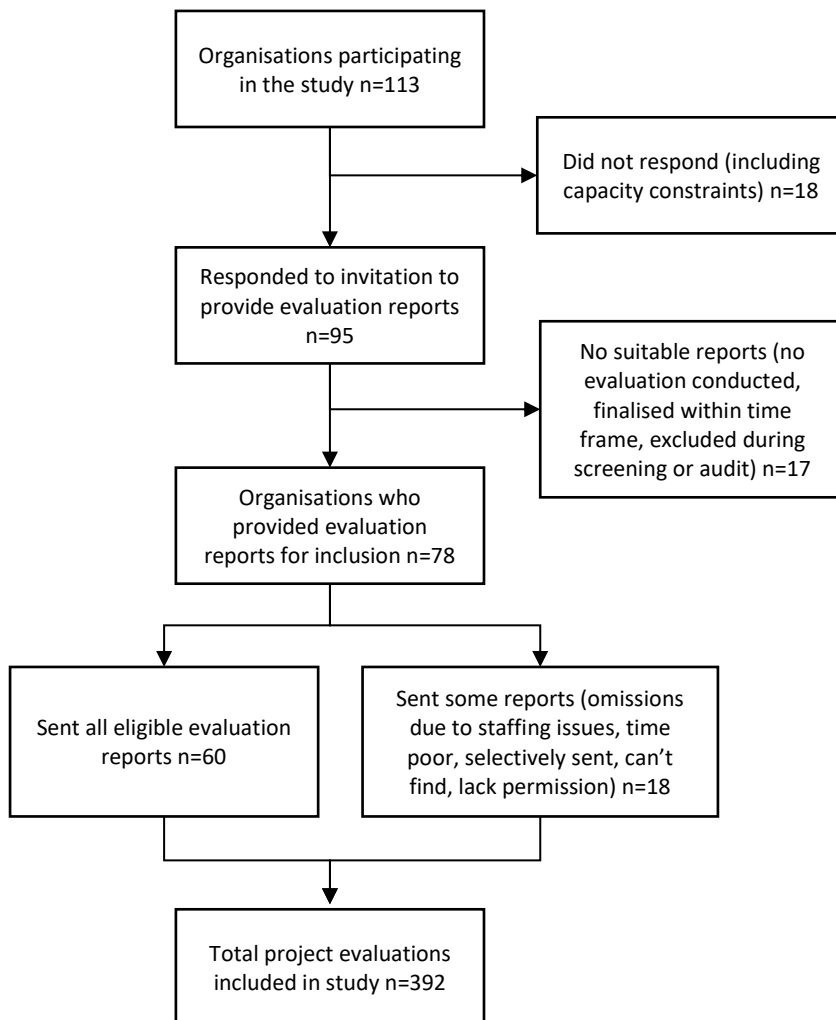
Multivariable linear regression using generalised estimating equations (GEE), with an exchangeable correlation matrix, was conducted to identify the important factors associated with evaluation quality. The GEE approach accounted for the multiple reports per organisation and was selected over multi-level effects techniques given the high levels of variation of number of reports per cluster, and the cross-sectional design. The independent variables were initially chosen based on qualitative research that explores the determinants of evaluation in health promotion. Univariate linear regression (using GEE) was conducted for each independent variable (evaluation capacity measures, organisational characteristics, report type). Variables were included in the multivariable regression model if the univariate model indicated an association defined as a $p\text{-value} < 0.1$.

8.4 Results

8.4.1 Survey sample and evaluation reports collected

The EPAS was emailed to 230 practitioners from 115 participating organisations. In total, 219 (95%) health promotion practitioners or managers completed the survey. Details of the survey sample have been described previously.[119] Three organisations withdrew before Phase 3 (evaluation report collection). Of the 113 organisations participating in the documentary review phase, 84% responded to the request for evaluation reports, including 78 organisations which provided eligible reports for inclusion in the study. Seventeen organisations indicated that they had no evaluation reports completed within the two-year time-frame or sent reports that were not considered program evaluation (i.e. research, needs assessment only or quality improvement activity), or evaluated programs of <12 months duration. These were excluded from analysis. Figure 4 summarises the process of evaluation report collection.

Figure 4. Organisational response to evaluation report collection



The final sample included 392 project evaluations collected from 78 organisations (range 1-24 per organisation). Details of the organisational participants are shown in Table 6. The median number of reports per organisation was 3 (inter-quartile range 1.75-7.00), and there were no significant differences in the median number of reports provided across the majority of organisational characteristics (Table 6). However, organisations that delivered local or regional programs provided a median of 4.5 reports compared to those that delivered state-wide or national programs which provided 2 reports ($p < 0.01$).

8.4.2 Evaluation characteristics

Table 7 shows the prevalence of evaluation report types, evaluators and methodologies, compared by organisational characteristics. Full evaluation reports (compared to templates or brief reports) were completed more often by organisations based in metropolitan areas, with the highest budgets and staffing levels, and those that delivered state-wide or national programs. Similarly, larger, metropolitan based organisations, conducting state-wide or national programs had more reports that were completed by an external evaluator compared to smaller, regional organisations, and those with smaller program jurisdictions. The differences in data collection methodology used in evaluations was statistically significant for all organisation characteristics (Table 7). Mixed-methods were the most frequently reported across all categories (65%), followed by quantitative methods (30%). The organisational characteristics associated with higher frequencies of mixed-methods evaluations were non-government organisations, those with lower or middle sized budgets, and lower staffing. Those with an internal evaluation role, or which worked with universities conducted a lower proportion of mixed-methods evaluations. Government agencies, those rural or regionally based, and small scale organisations conducted a higher proportion of quantitative only evaluations, as did those with higher budgets, more staff experience and those with access to evaluation support. Only 18 qualitative reports were reviewed. A higher proportion of qualitative only methodologies were used by organisations that were metropolitan based, covered state-wide or national programs, had higher budgets, higher levels of staffing, and more years of experience in health promotion

Table 6. Characteristics of participating organisations

	Organisations		Reports Total N	Reports per organisation		p value [†]
	n	%		Median	IQ range	
Total	78	100.0	392	3.00	1.75 – 7.00	
Organisation type						0.414
Government	32	41.0	168	3.50	2.00 – 7.75	
Non-government	46	59.0	224	3.00	1.00 – 6.00	
Organisational location						0.287
Metropolitan	58	74.4	284	3.00	1.00 – 6.00	
Regional or rural	20	25.6	108	4.00	2.00 – 7.00	
Size region covered by organisation						0.008*
Local / regional	44	56.4	283	4.50	2.00 – 8.75	
State / national	34	43.6	109	2.00	1.00 – 4.25	
Annual health promotion budget [‡]						0.814
< AU\$500 000	29	38.2	134	4.00	1.50 – 5.50	
AU\$500 000 to \$1 million	21	27.6	96	2.00	1.50 – 7.00	
> AU\$1 million	26	34.2	147	3.50	1.00 – 9.00	
Health promotion staffing (full time equivalent)						0.062
7 or less	46	59.0	188	3.00	1.00 – 5.00	
8 or more	32	41.0	204	4.50	2.00 – 9.75	
Organisational experience in health promotion						0.434
10 years or less	19	24.4	96	3.00	1.00 – 5.00	
More than 10 years	59	75.6	296	3.00	2.00 – 7.00	
Evaluation role in the organisation						0.513
Yes	23	70.5	264	3.00	2.00 – 7.00	
No	55	29.5	128	3.00	1.00 – 7.00	
Worked with University evaluators						0.226
Yes	69	88.5	333	3.00	1.50 – 6.50	
No	9	11.5	59	6.00	2.00 – 11.50	

[†]Mann-Whitney-Wilcoxon Test, except for Annual health promotion budget (Kruskal-Wallis); *p<0.05. [‡]"Don't know" responses not included n=2 orgs. IQ = inter-quartile range (Q1-Q3)

EVALUATION AUDIT & APPRAISAL

Table 7. Evaluation characteristics by key organisational attributes

	Total	(%)	Report type†			p-value	Evaluation conducted by‡			p-value	Methodology used†			p-value
			Full report (%)	Template report (%)	Brief report (%)		HP team (%)	HP team plus external (%)	External evaluator (%)		Qualitative (%)	Quantitative (%)	Mixed methods (%)	
Total number	392		131	193	46		259	65	68		18	117	255	
Total %	100		35.4	52.2	12.4		66.1	16.6	17.3		4.6	30.0	65.4	
Organisation type						0.068				0.037*				<0.001*
Non-government	224	57.1	37.0	53.9	9.1		61.6	20.5	17.9		4.0	19.3	76.7	
Government	168	42.9	33.1	49.7	17.2		72.0	11.3	16.7		5.4	44.3	50.3	
Location						<0.001*				<0.001*				0.028*
Metropolitan	284	72.4	41.4	45.5	13.2		62.0	15.5	22.5		6.0	27.3	66.7	
Regional or rural	108	27.6	20.2	69.2	10.6		76.9	19.4	3.7		0.9	37.0	62.0	
Region covered by organisation						<0.001*				<0.001*				<0.001*
Local or regional	283	72.2	20.9	67.4	11.7		78.8	16.3	4.9		1.8	32.2	66.1	
State or National	109	27.8	76.3	9.3	14.4		33.0	17.4	49.5		12.1	24.3	63.6	
Annual health promotion budgets§						<0.001*				<0.001*				0.006*
< \$500 000	134	34.2	31.0	58.9	10.1		81.3	7.5	11.2		2.3	25.6	72.2	
\$500 000 to \$1 mill	96	24.5	20.9	68.1	11.0		67.7	17.7	14.6		3.1	20.8	76.0	
> \$1 million	147	37.5	51.9	31.9	16.3		48.3	25.9	25.9		8.2	34.9	56.8	
Health promotion staff (FTE)						<0.001*				<0.001*				<0.001*
7 or less	188	48.0	25.1	63.1	11.7		77.1	12.2	10.6		2.7	21.4	75.9	
8 or more	204	52.0	45.0	41.9	13.1		55.9	20.6	23.5		6.4	37.9	55.7	
Organisational experience in health promotion						0.640				0.110				<0.001
10 years or less	96	24.5	33.7	51.1	15.2		72.9	16.7	10.4		9.4	16.7	74.0	
More than 10 years	296	75.5	36.0	52.5	11.5		63.9	16.6	19.6		3.1	34.4	62.6	
Internal evaluation role						0.440				0.084				0.009*
Yes	128	32.7	37.2	47.9	14.9		61.7	14.8	23.4		7.1	37.8	55.1	
No	264	67.3	34.5	54.2	11.2		68.2	17.4	14.4		3.4	26.2	70.3	
Work with university evaluators						<0.001*				<0.001*				0.001*
Yes	333	84.9	40.4	52.2	11.5		61.6	18.0	20.4		5.4	32.9	61.6	
No	59	15.1	8.6	74.1	17.2		91.5	8.5	0.0		0.0	13.6	86.4	

†Categories with small numbers excluded from Table (Report types 'Peer-reviewed', 'Presentation', 'Data sheet', 'Other' n=22; Methodology 'Economic' and 'Not-specified' n=2). ‡Where there were no authors or evaluators specified it was assumed to have been conducted internally by the health promotion team. §Does not include 'Don't know' responses (n=15). *p<0.05. HP (health promotion); FTE (Full time equivalent).

The frequency of each level of evaluation, and the top three data collection methods used at each level, is presented in Table 8. The most frequently reported levels of evaluation were the process evaluation elements of delivery (86%) and reach (88%). The next most common was impact evaluation (69%), followed by evaluation of participant satisfaction (56%). Formative evaluation was included in less than a third of reports, and outcome evaluation was only reported in 9%. Surveys were the most commonly used data collection tool for all levels of evaluation, except for the process evaluation of delivery, reach and context where program records or logs were used most often.

Table 8. Summary of data collection methods used at each level of evaluation

Level of evaluation	Reports [†]		Top 3 methods (% at this level)
	No.	%	
Formative	108	28	survey (40); focus groups (21); interviews (20);
Process			
- Delivery	338	86	program records/log (94); document review (9); survey (2); interview (2);
- Reach	344	88	program records/log (51); attendance records (41); social media analytics (16);
- Exposure	66	17	survey (48); program records/log (18); attendance records (17);
- Context	178	45	program records/log (31); interview (24); reflection (22);
- Satisfaction	221	56	survey (73); feedback (14); interviews (11);
Impact	269	69	survey (72); routine data (14); interviews (14);
Outcome	37	9	survey (57); physical assessment (22); routine data (16).

[†]No. reports and % are greater than total reports (or 100%) due to multiple levels or data collection methods per evaluation.

As shown in Table 9, metropolitan based organisations provided a higher proportion of reports that incorporated formative evaluation than organisations in regional locations. Organisations conducting local and regional programs conducted significantly higher proportions of process evaluations of delivery and reach, whereas those undertaking state-wide and national programs had higher prevalence of exposure evaluation. Participant satisfaction evaluation was included in a greater proportion of non-government evaluations, and also those of organisations with lower budgets, fewer staff, less experience and those which did not work with university evaluators. While the total number of outcome evaluations was small (n=37), this was most common in organisations based in metropolitan locations, with higher budgets, more staff, having an internal evaluation role and engaging university evaluators.

EVALUATION AUDIT & APPRAISAL

Table 9. Levels of evaluation by key organisational attributes

	Formative			Process						Impact		Outcome				
	n	P-value	Delivery	p-value	Reach	p-value	Exposure	p-value	Context	p-value	Satisfaction	p value	p-value	p-value		
Total number	n=109		n=340		n=346		n=67		n=179		n=223		n=271		n=37	
Organisation type (%)																
Non-government	30.8	0.096	88.8	0.083	89.3	0.286	16.5	0.846	44.6	0.725	63.4	0.001*	76.3	<0.001*	6.7	0.032*
Government	23.2		82.7		85.7		17.3		46.4		47.0		58.3		13.1	
Location (%)																
Metropolitan	31.0	0.014*	85.6	0.538	87.3	0.673	18.0	0.336	46.8	0.359	55.6	0.630	70.4	0.213	11.3	0.045*
Regional or rural	18.5		88.0		88.9		13.9		41.7		58.3		63.9		4.6	
Region covered by organisation (%)																
Local or regional	26.5	0.454	92.6	<0.001*	92.3	<0.001*	11.3	<0.001*	43.1	0.141	56.5	0.918	69.3	0.662	4.9	<0.001*
State or National	30.3		69.7		76.1		31.2		51.4		56.0		67.0		21.1	
Health promotion budget (%)†																
< \$500 000	20.1	0.080	85.1	0.007*	87.3	0.442	18.7	0.003*	48.5	0.564	61.9	0.009*	61.2	0.038*	8.2	0.007*
\$500 000 to \$1 million	27.1		94.8		90.6		6.3		45.8		66.7		77.1		3.1	
> \$1 million	32.0		80.3		85.0		23.1		42.2		48.3		68.7		15.0	
Health promotion staff (FTE) (%)																
7 or less	23.9	0.124	88.3	0.253	87.8	0.995	14.4	0.209	50.5	0.050	68.1	<0.001*	67.6	0.661	6.4	0.047*
8 or more	30.9		84.3		87.7		19.1		40.7		45.6		69.6		12.3	
Organisation's years HP experience (%)																
10 years or less	34.4	0.085	92.7	0.034*	90.6	0.324	15.6	0.715	59.4	0.002*	74.0	<0.001*	74.0	0.195	10.4	0.706
More than 10 years	25.3		84.1		86.8		17.2		40.9		50.7		66.9		9.1	
Internal evaluation role (%)																
Yes	31.5	0.368	83.6	0.293	85.9	0.445	17.2	0.897	40.3	0.185	50.0	0.076	64.8	0.262	14.8	0.011*
No	26.1		87.5		88.6		16.7		47.7		59.5		70.5		6.8	
Work with university evaluators (%)																
Yes	27.9	0.692	85.6	0.383	87.4	0.598	18.0	0.138	45.0	0.732	53.2	0.002*	67.9	0.444	11.1	0.007*
No	25.4		89.8		89.8		10.2		47.5		74.6		72.9		0.0	

†Does not include 'Don't know' responses (n=15). *p<0.05. FTE (Full time equivalent); HP (Health promotion)

8.4.3 Factors associated with evaluation quality

The median quality score for all evaluations was 24.5% (interquartile range 13.3 to 46.6; mean 33.2%). Univariate GEE models identified one evaluation capacity variable and five organisational variables for inclusion in the multivariable model at the $p < 0.1$ level: organisational leadership and culture, health promotion budget, staffing number, size of program delivery region, designation of an evaluation role in the organisation and working with university evaluators. A final sample of 377 reports, grouped by 76 organisations were included in the multivariable analysis. Table 10 shows that organisational characteristics that were associated with higher evaluation quality scores were having an annual health promotion budget $> \$1$ million ($p = 0.04$) and conducting programs at a state-wide or national level ($p = 0.02$). The presence of an evaluation role within the organisation was close to the significance threshold ($p = 0.05$).

Table 10. Generalised Estimating Equation (GEE) analysis of factors associated with evaluation quality score[†]

Organisational attributes	Evaluation quality scores [†]		Ratio of geometric means (95% CI)
	Mean (SD)	Median (IQ range)	
<i>Organisational leadership and culture</i> [‡]			
Low	24.8 (16.1)	20.0 (20.0)	REF
Medium	32.7 (21.6)	24.0 (30.0)	1.00 (0.72-1.41)
High	41.5 (27.2)	33.3 (51.0)	1.12 (0.79-1.60)
<i>Annual health promotion budget (AUD)</i> [‡]			
< \$500 000	26.9 (16.1)	21.7 (23.3)	REF
\$500 000 to \$1 million	27.9 (21.6)	20.0 (16.9)	1.18 (0.86-1.63)
> \$1 million	43.5 (26.2)	36.7 (44.3)	1.53 (1.02-2.29)*
<i>Health promotion staff (FTE)</i> [‡]			
7 or less	28.0 (18.4)	20.4 (21.4)	REF
8 or more	38.0 (26.1)	27.6 (46.7)	0.82 (0.56-1.19)
<i>Region covered by organisation</i> [‡]			
Local or regional	25.5 (17.2)	20.0 (20.0)	REF
State-wide or National	53.4 (24.9)	53.3 (41.7)	1.38 (1.05-1.82)*
<i>Internal evaluation role</i> [‡]			
No	30.2 (20.7)	20.4 (26.7)	REF
Yes	39.6 (26.7)	26.7 (44.7)	1.35 (1.00-1.84)
<i>Work with university evaluators</i> [‡]			
No	22.8 (11.5)	20.0 (20.0)	REF
Yes	35.1 (24.3)	26.7 (39.5)	1.39 (0.92-2.09)

[†]A higher evaluation quality score indicates better evaluation quality; [‡]Mann-Whitney U or Kruskal Wallis test significant at p<0.01; *Multiple regression significant at p<0.05; SD (standard deviation); IQ range (Interquartile range Q3-Q1); REF (Reference category); FTE (Full time equivalent).

8.5 Discussion

This is the first study to our knowledge that examines the content and quality of a large body of health promotion evaluation, drawn from multiple jurisdictions. The findings provide an overview of recent evaluation practice, the levels of evaluation, and data collection methods used by organisations with different attributes. Further, this study identified which organisational and evaluation capacity factors were associated with evaluation quality, which contributes to addressing a gap in the literature that will assist those seeking to improve the evidence base for prevention strategies.

We found process evaluation was common across all reports, with particularly high rates of

reporting program delivery and reach, followed by evaluation of participant satisfaction, particularly in organisations conducting small scale programs. Similarly, evaluation of participant satisfaction was conducted more often by non-government organisations, those less well resourced, less experienced and who did not work with university evaluators. While reporting on program delivery, reach and participant satisfaction are valuable components of program evaluation,[31] it is possible that these levels of evaluation are used by smaller organisations to meet mandatory reporting requirements.[44] Impact evaluation was included in 69% of evaluation reports, and the rate of impact evaluation was higher in non-government than government evaluation reports. This may result from explicit funding body requirements for non-government organisations to report on program effects, the imperative to demonstrate effectiveness to justify funding and compete for future funding opportunities,[56,105,107] or the nature of the intervention and type of impacts expected from the program.

Formative evaluation was identified in less than 30% of reports. This rate is consistent with a previous Australian study,[44] and indicates an important evaluation weakness as pre-testing program materials and methods prior to implementation can improve the likelihood of program acceptability, appropriateness and overall success.[31,123] Previous studies have identified a range of pressures on program teams to initiate implementation, with the consequence that time and resources are not allocated to formative evaluation.[59,104]

In this study, mixed-methods was found to be the most common approach reported in evaluations, followed by quantitative methods. In light of this finding, it is unsurprising that surveys were the most frequently used data collection used for formative, exposure, satisfaction, impact and outcome evaluation. Program records were the most common method used in process evaluation of delivery, reach and context. Reviews of published and unpublished evaluations have also identified the widespread use of surveys.[44,63] The predominance of surveys reflects the high number of impact evaluations and is likely because self-reported measures are a feasible and affordable way to collect data on a range of

indicators.[16] Contrary to this, studies exploring the barriers and facilitators to evaluation in health promotion organisations indicated staff were reluctant to use surveys to collect data from disadvantaged or youth populations, and identified a need for alternative data collection methods that were suitable for these specific contexts.[71,72]

In line with previous studies,[43,63] we identified the overall quality of evaluation reported to be low. The median evaluation quality score was 24.5% (mean 33.2%). Organisations which had annual health promotion budgets of over \$1 million had adjusted mean evaluation quality scores 53% higher than those with budgets under \$500 000. Having access to adequate resources has been frequently identified as an important facilitator of evaluation practice.[44,59,71,72,104,105,119] Additionally, organisations that delivered state-wide and national programs were found to have adjusted mean quality scores 38% higher than locally-oriented organisations. In some regards, this is not surprising, and even acceptable as some guidelines suggest program budget thresholds for evaluation,[50] and evidence from practitioners indicate this can occur in the health promotion field in Australia.[104] Similarly, state-wide and national programs may also attract more political attention which has also been identified as a driver of evaluation in health promotion.[105] Indeed, initiatives that have the potential to impact large populations, should have rigorous evaluation of their effects.[116] In this study, we did not take into account the nature of health promotion programs, or the populations groups targeted. The challenges entailed in data collection and evaluation for disadvantaged and youth populations[71,72] may also affect the quality of evaluation. The challenge for the health promotion field is to improve evaluation quality across all organisational types, sizes and program areas.

Given the pivotal influence of leadership in evaluation practice and capacity,[74,83,84,104,110,119] it was surprising that the variable 'organisational leadership and culture' was not an independent predictor of evaluation quality in this study. This may be due to the overriding influences of financial resources and scale of programs upon evaluation. However, it is possible that the construct 'leadership and culture' is not the best

representation of the key components of leadership, as in this study the scale includes items measuring partnerships and organisational influence. Further investigation of the effects of leadership upon evaluation quality are warranted. We were also surprised to find that having an internal evaluation role, or working with university evaluators was not associated with evaluation quality. A lack of access to evaluation expertise and insufficient resources have been frequently identified as barriers to health promotion evaluation[44,59,60,71,72,104] and working with external evaluators has been identified as a facilitator to evaluation, particularly if there is an ongoing collaborative relationship with the program team.[103,104] In our study we did not examine the nature of support provided by the internal or external evaluation specialists, and further research should explore the particular functions of these relationships that provide the most benefit to evaluation quality.

A strength of the study is the inclusion of health promotion organisations from multiple Australian jurisdictions, and the collection and appraisal of both published and unpublished evaluation reports. In the development of the evaluation audit and appraisal tool, we undertook systematic steps to enhance content and face validity of the items, and inter-rater reliability. However we also acknowledge several limitations. Small health promotion organisations were not included in this study, therefore the findings may not be representative of these organisations. Organisations voluntarily provided evaluation reports, and while the majority were assumed to be complete samples of evaluation from a two-year period, a small number of organisations indicated they had selectively provided reports. Additionally, by requesting only documented, written evaluation reports we may not have captured the full range of evaluation activities completed by organisations.

8.6 Conclusions

While evaluation is widely understood to play a critical role in strengthening health promotion policies and programs, there have been few studies which have systematically examined evaluation methods and quality in this field. We have found that there is significant scope for improving the breadth and rigour of evaluation in health promotion and

prevention organisations in Australia, particularly in smaller, locally-oriented organisations and government agencies. Achieving this will require greater investment by funding agencies, commitment by managers, and strategic capacity building initiatives that address the system-, organisational-, and practitioner-level influences upon evaluation practice and quality. It will be necessary to ensure that evaluation capacity building efforts are themselves systemically evaluated, so that the enablers for program learning and evidence generation can be identified and embedded within health promotion and prevention organisations.

8.7 Acknowledgements

The authors would like to acknowledge the contributions of Keren Shanley who recruited participants for the survey and collected evaluation reports from participating organisations. This work was supported by the Australian Research Council (ARC) [DP150103575]. Prof. Gabbe was supported by an ARC Future Fellowship [FT170100048] during the preparation of this manuscript.

9 Integrated discussion

The research presented in this thesis was used to determine the reported strengths and limitations of evaluation practice in Australian health promotion and identify factors that need to be addressed to improve evaluation quality. Evidence to meet these aims was obtained through a nationwide mixed-methods study that engaged 116 government agencies and non-government health promotion organisations, and over 400 health promotion practitioners and managers.

The rationale for this thesis emerged from the need to grow the evidence base for prevention programs through improved quality and comprehensiveness of program evaluation. As outlined in Chapter One, reviews of published and some unpublished health promotion evaluation reports in discrete area of health promotion practice, describe evaluation to be of low quality and have limited usefulness in decision making.

The literature review presented in Chapter Two examined the known barriers and facilitators to evaluation within the health promotion field. These barriers and facilitators to health promotion evaluation appeared to act at the practitioner-, organisational- and system-levels. Studies from the wider evaluation field focused predominantly on organisational and individual evaluation capacity, and have progressed attempts to measure components of evaluation capacity. In Chapters One and Two, a number of gaps in the literature were identified, specifically:

1. Studies of the influences upon health promotion evaluation have been conducted in limited areas of practice, and used only qualitative methods.
2. In the wider evaluation literature the concept of evaluation capacity is limited to organisational and individual components, and the applicability of these components to health promotion organisations is not known.

3. Studies from outside the health promotion field describe the measurement of evaluation capacity, and three attempts to validate conceptual models of evaluation capacity were found, using the perspectives of evaluation experts, or human service program managers. There are gaps in the assessment of the psychometric properties of these measures, and their validity for the health promotion field is not well-described.
4. Finally, no studies were identified that assessed the relationship between aspects of evaluation capacity and evaluation quality in health promotion, or the wider evaluation field.

In this final chapter, an integrated discussion of key findings and implications for ECB approaches in the health promotion field is provided. Next, the strengths and limitations of this research project are addressed. Future directions for ECB practice and research are discussed and, to conclude, recommendations are made to guide the priorities for ECB.

9.1 Key findings and implications

The research presented in Chapters Five to Eight makes a number of contributions to the literature. In this section, a summary of important findings is presented in relation to their implications for ECB in health promotion.

9.1.1 Health promotion evaluation is influenced by factors acting at the system-level, as well as at the practitioner- and organisational-levels

The studies in this thesis initially used qualitative methods to identify the influences upon evaluation practice and capacity across the health promotion field (Chapter Five). The depth and breadth of understanding that this provided about the determinants of evaluation capacity specific to health promotion practice, is an important contribution to the literature.

While some of the organisational- and practitioner-level determinants identified in this thesis were similar to the broader evaluation capacity literature, including organisational

systems, culture and resources for evaluation,[68,76,84,86] system-level factors that directly or indirectly impact health promotion evaluation capacity and practice were also identified. In particular, policy priorities, funding amounts and accountability requirements were found to be critical determinants of evaluation in both qualitative[105] and quantitative[119] components of this thesis. The identification of these external influences upon health promotion evaluation highlights a number of significant implications for ECB in the field.

Firstly, evidence that the political and policy environment directly affects evaluation approaches and funding, and that commitment to health promotion more broadly has an indirect influence,[105,119] points to the need for ongoing advocacy at the all levels of decision making. Evidence-based public health requires practitioners and researchers to develop skills in communication and advocacy to engage more effectively with policy makers and shape evidence-based decision making within the wider funding and political environment.[15,52]

Evaluation is often used for accountability through reporting on outputs or activities,[14,58,124] and this should be extended to contribute robust evidence to inform program design and improvement.[16,56,57,89] There is significant potential to transform reporting and evaluation requirements for health promotion programs to encourage useful and rigorous evaluation, however this must occur in negotiation with prevention organisations, to encourage evaluation that is both meaningful and achievable.[56]

9.1.2 An instrument developed to measure health promotion evaluation capacity was found to have strong psychometric properties

The EPAS developed in this research was found to reliably measure the components of health promotion evaluation capacity (Chapters Six and Seven).[110,119] The EPAS can be used by those seeking to undertake evaluation capacity needs assessment in a range of health promotion organisations. It should also be used to assess the effectiveness of ECB initiatives implemented within the health promotion field, where evaluation of ECB impact is a

recognised priority.[65,66,87,88]

ECB experts have called for progress in the development of valid and reliable evaluation capacity measurement instruments,[65,66,68,87,88] for a range of organisational contexts.[79] The EPAS is the first evaluation capacity measurement instrument developed specifically for health promotion organisations. The comprehensiveness of the EPAS is a particular strength; the scales measure 25 aspects of evaluation capacity across practitioner-, organisational-, and system-level domains.[110] Specifically, scales were developed to measure skills and knowledge, attitudes and beliefs, organisational leadership, systems and structures, resources, time, support, partnerships, funding body expectations, political and funding influences upon evaluation.[110,119] A further strength of the EPAS is that the scale structure was assessed using both PCA[110] and CFA,[119] in separate samples of health promotion practitioners. The EPAS scales were found to have good internal consistency, predictive and construct validity.[110,119]

9.1.3 The factors that affect evaluation practice are interrelated and there are pivotal influences which present opportunities for capacity building.

The interactions between the important determinants of evaluation practice were identified using qualitative methods (Chapter Five) and confirmed using statistical methods (Chapter Seven). Chapter Seven describes the development and assessment of the first explanatory model for evaluation capacity and practice in the health promotion and primary prevention field.[119] In this study, the role of organisational leadership and culture, organisational systems and support, and the funding and accountability environment were confirmed as pivotal and interrelated determinants of evaluation practice.[119] These findings highlight significant potential to improve evaluation capacity by targeting factors within health promotion organisations, and the wider prevention system.

Leadership for health promotion and evaluation was found to be positively associated with organisational culture, resource allocation, development of systems, engagement with

evaluation and practice partners, and provision of staff support.[104] Leadership has been identified as a determinant of organisational evaluation capacity in other studies,[76,81,91] however the study reported in this thesis also identified how leaders engaged strategically with policy makers and elected officials to demonstrate and promote the value of evaluation and evidence use.[105] Similarly, leaders who understood and valued evaluation were able to moderate the challenges posed by accountability requirements for evaluation by negotiating more favourable conditions with their funding body.[105] The cultivation of health promotion leaders who are effective advocates and role models for the value of prevention, program evaluation and use of evidence has the potential to increase evaluation capacity both within and outside of the organisation.[52,65,83,125,126]

The validated model of evaluation capacity also shows a number of interrelated organisational factors to be important components of evaluation capacity.[119] The relative importance of organisation factors to evaluation capacity is not new.[76,84] However, evidence of the interactions between organisational culture, leadership, systems, support and allocation of resources in health promotion organisations is an important contribution to the literature.[119] Qualitative evidence (Chapter Five) helps explain these interactions: organisations with strong leadership, culture of evaluation, and systems and resourcing to support evaluation were able to moderate the effects of mandatory reporting requirements and the wider political and funding environment.[119] Those seeking to build evaluation capacity within health promotion organisations should be heartened by the results that improvements to internal evaluation capacity may protect against challenges to evaluation from outside the organisation.

9.1.4 Reported evaluations in the health promotion field are of low quality and show consistent gaps

In the final phase of this mixed-methods study an audit and appraisal of 392 evaluation reports was conducted (Chapter Eight), reflecting two years' of health promotion evaluation

from government agencies and NGOs.[127] The overall quality of evaluation reports was low, with a median quality score of 25.4%,[127] which was consistent with previous reviews.[43,44,60-63] Almost all of the reports included in this study were unpublished. Previous health promotion evaluation quality appraisals have been limited, however, to published peer-reviewed evaluations,[61-63] or selected organisational environments and areas of practice.[43,44,60] As health promotion evaluation reports are not typically published, the audit and appraisal of reports collected from multiple organisations and jurisdictions enabled more thorough analysis of current health promotion evaluation practice.

The audit of evaluation reports found that delivery, reach and satisfaction aspects of process evaluation was commonplace, as was impact evaluation, although this did differ by organisation type.[127] Additionally, it was found that reporting of formative evaluation remains a weakness in the field.[127] Although it is possible that formative evaluation of health promotion programs is conducted, yet not documented or reported. Formative evaluation can be often overlooked in the context of pressures to initiate program delivery,[31] despite the potential that this level of evaluation can offer through insight into the early success (or failure) of a strategy.[123]

The findings from the audit also indicate that evaluation conducted in some organisations may be led, and limited by, reporting requirements as over half of the reports were presented in a template format.[44,127] Evaluation conducted for accountability can focus on reporting of outputs and activities that are easily measured,[14,58] at the expense of evaluation that may be useful to inform program improvements[56,128] or demonstrate program impact.[44] This finding reinforces the potential to improve evaluation comprehensiveness and quality through well-designed and jointly negotiated funding agreements, associated reporting requirements[56] and evaluation reporting templates. Others have found that the quality of evaluation reporting can be strengthened by increased use of program logic models, realistic and relevant program objectives, and evaluability assessment at the

planning stage.[43] To address the utility and quality of health promotion evaluation, this study has highlighted the need to consider practitioner and organisational evaluation capacity, resourcing for both planning and evaluation, as well as policy, funding and administrative drivers of evaluation.

9.1.5 Evaluation quality is associated with large health promotion budgets, and delivering large scale programs.

This research found that larger health promotion budgets and conducting programs on a large scale were independently associated with higher quality evaluation in Australian health promotion organisations (Chapter Eight). Evaluation capacity constructs were not found to be associated with evaluation quality in the model,[127] despite the emphasis placed on leadership, organisational culture and systems in qualitative studies[44,60,71,72,74,75,104] and evidence of an association with reported evaluation practice.[110,119] No other studies assessing the relationship between evaluation capacity and evaluation quality have been identified (Chapter Two). Indeed, evaluation quality is infrequently considered in definitions and aims of ECB,[67-70,83] despite the critical importance of high quality evidence of program implementation and effectiveness to inform decision making, resource allocation and program improvement in public health.[15,16,18]

This research highlights that the greatest potential for gains in health promotion evaluation quality may rest with improving evaluation capacity of smaller, locally or regionally oriented organisations (Chapter Eight).[127] However, there is also need for further investigation into the aspects of evaluation capacity that should be addressed in these organisations, compared to larger organisations. Studies of organisational evaluation capacity have postulated that organisations may fit into clusters that experience common barriers and facilitators to evaluation.[69,79,91] Those with limited capacity may conduct evaluation primarily for compliance purposes, those with a moderate level of capacity may use evaluation findings to make program improvements, and those with advanced capacity have most likely to have

embedded evaluation within the organisations and use more rigorous methods to advance knowledge and theory within the field.[69,79,91] Identifying patterns of evaluation capacity across locally focussed organisations with smaller budgets will assist the design of ECB initiatives in these contexts.

9.2 Strengths and limitations of this research

The research that contributes to this thesis has a number of strengths. Specially, this study was conducted nationwide, involving all state and territory jurisdictions, government agencies and NGOs that deliver health promotion in Australia. In a number of stages of this study, participation rates were excellent. The use of a sequential, mixed-methods design allowed for the triangulation of findings between qualitative and quantitative stages.

Further, systematic and rigorous methods were used to develop the survey instrument used for data collection, which subsequently demonstrated internal consistency, predictive and construct validity of scales. Another strength of this component of the research was that separate samples were recruited for each stage of the instrument validation, reducing bias in the development of the final model of evaluation capacity presented in Chapter Seven.

The limitations of the research should also be acknowledged. While every effort was made during the screening of organisations for participation, it is unknown if every eligible health promotion organisation was contacted. It is also possible that organisational eligibility may have fluctuated during the two year period of recruitment and data collection. The smallest health promotion organisations were not accounted for in this study, as eligibility criteria for budget and staffing were set to ensure a minimum probable level of evaluation. Subsequently the findings from this study may not apply to these organisations. Limitations of each phase of the research are identified in Chapters Five to Eight. It is worth noting that the sample size for PCA and SEM analysis in Chapters Six and Seven respectively, may be considered small and have implications for reliability of the findings, as both approaches are typically 'large sample' techniques.[115] The EPAS was validated based on responses from health promotion practitioners and managers in Australia, and further assessment would be require to establish

its validity in other sectors or from other perspectives, such as evaluators. Further, the measurement of evaluation capacity relied on self-report surveys. Despite seeking the perspectives of two representative of each organisation in Phase 2B of EPAS data collection (Chapters Seven and Eight), these responses may not accurately reflect the situation of the whole organisation.

9.3 Future directions for evaluation capacity building

In this thesis, a number of important new insights into the determinants of evaluation capacity and their interactions have been identified. These findings can now be used to inform the development and testing of ECB initiatives that address evaluation capacity at the practitioner-, organisational-, and system-levels.

To date, ECB has typically focused on addressing practitioner knowledge, skills and behaviours, often in combination with strategies to address organisational level factors.[65,68] Two review studies have described the nature of ECB interventions across a range of sectors,[65,68] and a number of subsequent ECB initiatives within the health promotion field have been published.[102,117,118,129] These studies indicate ECB frequently comprise training workshops or technical assistance to address practitioner knowledge and skills in evaluation,[65,68,117,118,129] or program planning.[102] Additionally, ECB strategies have also been implemented to improve organisational processes and policies, develop organisational partnerships with external evaluators, recruit skilled evaluation staff, or undertake evaluation capacity assessment.[65,68,117,129]

Despite an increasing number of studies describing ECB,[65,68] there has been limited evidence generated concerning strategies that are likely to be effective.[65] Studies that evaluate the impact of ECB have predominantly used observational or qualitative designs,[65,118,129] and small sample sizes.[102,117] In the absence of robust evidence of ECB strategy effectiveness, a number of factors that may contribute to the success of interventions have been postulated. Factors that may contribute to the success of ECB

include tailored approaches based on needs assessment, multi-level strategies, generating organisational 'buy-in' to evaluation and ECB, practice oriented training, and ongoing technical support for evaluation.[65,129]

There is potential for greater gains to evaluation capacity if a wider range of interventions in the organisation were considered in line with findings from this thesis. Future strategies should focus on promoting leadership (or champions[126]) for evaluation, changing organisational culture and modifying organisational systems to support evaluation. Further, few examples of ECB have addressed determinants of evaluation capacity within a wider funding or policy environment. The limited attention to system-level determinants of evaluation may arise from the emphasis that has been placed on evaluation knowledge, skills and behaviours,[67,102] and organisational changes,[68,84] rather than indicators of the usefulness of evaluation evidence to inform decision making,[65] or program improvement.[129] However there are exceptions: examples of centralised ECB support for state- and nation-wide tobacco control programs have been described in the USA.[130,131] Each ECB initiative was motivated by the need to improve the evidence available for decision making.[130,131] Based on these examples of jurisdiction-wide efforts, there is potential for the health promotion field to develop ECB strategies that address central funding, policy and administrative arrangements that impact on evaluation.

This thesis reported that dedicated evaluation roles within an organisation, or good connections to university or other skilled evaluators, could facilitate evaluation practice and capacity (Chapter Five).[104] Many authors recommend research-practice partnerships as critical to overcoming the challenges to evaluation.[43,44,56,59] However, surprisingly in this research the presence of an internal evaluation role or working with university evaluators were not independently associated with evaluation quality (as reported in Chapter Eight), although they approached significance. The lack of a clear relationship between access to evaluation expertise and evaluation quality may be due to the variation in the extent and nature of support provided by evaluation experts. There are a number of

examples of university involvement in ECB through provision of single training events, or support during a time-limited project.[131-134] A comprehensive example of university-based support is the Western Australian Sexual Health and Blood-borne Virus Applied Research and Evaluation Network (SiREN). SiREN provides mentoring, training, evaluation assistance and knowledge dissemination to stakeholder organisations in the sexual health and blood borne viruses sector. While the impact of SiREN has not yet been assessed, there are indications of positive changes in the generation and use of evidence.[103] The ongoing success of such initiatives will require partnerships, long term commitment and further engagement with organisations.[103] Differing funding mechanisms and the potentially divergent priorities of research and practice can pose challenges to these partnerships.[135,136]

9.4 Future directions for research

The research in this thesis has made a significant contribution to the ECB literature, potentially to improve evidence-based health promotion and evaluation practice. Future research should explore the distribution of evaluation capacity factors throughout Australia, and compare differences in evaluation capacity strengths between organisations with a wider range of attributes. Similarly, to further tailor ECB initiatives within health promotion, and ensure reported evaluation can contribute to the evidence base for decision making, there is a need to examine which levels of evaluation and methods are frequently overlooked. Given that some challenges to evaluation arise due to the nature of the program population,[72] or type of strategy used,[49] identification of common barriers to evaluation, and gaps in evaluation quality, could further assist in developing targeted ECB initiatives.

To progress ECB in health promotion,[65,68] rigorous assessment of ECB strategies implemented at the practitioner-, organisational-, and system-level in the health promotion field is essential. Further, to underpin the development of ECB strategies that engage partners, access to expertise and development of evaluation skills, there is a need for examination of the characteristics of effective research-practice-policy partnerships that

inform the health promotion evidence base.

9.5 Recommendations for evaluation capacity building

This section outlines recommendations to improve evaluation capacity in health promotion arising from this thesis. These recommendations identify important actions for health promotion organisations, funding bodies and policy makers.

Based on the findings from this thesis, those seeking to implement ECB strategies in health promotion must:

1. Design and implement ECB initiatives that can address practitioner-, organisational- and system-level capacity factors.
2. Address purpose, evaluation utility and quality in the goals of ECB initiatives.
3. Engage with smaller, locally focused health promotion organisations as a priority.
4. Use validated evaluation capacity measurement instruments in the planning and evaluation of ECB initiatives.

To facilitate improvements to evaluation capacity, prevention policy makers and funding administrators must continue to engage with prevention organisations and agencies. It is recommended that:

5. Health promotion program reporting and accountability requirements should be clearly communicated, and agreed between the organisation and the funding body. Agreement should be reached in regards to: A) the capacity of organisation to meet the requirements, and/or provision of additional support and resources to meet the requirements; and, B) the purpose and utility of evaluation, and associated indicators.
6. All mandated program evaluation and reporting should consider the opportunity cost of compliance reporting, and instead focus on evidence that can be used to improve program design, improvement, and demonstration of effectiveness. Attention should

be given to the role of formative evaluation, and for evaluation that can address gaps in the evidence base for programs that address health inequity.

7. Opportunities are created for smaller health promotion organisations to obtain access to evaluation support. This support should provide access to dedicated evaluation resources, including technical advice, mentoring, and evaluation guides. Support must be affordable within existing program budgets, and models that are based around geographic regions, areas of health promotion practice, or funding jurisdiction should be considered.
8. Incentives for organisations with strengths in evaluation, to support organisations with developing evaluation capacity should be implemented.

Health promotion and prevention organisations seeking to improve evaluation capacity must:

9. Invest in health promotion leaders who understand the value of evaluation conduct and use, develop their ability to embed evaluation within organisational systems and cultures, and engage with funders, policy makers and decision makers to improve the conditions for evaluation.
10. Consider the important and interrelated organisational determinants of evaluation capacity, and design ECB strategies to target culture, systems, support, and the allocation of resources within an organisation.
11. Encourage interested health promotion practitioners and managers to take up opportunities for evaluation training in the form of certificate, or post-graduate level qualifications and professional development activities.

Within the health promotion field, collaborative efforts to improve evaluation capacity will require:

12. Investment in ECB approaches and evaluation.

13. Ongoing advocacy to promote the benefits of prevention approaches, and demonstrate and promote the value of using evidence in decision making.

9.6 Conclusion

The aim of the program of research reported in this thesis was to assess the strengths and limitations of evaluation in the health promotion field, and identify opportunities for ECB. Previous studies have identified poor quality evaluation in some health promotion settings, and literature that describes the factors that influence health promotion evaluation practice and capacity and their interactions was found to be limited. No previous research has identified the factors associated with evaluation quality in health promotion. The research in this thesis has contributed to the evaluation capacity literature, by systematically identifying the important determinants of evaluation capacity and quality in the health promotion field. Further, this research has described the interactions between these determinants, and validated a model of evaluation capacity. To improve capacity for health promotion evaluation there is a need to address leadership, culture, systems and supports for evaluation within organisations, and to re-orient health promotion funding and administrative arrangements to facilitate evaluation. These contributions provide guidance to policy and decision makers, funding bodies and health promotion organisations to implement ECB strategies, and ultimately improve the quality and effectiveness of prevention programs in the field.

References

1. World Health Organization. The Bangkok Charter for Health Promotion in a Globalized World 6th Global Conference on Health Promotion; 2005; Bangkok, Thailand Available from:
https://www.who.int/healthpromotion/conferences/6gchp/bangkok_charter/en/.
2. Bloom M, Gullotta TP. Definitions of Primary Prevention. In: Gullotta TP, Bloom M, editors. *Encyclopedia of Primary Prevention and Health Promotion*. Boston, MA: Springer US; 2014. p. 3-12.
3. World Health Organization. Noncommunicable diseases country profiles 2018. Geneva; 2018. Available from: <https://www.who.int/nmh/publications/ncd-profiles-2018/en/>.
4. Crosland P, Ananthapavan J, Davison J, Lambert M, Carter R. The health burden of preventable disease in Australia: a systematic review. *Australian and New Zealand Journal of Public Health*. 2019;Online:1-8 Available from:
<https://onlinelibrary.wiley.com/doi/abs/10.1111/1753-6405.12882>.
5. Australian Institute of Health and Welfare. *Australia's health 2018*. Canberra: AIHW; 2018. Available from: <https://www.aihw.gov.au/reports/australias-health/australias-health-2018/contents/table-of-contents>.
6. Australian Health Ministers' Advisory Council. *Aboriginal and Torres Strait Islander Health Performance Framework 2017 Report*. Canberra: AHMAC; 2017 Available from:
<https://www.aihw.gov.au/reports/indigenous-health-welfare/health-performance-framework/contents/overview>.
7. Australian Institute of Health and Welfare. *Indicators of socioeconomic inequalities in cardiovascular disease, diabetes and chronic kidney disease*. Canberra: AIHW; 2019. Available from: <https://www.aihw.gov.au/reports/chronic-disease/indicators-of-socioeconomic-inequalities-in-cardio/contents/table-of-contents>.
8. Commission on Social Determinants of Health. *Closing the gap in a generation: Health equity through action on the social determinants of health. Final Report of the Commission on Social Determinants of Health*. Geneva: World Health Organization; 2008. Available from: https://www.who.int/social_determinants/thecommission/finalreport/en/.
9. Marmot M. Social determinants of health inequalities. *The Lancet*. 2005;365(9464):1099-104 Available from:
<http://www.sciencedirect.com/science/article/pii/S0140673605711466>.
10. Nutbeam D. Getting evidence into policy and practice to address health inequalities. *Health Promotion International*. 2004;19(2):137-40 Available from:
<https://dx.doi.org/10.1093/heapro/dah201>.
11. Jolley G. Evaluating complex community-based health promotion: Addressing the

- challenges. *Evaluation and Program Planning*. 2014;45:71-81 Available from: <http://www.sciencedirect.com/science/article/pii/S0149718914000317>.
12. Smith JA, Crawford G, Signal L. The case of national health promotion policy in Australia: where to now? *Health Promotion Journal of Australia*. 2016;27(1):61-5 Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1071/HE15055>.
 13. Higgins JW, Strange K, Scarr J, Pennock M, Barr V, Yew A, et al. "It's a feel. That's what a lot of our evidence would consist of ": public health practitioners' perspectives on evidence. *Evaluation and the Health Professions*. 2011;34(3):278-96 Available from: <http://ehp.sagepub.com/content/34/3/278.full.pdf>.
 14. Wimbush E, Watson J. An Evaluation Framework for Health Promotion: Theory, Quality and Effectiveness. *Evaluation*. 2000;6(3):301-21 Available from: <https://journals.sagepub.com/doi/pdf/10.1177/135638900000600302>.
 15. Brownson RC, Fielding JE, Green LW. Building Capacity for Evidence-Based Public Health: Reconciling the Pulls of Practice and the Push of Research. *Annual Review of Public Health*. 2018;39(1):27-53 Available from: <https://www.annualreviews.org/doi/abs/10.1146/annurev-publhealth-040617-014746>.
 16. Pettman TL, Armstrong R, Doyle J, Burford B, Anderson LM, Hillgrove T, et al. Strengthening evaluation to capture the breadth of public health practice: ideal vs. real. *Journal of Public Health*. 2012;34(1):151-5 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22362969>.
 17. Smith BJ. Evaluation of health promotion programs: are we making progress? *Health Promotion Journal of Australia*. 2011;22(3):165 Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1071/HE11165>.
 18. Rychetnik L, Frommer M, Hawe P, Shiell A. Criteria for evaluating evidence on public health interventions. *Journal of Epidemiology and Community Health*. 2002;56(2):119-27 Available from: <https://jech.bmj.com/content/56/2/119>.
 19. Galea S, Vaughan R. Making Decisions That Narrow, or Widen, Health Gaps: A Public Health of Consequence, February 2019. *American Journal of Public Health*. 2019;109(2):196-7 Available from: <https://www.ncbi.nlm.nih.gov/pubmed/30649952>.
 20. White M, Adams J, Heywood P. How and why do interventions that increase health overall widen inequalities within populations. In: Babones S, editor. *Social Inequality and Public Health: ProQuest EBook Central*; 2009.
 21. Wutzke S, Morrice E, Benton M, Milat A, Russell L, Wilson A. Australia's National Partnership Agreement on Preventive Health: Critical reflections from States and Territories. *Health Promotion Journal of Australia*. 2018;29(3):228-35 Available from: <http://dx.doi.org/10.1002/hpja.9>.
 22. Baugh Littlejohns L, Baum F, Lawless A, Freeman T. Disappearing health system building blocks in the health promotion policy context in South Australia (2003–2013). *Critical Public Health*. 2019;29(2):228-40 Available from:

<https://doi.org/10.1080/09581596.2017.1418501>.

23. Jones SC, Donovan RJ. Professional development activities and qualifications of health promotion practitioners in Australia. *Health Promotion Journal of Australia*. 2003;14(1):8-14 Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1071/HE03008>.
24. Jones-Roberts A, Phillips J, Tinsley K. Creating a sustainable health promotion workforce in Australia: a health promoting approach to professionalisation. *Health Promotion Journal of Australia*. 2014;25(2):150-2 Available from: <http://www.publish.csiro.au/he/pdf/HE13076>.
25. Sendall MC. Creating a regulated health promotion workforce in Australia. 15th World Congress on Public Health; 3-7 August; Melbourne 2017 Available from: <https://eprints.qut.edu.au/125522/>.
26. US Department of Health and Human Services Centers for Disease Control and Prevention. Introduction to program evaluation for public health programs: A self-study guide. Atlanta, GA: Office of the Director. Office of Strategy and Innovation. Centers for Disease Control and Prevention 2011. Available from: <https://www.cdc.gov/eval/guide/CDCEvalManual.pdf>.
27. Patton MQ. Utilization-focused evaluation. Thousand Oaks, California: SAGE Publications; 2008.
28. Smith BJ, Rissel C, Shilton T, Bauman A. Advancing evaluation practice in health promotion. *Health Promotion Journal of Australia*. 2016;27(3):184-6 Available from: http://www.publish.csiro.au/he/pdf/HEv27n3_ED2.
29. Sukovic S. HETI Research Blog [Internet]. Research Directorate, editor: Health Education & Training. NSW Ministry of Health. 2017. [cited 2019]. Available from: <http://blog.heti.nsw.gov.au/2017/10/25/evaluation-research-part-2-similarities-differences/>.
30. Patton MQ. Evaluation Flash Cards: Embedding Evaluative Thinking in Organizational Culture. St. Paul, MN: Otto Bremer Trust; 2014 Available from: https://ottobremer.org/news_stories/evaluation-flash-cards/.
31. Bauman AE, Nutbeam D. Evaluation in a nutshell: a practical guide to the evaluation of health promotion programs. 2nd ed. North Ryde, N.S.W.: McGraw-Hill; 2014.
32. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *American Journal of Public Health*. 1999;89(9):1322-7 Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1508772/pdf/amjph00009-0018.pdf>.
33. Springett J. Chapter 4. Participatory approaches to evaluation in health promotion. In: Rootman I GM, et al., editor. Evaluation in health promotion: principles and perspectives Evaluation in Health Promotion: Principles and Perspectives. WHO Reg Publ Eur Ser. Denmark: WHO Regional Publications; 2001. p. 83-105.

34. Nitsch M, Waldherr K, Denk E, Griebler U, Marent B, Forster R. Participation by different stakeholders in participatory evaluation of health promotion: a literature review. *Evaluation and Program Planning*. 2013;40:42-54 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23732440>.
35. Brownson RC, Fielding JE, Maylahn CM. Evidence-based public health: a fundamental concept for public health practice. *Annual Review of Public Health*. 2009;30:175-201 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19296775>.
36. Owen L, Morgan A, Fischer A, Ellis S, Hoy A, Kelly M. The cost-effectiveness of public health interventions. *Journal of Public Health*. 2012;34(1):37-45
37. Wolfenstetter SB. Conceptual Framework for Standard Economic Evaluation of Physical Activity Programs in Primary Prevention. *Prevention Science*. 2011;12(4):435-51 Available from: <https://www.ncbi.nlm.nih.gov/pubmed/21773728>.
38. Hueiming L, Janani M, Jaithri A, Penelope H, Alan S, Stephen J. Exploring the use of economic evidence to inform investment in disease prevention – a qualitative study. *Australian and New Zealand Journal of Public Health*. 2018;42(2):200-6 Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/1753-6405.12748>.
39. Underhill G. Healthy Together Mildura: Evaluation Paper 2012-2016. Mildura: Mildura Rural City Council and Sunraysia Community Health Services; 2016. Available from: http://healthytogethertomildura.com.au/wp-content/uploads/2013/10/HTM-Evaluation-Paper_270616_Final.pdf.
40. Fagen MC, Redman SD, Stacks J, Barrett V, Thullen B, Altenor S, et al. Developmental Evaluation: Building Innovations in Complex Environments. *Health Promotion Practice*. 2011;12(5):645-50 Available from: <http://journals.sagepub.com/doi/abs/10.1177/1524839911412596>.
41. Matheson A, Walton M, Gray R, Lindberg K, Shanthakumar M, Fyfe C, et al. Evaluating a community-based public health intervention using a complex systems approach. *Journal of Public Health*. 2018;40(3):606-13 Available from: <http://dx.doi.org/10.1093/pubmed/fox117>.
42. Green LW. Public health asks of systems science: to advance our evidence-based practice, can you help us get more practice-based evidence? *American Journal of Public Health*. 2006;96(3):406-9 Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1470512/pdf/0960406.pdf>.
43. Hanson S, Jones A. Missed opportunities in the evaluation of public health interventions: a case study of physical activity programmes. *BMC Public Health*. 2017;17(1):674 Available from: <https://doi.org/10.1186/s12889-017-4683-z>.
44. Francis LJ, Smith BJ. Toward Best Practice in Evaluation: A Study of Australian Health Promotion Agencies. *Health Promotion Practice*. 2015;16(5):715-23 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25690614>.
45. International Union of Health Promotion Education. IUHPE core competencies and

- professional standards for health promotion: Full version. International Union of Health Promotion Education; 2016. Available from: http://www.iuhpe.org/images/JC-Accreditation/Core_Competencies_Standards_linkE.pdf.
46. Prevention and Population Health Branch. Evaluation framework for health promotion and disease prevention programs. Melbourne: Victorian Government Department of Health; 2010. Available from: www.health.vic.gov.au/healthpromotion/evidence_evaluation/cdp_tools.htm.
47. NSW Government. NSW Government Evaluation Framework. Sydney: New South Wales Government; 2013. Available from: www.nsw.gov.au/strategic-centre.
48. O'Connor-Fleming ML, Parker E, Higgins H, Gould T. A framework for evaluating health promotion programs. *Health Promotion Journal of Australia*. 2006;17(1):61-6 Available from: <http://www.publish.csiro.au/HE/HE06061>.
49. Rootman I, Goodstadt M, Hyndman B, McQueen DV, Potvin L, Springett J, et al. Evaluation in health promotion: principles and perspectives. Denmark: WHO Regional Publications; 2001.
50. Learmonth AM. Utilizing research in practice and generating evidence from practice. *Health Education Research*. 2000;15(6):743-56 Available from: <https://academic.oup.com/her/article/15/6/743/580634>.
51. South J, Tilford S. Perceptions of research and evaluation in health promotion practice and influences on activity. *Health Education Research*. 2000;15(6):729-41 Available from: <https://academic.oup.com/her/article/15/6/729/580633>.
52. Bowen S, Zwi AB. Pathways to "evidence-informed" policy and practice: a framework for action. *PLoS Medicine*. 2005;2(7):e166 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15913387>.
53. Orton L, Lloyd-Williams F, Taylor-Robinson D, O'Flaherty M, Capewell S. The use of research evidence in public health decision making processes: systematic review. *PloS One*. 2011;6(7):e21704 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21818262>.
54. Dzewaltowski DA, Klesges LM, Estabrooks PA, Glasgow RE, Bull S. Behavior change intervention research in community settings: how generalizable are the results? *Health Promotion International*. 2004;19(2):235-45 Available from: <https://dx.doi.org/10.1093/heapro/dah211>.
55. Li V, Carter SM, Rychetnik L. Evidence valued and used by health promotion practitioners. *Health Education Research*. 2015;30(2):193-205 Available from: <http://her.oxfordjournals.org/content/30/2/193.full.pdf>.
56. Lobo R, Petrich M, Burns SK. Supporting health promotion practitioners to undertake evaluation for program development. *BMC Public Health*. 2014;14(1315):e1-8 Available from: <https://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-14-1315>.

57. Judd J, Frankish CJ, Moulton G. Setting standards in the evaluation of community-based health promotion programmes— a unifying approach. *Health Promotion International*. 2001;16(4):367-80 Available from: <http://dx.doi.org/10.1093/heapro/16.4.367>.
58. Dahler-Larsen P. Evaluation and Public Management. In: Ferlie E, Lynn LE, Pollitt C, editors. *The Oxford Handbook of Public Management*. Oxford Handbooks Online: Oxford University Press; 2009.
59. Brug J, Tak NI, Te Velde SJ. Evaluation of nationwide health promotion campaigns in The Netherlands: an exploration of practices, wishes and opportunities. *Health Promotion International*. 2011;26(2):244-54 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20739324>.
60. Jolley GM, Lawless AP, Baum FE, Hurley CJ, Fry D. Building an evidence base for community health: a review of the quality of program evaluations. *Australian Health Review*. 2007;31(4):603-10 Available from: http://www.publish.csiro.au/?act=view_file&file_id=AH070603.pdf.
61. Leavy JE, Bull FC, Rosenberg M, Bauman A. Physical activity mass media campaigns and their evaluation: a systematic review of the literature 2003-2010. *Health Education Research*. 2011;26(6):1060-85 Available from: <http://her.oxfordjournals.org/content/26/6/1060.full.pdf>.
62. Lim MSC, Wright CJC, Carrotte ER, Pedrana AE. Reach, engagement, and effectiveness: a systematic review of evaluation methodologies used in health promotion via social networking sites. *Health Promotion Journal of Australia*. 2016;27(3):187-97 Available from: <http://www.publish.csiro.au/he/pdf/HE16057>.
63. Chambers AH, Murphy K, Kolbe A. Designs and methods used in published Australian health promotion evaluations 1992-2011. *Australian and New Zealand Journal of Public Health*. 2015;39(3):222-6 Available from: <http://onlinelibrary.wiley.com/store/10.1111/1753-6405.12359/asset/azph12359.pdf?v=1&t=id83rmcz&s=3d40d445ddca2dbe8c6385bdf50832d32f184d66>.
64. Cavill N, Roberts K, Rutter H. *Standard Evaluation Framework for Physical Activity Interventions England: National Obesity Observatory*. Public Health Observatories in England; 2012. Available from: https://webarchive.nationalarchives.gov.uk/20170110171012/https://www.noo.org.uk/core/frameworks/SEF_PA.
65. Norton S, Milat A, Edwards B, Giffin M. Narrative review of strategies by organizations for building evaluation capacity. *Evaluation and Program Planning*. 2016;58:1-19 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27258905>.
66. Suarez-Balcazar Y, Taylor-Ritzler T. Moving from Science to Practice in Evaluation Capacity Building. *American Journal of Evaluation*. 2014;35(1):95-9 Available from: <https://journals.sagepub.com/doi/10.1177/1098214013499440>.

67. Naccarella L, Pirkis J, Kohn F, Morley B, Burgess P, Blashki G. Building evaluation capacity: definitional and practical implications from an Australian case study. *Evaluation and Program Planning*. 2007;30(3):231-6 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17689328>.
68. Labin SN, Duffy JL, Meyers DC, Wandersman A, Lesesne CA. A Research Synthesis of the Evaluation Capacity Building Literature. *American Journal of Evaluation*. 2012;33(3):307-38 Available from: <https://journals.sagepub.com/doi/10.1177/1098214011434608>.
69. Gibbs D, Napp D, Jolly D, Westover B, Uhl G. Increasing Evaluation Capacity within Community-Based HIV Prevention Programs. *Evaluation and Program Planning*. 2002;25(3):261-69 Available from: <https://www.sciencedirect.com/science/article/abs/pii/S0149718902000204>.
70. Stockdill SH, Baizerman M, Compton DW. Toward a Definition of the ECB Process: A Conversation with the ECB Literature. *New Directions for Evaluation*. 2002;2002(93):7-25 Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/ev.39>.
71. Napp D, Gibbs D, Jolly D, Westover B, Uhl G. Evaluation barriers and facilitators among community-based HIV prevention programs. *AIDS Education and Prevention*. 2002;14(3 Suppl A):38-48 Available from: <https://www.ncbi.nlm.nih.gov/pubmed/12092935>.
72. Lobo R, McManus A, Brown G, Hildebrand J, Maycock B. Evaluating peer-based youth programs: barriers and enablers. *Evaluation Journal of Australasia*. 2010;10(1):36-43 Available from: <http://journals.sagepub.com/doi/10.1177/1035719X1001000106>.
73. Klinner C, Carter SM, Rychetnik L, Li V, Daley M, Zask A, et al. Integrating relationship- and research-based approaches in Australian health promotion practice. *Health Promotion International*. 2015;30(4):891-902 Available from: <http://dx.doi.org/10.1093/heapro/dau026>.
74. Huckel Schneider C, Milat AJ, Moore G. Barriers and facilitators to evaluation of health policies and programs: Policymaker and researcher perspectives. *Evaluation and Program Planning*. 2016;58:208-15 Available from: <https://www.sciencedirect.com/science/article/pii/S0149718915301014>.
75. van Koperen TM, Renders CM, Spierings EJM, Hendriks A-M, Westerman MJ, Seidell JC, et al. Recommendations and Improvements for the Evaluation of Integrated Community-Wide Interventions Approaches. *Journal of Obesity*. 2016;2016(e2385698):1-13 Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5220506/pdf/JOBE2016-2385698.pdf>.
76. Bourgeois I, Cousins JB. Understanding Dimensions of Organizational Evaluation Capacity. *American Journal of Evaluation*. 2013;34(3):299-319 Available from: <https://journals.sagepub.com/doi/abs/10.1177/1098214013477235>.
77. Bourgeois I, Simmons L, Hotte N, Osseni R. Measuring Evaluation Capacity in Ontario Public Health Units. *The Canadian Journal of Program Evaluation*. 2016;31(2):1

- Available from: <https://journalhosting.ucalgary.ca/index.php/cjpe/article/view/31074>.
78. Bourgeois I, Toews E, Whynot J, Lamarche MK. Measuring Organizational Evaluation Capacity in the Canadian Federal Government. *The Canadian Journal of Program Evaluation*. 2013;28(2):1-19 Available from: <https://cdm.ucalgary.ca/index.php/cjpe/article/view/30828>.
 79. Bourgeois I, Whynot J, Thériault É. Application of an organizational evaluation capacity self-assessment instrument to different organizations: Similarities and lessons learned. *Evaluation and Program Planning*. 2015;50:47 Available from: <https://www.sciencedirect.com/science/article/pii/S0149718915000063>.
 80. Cousins JB, Elliott C, Amo C, Bourgeois I, Chouinard J, Goh SC, et al. Organizational Capacity to Do and Use Evaluation: Results of a Pan-Canadian Survey of Evaluators. *The Canadian Journal of Program Evaluation*. 2008;23(3):1-35 Available from: <https://evaluationcanada.ca/system/files/cjpe-entries/23-3-001.pdf>.
 81. Cousins JB, Goh SC, Elliott C, Aubry T, Gilbert N. Government and voluntary sector differences in organizational capacity to do and use evaluation. *Evaluation and Program Planning*. 2014;44:1-13 Available from: <https://www.sciencedirect.com/science/article/abs/pii/S0149718913001055>.
 82. Gagnon F, Aubry T, Cousins JB, Goh SC, Elliott C. Validation of the evaluation capacity in organizations questionnaire. *Evaluation and Program Planning*. 2018;68:166-75 Available from: <http://www.sciencedirect.com/science/article/pii/S014971891730263X>.
 83. Preskill H, Boyle S. A Multidisciplinary Model of Evaluation Capacity Building. *American Journal of Evaluation*. 2008;29(4):443-59 Available from: <https://journals.sagepub.com/doi/10.1177/1098214008324182>.
 84. Taylor-Ritzler T, Suarez-Balcazar Y, Garcia-Iriarte E, Henry DB, Balcazar FE. Understanding and Measuring Evaluation Capacity: A Model and Instrument Validation Study. *American Journal of Evaluation*. 2013;34(2):190-206 Available from: <https://journals.sagepub.com/doi/abs/10.1177/1098214012471421>.
 85. Fierro LA, Christie CA. Evaluator and program manager perceptions of evaluation capacity and evaluation practice. *American Journal of Evaluation*. 2017;38(3):376-92 Available from: <https://journals.sagepub.com/doi/abs/10.1177/1098214016667581>.
 86. Nielsen SB, Lemire S, Skov M. Measuring Evaluation Capacity—Results and Implications of a Danish Study. *American Journal of Evaluation*. 2011;32(3):324-44 Available from: <http://aje.sagepub.com/content/32/3/324.abstract>.
 87. Labin SN. Developing Common Measures in Evaluation Capacity Building: An Iterative Science and Practice Process. *American Journal of Evaluation*. 2014;35(1):107-15 Available from: <http://journals.sagepub.com/doi/abs/10.1177/1098214013499965>.
 88. Preskill H. Now for the Hard Stuff: Next Steps in ECB Research and Practice. *American Journal of Evaluation*. 2013;35(1):116-9 Available from: <https://journals.sagepub.com/doi/10.1177/1098214013499439>.

89. Carman JG. Evaluation practice among community-based organizations - Research into the reality. *American Journal of Evaluation*. 2007;28(1):60-75 Available from: <https://journals.sagepub.com/doi/10.1177/1098214006296245>.
90. Carman JG. Nonprofits, Funders, and Evaluation: Accountability in Action. *The American Review of Public Administration*. 2008;39(4):374-90 Available from: <https://doi.org/10.1177/0275074008320190>.
91. Carman JG, Fredericks KA. Evaluation Capacity and Nonprofit Organizations: Is the Glass Half-Empty or Half-Full? *American Journal of Evaluation*. 2009;31(1):84-104 Available from: <http://journals.sagepub.com/doi/10.1177/1098214009352361>.
92. Jackson H, Shiell A. Preventive health: How much does Australia spend and is it enough? Canberra: Foundation for Alcohol Research and Education; 2017. Available from: http://fare.org.au/wp-content/uploads/Preventive-health-How-much-does-Australia-spend-and-is-it-enough_FINAL.pdf.
93. Teddlie C, Tashakkori A. Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences. Thousand Oaks, California: SAGE Publications Inc.; 2009.
94. Curry L, Nunez-Smith M. Mixed Methods in Health Sciences Research: A Practical Primer. Thousand Oaks, California: SAGE Publications, Inc.; 2015. Available from: <http://methods.sagepub.com/book/mixed-methods-in-health-sciences-research-a-practical-primer>.
95. Leech N, Onwuegbuzie A. A typology of mixed methods research designs. *Quality & Quantity*. 2009;43(2):265-75 Available from: <http://dx.doi.org/10.1007/s11135-007-9105-3>.
96. Davis P, Scott A. Health research sampling methods. In: Saks M, Allsop J, editors. *Researching health: Qualitative, quantitative and mixed methods*. London: SAGE Publications Ltd.; 2007.
97. Streiner DL, Norman GR. Health measurement scales a practical guide to their development and use. 4th ed. New York: Oxford University Press; 2008.
98. McCann W. Review of non-hospital based services. Adelaide: Office of the Public Employment and Review: Government of South Australia; 2012 Available from: <https://www.sahealth.sa.gov.au/wps/wcm/connect/45cf08804dab7a169edebed1d6abeab7/nonhospital-report-PHCS-20121203.pdf?MOD=AJPERES&CACHE=NONE&CONTENTCACHE=NONE>.
99. Milat AJ, King L, Bauman AE, Redman S. The concept of scalability: increasing the scale and potential adoption of health promotion interventions into policy and practice. *Health Promotion International*. 2013;28(3):285-98 Available from: <http://www.ingentaconnect.com/content/oxup/09574824/2013/00000028/00000003/art00002>.
100. Datta J, Petticrew M. Challenges to evaluating complex interventions: a content analysis of published papers. *BMC Public Health*. 2013;13(568):e1-18 Available from:

<http://www.ncbi.nlm.nih.gov/pubmed/23758638>.

101. Edwards B, Stickney B, Milat A, Campbell D, Thackway S. Building research and evaluation capacity in population health: the NSW Health approach. *Health Promotion Journal of Australia*. 2016;27(3):264-7 Available from: <http://www.publish.csiro.au/he/pdf/HE16045>.
102. Lindeman PT, Bettin E, Beach LB, Adames CN, Johnson AK, Kern D, et al. Evaluation capacity building—Results and reflections across two years of a multisite empowerment evaluation in an HIV prevention context. *Evaluation and Program Planning*. 2018;71:83-8 Available from: <http://www.sciencedirect.com/science/article/pii/S0149718918301277>.
103. Lobo R, Crawford G, Hallett J, Laing S, Mak DB, Jancey J, et al. A research and evaluation capacity building model in Western Australia. *Health Promotion International*. 2018;33(3):468-78 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28028011>.
104. Schwarzman J, Bauman A, Gabbe B, Rissel C, Shilton T, Smith BJ. Organizational determinants of evaluation practice in Australian prevention agencies. *Health Education Research*. 2018;33(3):243–55 Available from: <http://dx.doi.org/10.1093/her/cyy015>.
105. Schwarzman J, Bauman A, Gabbe BJ, Rissel C, Shilton T, Smith BJ. The Funding, Administrative, and Policy Influences on the Evaluation of Primary Prevention Programs in Australia. *Prevention Science*. 2019:e1-11 Available from: <https://doi.org/10.1007/s11121-019-00997-4>.
106. Baum F, Fisher M. Why behavioural health promotion endures despite its failure to reduce health inequities. *Sociology of Health and Illness*. 2014;36(2):213-25 Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9566.12112>.
107. Carman JG. Understanding Evaluation in Nonprofit Organizations. *Public Performance & Management Review*. 2011;34(3):350-77 Available from: <https://www.tandfonline.com/doi/abs/10.2753/PMR1530-9576340302>.
108. Liverani M, Hawkins B, Parkhurst JO. Political and Institutional Influences on the Use of Evidence in Public Health Policy. A Systematic Review. *PloS One*. 2013;8(10):e77404 Available from: <https://doi.org/10.1371/journal.pone.0077404>.
109. Soderberg E, Wikstrom E. The policy process for health promotion. *Scandinavian Journal of Public Health*. 2015;43(6):606-14 Available from: <https://journals.sagepub.com/doi/abs/10.1177/1403494815586327>.
110. Schwarzman J, Bauman AE, Gabbe BJ, Rissel C, Shilton T, Smith BJ. Understanding the factors that influence health promotion evaluation: the development and validation of the Evaluation Practice Analysis Survey. *Evaluation and Program Planning*. 2019;74:76-83 Available from: <https://www.sciencedirect.com/science/article/pii/S0149718918302635>.
111. McCullough JM. Successful Academic–Public Health Practice Collaboration: What Works From the Public Health Workforce's Perspective. *Journal of Public Health Management and Practice*. 2015;21(6 Supp):S121-S9 Available from:

https://journals.lww.com/jphmp/Fulltext/2015/11001/Successful_Academic_Public_Health_Practice.17.aspx.

112. Harrington D. *Confirmatory Factor Analysis*. Oxford: Oxford University Press; 2009.
113. Kline RB. *Principles and practice of structural equation modeling*. 4th ed. New York: Guilford Press; 2016.
114. Schumacker RE, Lomax RG. *A beginner's guide to structural equation modeling*. 2nd ed. Mahwah N.J.: Lawrence Erlbaum Associates; 2004.
115. Tabachnick BG, Fidell LS. *Using multivariate statistics*. 6th ed. Boston: Pearson Education; 2013.
116. Oxman AD, Bjørndal A, Becerra-Posada F, Gibson M, Block MAG, Haines A, et al. A framework for mandatory impact evaluation to ensure well informed public policy decisions. *The Lancet*. 2010;375(9712):427-31 Available from: [http://dx.doi.org/10.1016/S0140-6736\(09\)61251-4](http://dx.doi.org/10.1016/S0140-6736(09)61251-4).
117. Nichols N, McFarlane K, Gibson P, Millard F, Packer A, McDonald M. Skills, systems and supports: An Aboriginal Community Controlled Health Service (Apunipima) approach to building health promotion evaluation capacity of staff. *Health Promotion Journal of Australia*. 2018;29(1):18-22 Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/hpja.18>.
118. Valenti M, Campetti R, Schoenborn N, Quinlan K, Dash K. Building evaluation capacity of local substance abuse prevention programs serving LGBQ populations. *Evaluation and Program Planning*. 2017;63:101 Available from: <https://www.ncbi.nlm.nih.gov/pubmed/28456016>.
119. Schwarzman J, Bauman A, Gabbe B, Rissel C, Shilton T, Smith BJ. How practitioner, organisational and system-level factors act to influence health promotion evaluation practice: validation of a conceptual framework. *Health Education Research* (under review).
120. Spencer L, Ritchie J, Lewis J, Dillon L. *Quality in Qualitative Evaluation: A framework for assessing research evidence*. London: Cabinet Office; 2003. Available from: <https://www.gov.uk/government/publications/government-social-research-framework-for-assessing-research-evidence>.
121. United Nations Evaluation Group. *UNEG Quality Checklist for Evaluation Reports*. New York: UN Evaluation Group; 2010. Report No.: UNEG/G(2010)2 Available from: <http://www.unevaluation.org/document/detail/607>.
122. Rychetnik L, Frommer M. *A Schema for Evaluating Evidence on Public Health Interventions: Version 4*. Melbourne: National Public Health Partnership; 2002.
123. Brown JL, Kiernan NE. Assessing the subsequent effect of a formative evaluation on a program. *Evaluation and Program Planning*. 2001;24(2):129-43 Available from: <https://www.sciencedirect.com/science/article/pii/S0149718901000040>.

124. Hawe P, Shiell A. Preserving innovation under increasing accountability pressures: the health promotion investment portfolio approach. *Health Promotion Journal of Australia*. 1995;5(2):4-9
125. Carlton EL, Holsinger JW, Riddell MC, Bush H. Full-Range Public Health Leadership, Part 2: Qualitative Analysis and Synthesis. *Frontiers in Public Health*. 2015;3(174):e1-9 Available from: <https://www.frontiersin.org/article/10.3389/fpubh.2015.00174>.
126. Silliman B, Crinion P, Archibald T. Evaluation Champions: What They Do, Why They Do It, and Why It Matters to Organizations. *Journal of Human Sciences and Extension*. 2016;4(3):1
127. Schwarzman J, Nau T, Bauman A, Gabbe B, Rissel C, Shilton T, et al. An assessment of program evaluation methods and quality in Australian prevention agencies. *Health Promotion Journal of Australia* (under review).
128. Carman JG. The Accountability Movement: What's Wrong With This Theory of Change? *Nonprofit and Voluntary Sector Quarterly*. 2009;39(2):256-74 Available from: <https://doi.org/10.1177/0899764008330622>.
129. Bourgeois I, Simmons L, Buetti D. Building evaluation capacity in Ontario's public health units: promising practices and strategies. *Public Health*. 2018;159:89 Available from: <https://www.sciencedirect.com/science/article/pii/S0033350618300428>.
130. Compton DW, MacDonald G, Baizerman M, Schooley M, Zhang L. Using Evaluation Capacity Building (ECB) to Interpret Evaluation Strategy and Practice in the United States National Tobacco Control Program (NTCP): A Preliminary Study. *The Canadian Journal of Program Evaluation*. 2008;23(3):199
131. Treiber J, Cassady D, Kipke R, Kwon N, Satterlund T. Building the evaluation capacity of California's local tobacco control programs. *Health Promotion Practice*. 2011;12(6 Suppl 2):118S-24S Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22068574>.
132. Fleming ML, Easton J. Building environmental educators' evaluation capacity through distance education. *Evaluation and Program Planning*. 2010;33(2):172-7 Available from: <http://www.sciencedirect.com/science/article/pii/S014971890900072X>.
133. McDonald B, Rogers P, Kefford B. Teaching People to Fish? Building the Evaluation Capability of Public Sector Organizations. *Evaluation*. 2003;9(1):9-29 Available from: <https://journals.sagepub.com/doi/abs/10.1177/1356389003009001002>.
134. García-Iriarte E, Suarez-Balcazar Y, Taylor-Ritzler T, Luna M. A Catalyst-for-Change Approach to Evaluation Capacity Building. *American Journal of Evaluation*. 2010;32(2):168-82 Available from: <https://doi.org/10.1177/1098214010387114>.
135. Wright CJC, Schwarzman J, Dietze PM, Crockett B, Lim MSC. Barriers and opportunities in the translation of mobile phone and social media interventions between research and health promotion practice in Australia: a qualitative study of expert perspectives. *Health Res Policy Syst*. 2019;17(1):5 Available from:

<https://doi.org/10.1186/s12961-018-0406-x>.

136. Spoth R, Rohrbach L, Greenberg M, Leaf P, Brown C, Fagan A, et al. Addressing Core Challenges for the Next Generation of Type 2 Translation Research and Systems: The Translation Science to Population Impact (TSci Impact) Framework. *Prevention Science*. 2013;14(4):319-51 Available from: <https://www.ncbi.nlm.nih.gov/pubmed/23430579>.

Appendices

The following pages contain supplementary material published or submitted with the manuscripts in Chapters Five to Eight.

Appendix A. Chapter Five Supplementary Material

A-1. Electronic Supplementary Material

Appendix B. Chapter Six Supplementary Materials

B-1. Final Evaluation Practice Analysis Survey (EPAS): scales and survey items

Appendix C. Chapter Seven Supplementary Material

C-1. EPAS scale distributions, Cronbach's alpha, confirmatory factor analysis results.

C-2. Evaluation Practice Analysis Survey (EPAS) scales and item factor loadings

C-3. Details of model modification

C-4. Final path analysis - direct, indirect and total path coefficients (standardized and unstandardized).

Appendix D. Chapter Eight Supplementary Material

D-1. Quality appraisal tool

Appendix A. Chapter Five Supplementary Material

This appendix contains the semi-structured interview guide used in Phase 1 of data collection, described in Chapter Five. This appendix was published as electronic supplementary material for the paper “The Funding, Administrative, and Policy Influences on the Evaluation of Primary Prevention Programs in Australia”.

A-1. Electronic Supplementary Material

Semi-structured interview guide

1. Can you tell me about the evaluation that your organization has undertaken in the past 12 months?
2. What are the reasons that your organization evaluates health promotion strategies/interventions?
3. What are the things that facilitate good evaluation practice in your organization?
4. What are the barriers to effectively evaluating health promotion programs in your organization?
5. Has your organization undertaken economic evaluation in the past 12 months (or longer)? What affects your ability to do that?
6. You've described your program as [nature of program]. What influence does the nature of the intervention have on the ability to evaluate it?
7. Are there any individual or staff factors that influence the ability to conduct evaluation?
8. What about factors at the organizational level?
9. What are the external influences on your organizations evaluation practice?
10. What are the main strategies that you would recommend to other health promotion organizations looking to improve their evaluation capacity?

Appendix B. Chapter Six Supplementary Materials

This appendix contains the full Evaluation Practice Analysis Survey, including scales, internal consistency coefficients (Cronbach's alpha), items, and item loadings, as developed in Chapter Six. This appendix was published along with the paper "Understanding the factors that influence health promotion evaluation: the development and validation of the Evaluation Practice Analysis Survey".

B-1. Final Evaluation Practice Analysis Survey (EPAS): scales and survey items

Survey domains Scale names	Cronbach's α	Component loading
Evaluation practice		
Self-reported evaluation practice	0.79	
a) Percentage of projects evaluated in the past two years at the formative level		0.74
b) Percentage of projects evaluated in the past two years at the process level		0.79
c) Percentage of projects evaluated in the past two years at the impact level		0.84
d) Percentage of projects evaluated in the past two years at the outcome level		0.75
Individual level		
Attitude and beliefs - Program and staff level benefits <i>I think that evaluation...</i>	0.83	
a) Will help me understand my program		0.83
b) Will inform the decisions I make about my program		0.91
c) Is absolutely necessary to improve my program		0.56
d) Should involve program participants in the evaluation process		0.73
e) Is worth spending time on		0.67
Attitude and beliefs - Organisation and wider benefits <i>I think that evaluation...</i>	0.83	
a) Will influence policy relevant to my program (e.g. stakeholder/partnership, local, state or national policy)		0.87
b) Will help improve programs for people from minority and hard to reach populations		0.78
c) Will justify funding for my program		0.50
d) Will help to convince managers that changes are needed in my program		0.75
e) Will inform changes in our internal processes and documentation		0.66
Skills for different levels of evaluation	0.90	
a) Formative evaluation (i.e. pretesting or piloting)		0.87
b) Process evaluation (i.e. monitoring of implementation)		0.90
c) Impact (i.e. achievement of objectives)		0.94
d) Outcome evaluation (i.e. achievement of project goals)		0.93
e) Economic evaluation (eg. cost benefit analysis)		0.60
Skills for different evaluation tasks	0.95	
a) Clearly stating goals and objectives for my project		0.74
b) Deciding what questions to answer in an evaluation		0.86
c) Choosing designs to evaluate project impacts or outcomes		0.86
d) Developing data collection tools		0.85
e) Identifying from whom to collect the information (i.e. a sampling strategy)		0.83
f) Implementing or appraising ethical standards in evaluation		0.68
g) Analysing quantitative evaluation data		0.79
h) Analysing qualitative evaluation data		0.79
i) Developing recommendations based on evaluation results		0.82
j) Writing an evaluation report		0.85
k) Disseminating findings to relevant stakeholders		0.83
l) Managing an evaluation		0.89

Survey domains	Cronbach's α	Component loading
Scale names		
Leadership for evaluation	0.96	
<i>The leaders in my organisation...</i>		
a) Have a clear vision for evaluation		0.91
b) Demonstrate a good understanding of program evaluation		0.92
c) Are able to build good partnerships with those who can support evaluation when required		0.87
d) Can successfully negotiate evaluation and reporting requirements with our funding source		0.83
e) Ensure our evaluation findings are useful to decision makers		0.90
f) Champion evaluation practice and use internally		0.93
g) Have a strong desire to share our findings		0.86
h) Lack interest in evaluation		0.78
i) Advocate in support of our evaluation efforts when required		0.85
Organisational culture to create the right environment	0.87	
<i>My organisation...</i>		
a) Encourages a learning environment		0.77
b) Is an innovative organisation		0.76
c) Uses evidence from evaluations to inform decision making		0.69
d) Shares evaluation findings with other organisations		0.74
e) Has a system that allows us to learn successful practices from other organisations		0.65
f) Is worried about evaluation findings making the organisation look bad		0.69
Organisational culture for systems and structures	0.91	
a) Evaluation is an important part of our organisational culture		0.81
b) Evaluation of our programs is part of our core business		0.92
c) Everyone in the organisation has some responsibility to contribute to evaluation		0.83
<i>My organisation...</i>		
d) Embeds evaluation in everything we do		0.67
e) Uses evaluation findings to inform the next phase of work		0.58
f) Actively provides staff with opportunities to upskill in the area of evaluation		0.54
Systems and structures	0.85	
a) It is clear from my job description what is expected of me in terms of evaluation		0.74
b) We have an evaluation framework in place for the health promotion work that we do		0.82
c) We have a user friendly way to store and manage our program data (electronic, or paper based)		0.63
d) It is easy to access evaluation expertise from within the organisation		0.75
e) There is a process for seeking advice from external evaluation experts if needed		0.71
f) We have regular meetings to facilitate shared understanding regarding the evaluation plan		0.77
g) Our organisation has clear expectations and timeframes for reporting		0.68
Internal support	0.88	
a) There is adequate support within my organisation for the evaluation I'm involved in		0.90
b) My organisation arranges structured support and training for evaluation as required		0.82
c) Management is happy to support evaluation and allows time and resources for it		0.94
d) We have a strong working relationship with people who can support our team in evaluation		0.71
External support	0.77	
a) I am comfortable contacting people outside of my organisation for evaluation support and advice		0.94
b) Formal arrangements with other organisations support our evaluation practice		0.66
c) I get evaluation support from a network of practitioners I am involved with		0.74

Survey domains	Cronbach's α	Component loading
Scale names		
a) We have access to data collection instruments appropriate to our work		0.75
b) We have a functioning database to collate information		0.87
c) We have appropriate tools for evaluation involving hard to reach and vulnerable groups		0.82
d) Our organisation provides us with the basic resources to support evaluation (e.g., computers, software, copying, administrative support)		0.78
e) We can access a range of good templates for reporting on our evaluation		0.70
Time for evaluation	0.86	
a) We have time to engage with our community and stakeholders for evaluation		0.60
b) We have adequate time for evaluation planning		0.73
c) We have enough time to think and reflect on our work		0.97
d) We have time to write up our findings to share them		0.92
Resources dedicated to evaluation	0.85	
a) Our organisation has long-term, dedicated financial support for our evaluation activities		0.81
b) Evaluation funding is set aside as part of every program budget		0.93
c) Even when there are budget constraints we still allocate funds to evaluation		0.91
d) Staff are encouraged to spend time on program evaluation		0.51
Evaluation role in the organisation	0.86	
<i>The evaluation person or team...</i>		
a) Champions evaluation of health promotion projects in our organisation.		0.82
b) Supports the health promotion program staff to undertake elements of the evaluation		0.89
c) Mentors staff members in evaluation practice		0.75
d) Is involved in helping to plan and design our evaluations right from the beginning		0.76
Partnerships	0.86	
a) Partnerships allow us to attract additional resources for evaluation		0.77
b) Involving our partners in evaluation provides additional skills for evaluation		0.81
c) Our partners help to facilitate evaluation involving hard to reach population groups		0.83
d) Our partnerships are valuable for obtaining data needed in our evaluation		0.81
e) There are clear expectations about evaluation between our partners		0.62
f) Involving our partners in evaluation of the programs is worth the effort		0.74
External Evaluators	0.73	
<i>We engage external evaluators...</i>		
a) To ensure the findings are independent		0.69
b) To ensure a high quality of evaluation		0.80
c) When we don't have the required skills internally		0.87
University partnerships	0.75	
a) We work with university partners to help obtain funds for evaluation		0.81
b) Our university partners understand how we work		0.84
c) We have the opportunity to build evaluation capacity through our contacts with university partners		0.72
System level		
Adequate funding for evaluation	0.94	
<i>We have sufficient funding for...</i>		
a) Health promotion evaluation (overall)		0.91
b) Formative evaluation (i.e. pretesting and piloting)		0.88
c) Process evaluation (i.e. monitoring of implementation)		0.90
d) Impact and/or outcome evaluation (i.e. achievement of objectives)		0.92
e) Outcome evaluation (i.e. achievement of project goals)		0.89
Clear and realistic expectations	0.78	
a) There are clear evaluation requirements from our funding body		0.78
b) There are clear expectations from our funding body about the level of investment required in evaluation		0.61
c) The evaluation activities I engage in are consistent with funders' expectations		0.76
d) The funding body has realistic expectations for evaluation given available resources		0.90

Survey domains	Cronbach's α	Component loading
Scale names		
Control and flexibility	0.66	
a) We have good flexibility from our funding body to evaluate the way we think is most appropriate		0.75
b) We have a high level of control over the indicators chosen to evaluate and report on		0.80
c) There is sufficient time available to develop an evaluation plan that meets stakeholder requirements		0.77
Competing demands	0.77	
a) It is challenging to meet funding body requirements for evaluation due to the complex nature of our programs		0.72
b) There is pressure to deliver the programs at the expense of quality evaluation		0.80
c) Evaluation is cut or reduced when the budget is tight overall		0.71
d) Our funding body is generally more concerned with health service delivery than health promotion or prevention		0.62
e) There is tension between what is important for us to evaluate locally and the funding body's priorities for evaluation		0.75
Organisational influence in the system	0.74	
a) We've built political support for our health promotion or disease prevention work to continue and grow		0.68
b) We go to extra lengths to secure funding for evaluation beyond the usual program funding		0.80
c) We distil our evaluation findings into key messages that help us engage politicians		0.66
d) We have had success in securing additional funds for program evaluation		0.83
Learning and sharing	0.76	
a) Our funding body communicates clearly about how they will use our evaluation		0.77
b) Our funding body encourages us to share our evaluation findings		0.77
c) Our funding body provides leadership for conducting evaluation		0.78
d) The funding body supports learning from evaluation findings		0.73
Political decision making	0.63	
a) Changes in funding priorities have impacted on the evaluation of our programs		0.53
b) Despite having evaluation findings available, decisions appear to be made based on political factors		0.58
c) Despite demonstrating effectiveness, projects have not been refunded		0.51
d) It's more important for us to evaluate using rigorous methods in the current political climate		0.64
e) Evaluation findings that include an economic component are used more often by decision makers		0.79
Reporting facilitates evaluation	0.78	
a) I am confident to use the reporting templates that we have been provided with		0.74
b) Evaluation is expected as part of the funding arrangements		0.83
c) Our reporting requirements and key performance indicators drive our evaluation work		0.78
d) Our funding body provides a reporting template		0.76

Appendix C. Chapter Seven Supplementary Material

This appendix contains supplementary material submitted for online-only publication with the manuscript in Chapter Seven, “How practitioner, organisational and system-level factors act to influence health promotion evaluation capacity: validation of a conceptual framework”. The following pages include supplementary results tables (C-1, C-4), data collection instrument (C-2) and additional details of the model modification in the validation of the health promotion evaluation capacity framework using path analysis (C-3).

C-1. EPAS scale distributions, Cronbach's alpha, confirmatory factor analysis results.

Scales	Items / scale	N	Initial alpha	Missing (mean%/scale)	Modification (covariance added)	Df Final	Chi 2	P	RMSEA (90% CI)	CFI	SRMR	Final scale Mean	Final scale SD
Attitude and beliefs - Program	5	219	0.89	0	3	2	2.05	0.36	0.01 (0.00 – 0.14)	1.00	0.01	5.64	0.53
Attitude and beliefs - Wider	5	219	0.87	0	n/a	5	9.22	0.10	0.06 (0.00 – 0.12)	0.99	0.02	5.24	0.68
Skills and knowledge - Levels**	4	219	0.92	0	1	1	2.79	0.10	0.09 (0.00 – 0.22)	1.00	0.01	4.33	1.03
Skills and knowledge- Tasks	12	219	0.94	0	10	44	104.94	0.00	0.08 (0.06 – 0.10)	0.97	0.04	4.31	0.83
Leadership	9	219	0.95	0	2	25	55.48	0.00	0.08 (0.05 – 0.10)	0.98	0.02	4.56	0.88
Organisational culture - right environment*	5	219	0.86	0	2	3	3.89	0.27	0.04 (0.00 – 0.13)	1.00	0.01	4.69	0.76
Organisational culture - for systems	6	219	0.90	0	2	7	7.91	0.34	0.02 (0.00 – 0.09)	1.00	0.02	4.54	0.88
Systems and structures	7	219	0.86	0	n/a	14	29.74	0.01	0.07 (0.04 – 0.11)	0.97	0.03	4.31	0.88
Partnerships	6	219	0.88	0	2	7	9.55	0.22	0.04 (0.00 – 0.10)	1.00	0.02	4.77	0.77
Resources – Time	4	217	0.91	1	1	1	0.02	0.88	0.00 (0.00 – 0.09)	1.00	0.00	4.37	0.93
Resources – Dedicated to evaluation	4	217	0.87	1	1	1	0.33	0.57	0.00 (0.00 – 0.15)	1.00	0.01	3.99	1.08
Tools	5	217	0.79	1	1	4	6.30	0.18	0.05 (0.00 – 0.12)	0.99	0.02	4.32	0.75
Support - Internal	4	219	0.88	0	n/a	2	2.14	0.34	0.02 (0.00 – 0.14)	1.00	0.01	4.37	0.95
Support - External	3	219	0.78	0	n/a	0	0.00	n/a	0.00 (0.00 – 0.00)	1.00	0.00	4.28	1.07
Clear expectations	4	203	0.85	5.1	1	1	1.15	0.28	0.03 (0.00 – 0.19)	1.00	0.01	4.34	0.98
Competing demands	5	175	0.82	8.9	2	3	4.60	0.20	0.06 (0.00 – 0.15)	1.00	0.02	3.52	1.01
Funding for evaluation	5	212	0.93	3	3	2	3.26	0.20	0.06 (0.00 – 0.16)	1.00	0.01	3.89	1.12
Learn and share	4	196	0.87	7.5	1	1	0.27	0.61	0.00 (0.00 – 0.15)	1.00	0.00	3.82	1.08
Organisational influence	4	190	0.81	5.8	1	1	3.54	0.06	0.12 (0.00 – 0.26)	0.99	0.02	3.78	0.98
Control and flexibility	3	200	0.81	7	n/a	0	0.00	n/a	0.00 (0.00 – 0.00)	1.00	0.00	4.12	0.93
Political decision making*	4	179	0.74	6.5	n/a	2	1.36	0.51	0.00 (0.00 – 0.13)	1.00	0.01	3.39	0.69
Reporting facilitates evaluation	4	188	0.79	7.5	1	1	0.01	0.90	0.00 (0.00 – 0.08)	1.00	0.00	4.66	0.94
Evaluation practice	5	216	0.74	1	n/a	5	69.04	0.00	0.24 (0.19 – 0.30)	0.77	0.09	67.75	27.17

SD=Standard deviation, df=degrees of freedom, RMSEA=root mean square error of approximation, CFI=comparative fit index, SRMR=standardised root mean square residual.

*Item deleted from scale prior to CFA due to poor correlation (Item total R<0.2, or R<0.2 for >50% of pairs)

** Item deleted due to poor factor loading (in prelim data check)

C-2. Evaluation Practice Analysis Survey scales and item factor loadings

Evaluation capacity constructs Scale names	Item factor loading
Evaluation practice	
Self-reported evaluation practice (NB. Scale did not fit the data)	
e) Percentage of projects evaluated in the past two years at the formative level	0.46
f) Percentage of projects evaluated in the past two years at the process level	0.63
g) Percentage of projects evaluated in the past two years at the impact level	0.89
h) Percentage of projects evaluated in the past two years at the outcome level	0.59
i) Percentage of projects evaluated in the past two years at the economic level	0.38
Attitude and beliefs	
Program and staff level benefits	
<i>I think that evaluation...</i>	
f) Will help me understand my program	0.77
g) Will inform the decisions I make about my program	0.81
h) Is absolutely necessary to improve my program	0.80
i) Should involve program participants in the evaluation process	0.66
j) Is worth spending time on	0.83
Organisation and wider benefits	
<i>I think that evaluation...</i>	
f) Will influence policy relevant to my program (e.g. stakeholder/partnership, local, state or national policy)	0.78
g) Will help improve programs for people from minority and hard to reach populations	0.72
h) Will justify funding for my program	0.76
i) Will help to convince managers that changes are needed in my program	0.78
j) Will inform changes in our internal processes and documentation	0.79
Skills and knowledge	
Skills and knowledge for different levels of evaluation	
f) Formative evaluation (i.e. pretesting or piloting)	0.76
g) Process evaluation (i.e. monitoring of implementation)	0.86
h) Impact (i.e. achievement of objectives)	0.97
i) Outcome evaluation (i.e. achievement of project goals)	0.85
Skills and knowledge for different evaluation tasks	
m) Clearly stating goals and objectives for my project	0.58
n) Deciding what questions to answer in an evaluation	0.66
o) Choosing designs to evaluate project impacts or outcomes	0.73
p) Developing data collection tools	0.77
q) Identifying from whom to collect the information (i.e. a sampling strategy)	0.68
r) Implementing or appraising ethical standards in evaluation	0.75
s) Analysing quantitative evaluation data	0.74
t) Analysing qualitative evaluation data	0.81
u) Developing recommendations based on evaluation results	0.83
v) Writing an evaluation report	0.84
w) Disseminating findings to relevant stakeholders	0.71
x) Managing an evaluation	0.84
Organisational leadership and culture	
Leadership for evaluation	
<i>The leaders in my organisation...</i>	
j) Have a clear vision for evaluation	0.81
k) Demonstrate a good understanding of program evaluation	0.87
l) Are able to build good partnerships with those who can support evaluation when required	0.82
m) Can successfully negotiate evaluation and reporting requirements with our funding source	0.78
n) Ensure our evaluation findings are useful to decision makers	0.86
o) Champion evaluation practice and use internally	0.89
p) Have a strong desire to share our findings	0.79
q) Lack interest in evaluation	0.77
r) Advocate in support of our evaluation efforts when required	0.77

Evaluation capacity constructs	Item factor loading
Scale names	
Organisational culture to create the right environment	
<i>My organisation...</i>	
g) Encourages a learning environment	0.62
h) Is an innovative organisation	0.71
i) Uses evidence from evaluations to inform decision making	0.85
j) Shares evaluation findings with other organisations	0.71
k) Has a system that allows us to learn successful practices from other organisations	0.69
Organisational influence in the system	
e) We've built political support for our health promotion or disease prevention work to continue and grow	0.72
f) We go to extra lengths to secure funding for evaluation beyond the usual program funding	0.69
g) We distil our evaluation findings into key messages that help us engage politicians	0.66
h) We have had success in securing additional funds for program evaluation	0.93
Partnerships	
g) Partnerships allow us to attract additional resources for evaluation	0.66
h) Involving our partners in evaluation provides additional skills for evaluation	0.75
i) Our partners help to facilitate evaluation involving hard to reach population groups	0.75
j) Our partnerships are valuable for obtaining data needed in our evaluation	0.76
k) There are clear expectations about evaluation between our partners	0.61
l) Involving our partners in evaluation of the programs is worth the effort	0.79
Organisational systems and support	
Organisational culture for systems and structures	
g) Evaluation is an important part of our organisational culture	0.89
h) Evaluation of our programs is part of our core business	0.82
i) Everyone in the organisation has some responsibility to contribute to evaluation	0.67
<i>My organisation...</i>	
j) Embeds evaluation in everything we do	0.86
k) Uses evaluation findings to inform the next phase of work	0.75
l) Actively provides staff with opportunities to upskill in the area of evaluation	0.59
Internal support	
e) There is adequate support within my organisation for the evaluation I'm involved in	0.86
f) My organisation arranges structured support and training for evaluation as required	0.77
g) Management is happy to support evaluation and allows time and resources for it	0.81
h) We have a strong working relationship with people who can support our team in evaluation	0.79
Systems and structures	
h) It is clear from my job description what is expected of me in terms of evaluation	0.66
i) We have an evaluation framework in place for the health promotion work that we do	0.76
j) We have a user friendly way to store and manage our program data (electronic, or paper based)	0.58
k) It is easy to access evaluation expertise from within the organisation	0.67
l) There is a process for seeking advice from external evaluation experts if needed	0.69
m) We have regular meetings to facilitate shared understanding regarding the evaluation plan	0.77
n) Our organisation has clear expectations and timeframes for reporting	0.63
Tools for evaluation	
f) We have access to data collection instruments appropriate to our work	0.70
g) We have a functioning database to collate information	0.75
h) We have appropriate tools for evaluation involving hard to reach and vulnerable groups	0.69
i) Our organisation provides us with the basic resources to support evaluation (e.g., computers, software, copying, administrative support)	0.50
j) We can access a range of good templates for reporting on our evaluation	0.73

Evaluation capacity constructs	Item factor loading
Scale names	
Evaluation resources	
Resources dedicated to evaluation	
e) Our organisation has long-term, dedicated financial support for our evaluation activities	0.79
f) Evaluation funding is set aside as part of every program budget	0.84
g) Even when there are budget constraints we still allocate funds to evaluation	0.97
h) Staff are encouraged to spend time on program evaluation	0.62
Time for evaluation	
e) We have time to engage with our community and stakeholders for evaluation	0.73
f) We have adequate time for evaluation planning	0.85
g) We have enough time to think and reflect on our work	0.93
h) We have time to write up our findings to share them	0.82
Political and funding environment	
Adequate funding for evaluation	
<i>We have sufficient funding for...</i>	
f) Health promotion evaluation (overall)	0.92
g) Formative evaluation (i.e. pretesting and piloting)	0.81
h) Process evaluation (i.e. monitoring of implementation)	0.82
i) Impact evaluation (i.e. achievement of objectives)	0.85
j) Outcome evaluation (i.e. achievement of project goals)	0.75
Competing demands	
f) It is challenging to meet funding body requirements for evaluation due to the complex nature of our programs	0.57
g) There is pressure to deliver the programs at the expense of quality evaluation	0.79
h) Evaluation is cut or reduced when the budget is tight overall	0.74
i) Our funding body is generally more concerned with health service delivery than health promotion or prevention	0.72
j) There is tension between what is important for us to evaluate locally and the funding body's priorities for evaluation	0.68
Political decision making	
f) Changes in funding priorities have impacted on the evaluation of our programs	0.60
g) Despite having evaluation findings available, decisions appear to be made based on political factors	0.76
h) Despite demonstrating effectiveness, projects have not been refunded	0.68
i) Evaluation findings that include an economic component are used more often by decision makers	-0.52
Control and flexibility	
d) We have good flexibility from our funding body to evaluate the way we think is most appropriate	0.79
e) We have a high level of control over the indicators chosen to evaluate and report on	0.82
f) There is sufficient time available to develop an evaluation plan that meets stakeholder requirements	0.68
Funding body expectations	
Clear and realistic expectations	
e) There are clear evaluation requirements from our funding body	0.71
f) There are clear expectations from our funding body about the level of investment required in evaluation	0.63
g) The evaluation activities I engage in are consistent with funders' expectations	0.89
h) The funding body has realistic expectations for evaluation given available resources	0.72
Reporting facilitates evaluation	
e) I am confident to use the reporting templates that we have been provided with	0.64
f) Evaluation is expected as part of the funding arrangements	0.71
g) Our reporting requirements and key performance indicators drive our evaluation work	0.81
h) Our funding body provides a reporting template	0.58

Evaluation capacity constructs	Item factor loading
Scale names	
Promote learning and sharing	
e) Our funding body communicates clearly about how they will use our evaluation	0.78
f) Our funding body encourages us to share our evaluation findings	0.72
g) Our funding body provides leadership for conducting evaluation	0.91
h) The funding body supports learning from evaluation findings	0.80
Additional scales (not included in final path model)	
External support	
d) I am comfortable contacting people outside of my organisation for evaluation support and advice	0.79
e) Formal arrangements with other organisations support our evaluation practice	0.73
f) I get evaluation support from a network of practitioners I am involved with	0.68
Evaluation role in the organisation	
<i>The evaluation person or team...</i>	
e) Champions evaluation of health promotion projects in our organisation.	n/a
f) Supports the health promotion program staff to undertake elements of the evaluation	n/a
g) Mentors staff members in evaluation practice	n/a
h) Is involved in helping to plan and design our evaluations right from the beginning	n/a
External Evaluators	
<i>We engage external evaluators...</i>	
d) To ensure the findings are independent	n/a
e) To ensure a high quality of evaluation	n/a
f) When we don't have the required skills internally	n/a
University partnerships	
d) We work with university partners to help obtain funds for evaluation	n/a
e) Our university partners understand how we work	n/a
f) We have the opportunity to build evaluation capacity through our contacts with university partners	n/a

C-3. Details of model modification

This appendix provides additional detail of baseline and interim model fit statistics. Detailed descriptions of modifications made at each stage are provided, along with the theoretical justification for the changes.

Baseline model estimation and fit

The baseline model of the factors that influence evaluation practice was estimated using MLE, with a sample of $n=216$. The baseline model did not fit the data well: $\chi^2_{14} = 155.42$, $p < 0.001$, RMSEA = 0.26, 90% CI (0.19-0.25), CFI = 0.87, SRMR = 0.11. Four relationships within the model were not significant: Skills and Knowledge -> Evaluation Practice ($p=0.38$), Organisational Systems and Support -> Evaluation Practice ($p=0.93$), Funding and Reporting -> Evaluation Practice ($p=0.11$) and Organisational Leadership -> Funding Body Expectations ($p=0.06$).

Modifications to baseline model

The three clearly non-significant parameters were deleted (Organisational Leadership --> Funding Body Expectations was retained), and three parameters were added. Firstly, the relationship between Funding body expectations -> Organisational systems and support highlights the role of clear, relevant and agreed upon reporting and evaluation requirements that can facilitate the establishment of data collection and reporting systems for within an organisations.[44,91] Second, a direct relationship between Organisational leadership -> Evaluation resources is clearly supported as leaders supportive of evaluation, particularly those with budgetary responsibility, allocate funding and staff time to evaluation.[104] Finally, the direct pathway between Political and funding environment -> Evaluation practice was considered justifiable in the health promotion context despite limited evidence from the field.[105] Additional support for this relationship can be found in a systematic review of political influences upon evidence use in policy, where the authors found the political systems, framing and values are often overlooked, despite playing a key role in the

use of evidence in decision making.[108]

Model 2 estimation and fit

The second model was estimated using the same procedures. All fit criteria improved, although the χ^2 value remained significant: $\chi^2_{14} = 25.13$, $p = 0.03$, RMSEA = 0.06 (90%CI 0.03-0.10), CFI = 0.99; SRMR = 0.05. Again, the model was assessed for non-significant paths, and modification indices reviewed for potential additions. The modifications indices suggested the direction of the parameter was changed to express Skills and Knowledge -> Attitudes and Beliefs.

Modifications to model 2

The final model incorporated the suggested changes to model 2, specifically the parameter direction between Attitudes and beliefs about evaluation -> Skills and knowledge. This change in this relationship highlights the potential for those interested in evaluation may personally seek out skill development in evaluation.[104]

C-4. Final path analysis - direct, indirect and total path coefficients (standardized and unstandardized).

Structural path	Direct effects				Indirect effects				Total effects			
	β	S.E.	P>z	Std. β	β	S.E.	P>z	Std. β	β	S.E.	P>z	Std. β
Skills and knowledge												
Organisational resources	0.60	0.08	0.00	0.46	(no path)				0.60	0.08	0.00	0.46
Organisational systems and support	(no path)				0.34	0.02	0.00	0.35	0.34	0.02	0.00	0.35
Organisational leadership	(no path)				0.43	0.03	0.00	0.33	0.43	0.03	0.00	0.33
Funding body expectations	(no path)				0.12	0.03	0.00	0.10	0.12	0.03	0.00	0.10
Political and funding environment	(no path)				0.30	0.05	0.00	0.23	0.30	0.05	0.00	0.23
Attitudes and beliefs												
Skills and Knowledge	0.10	0.03	0.00	0.24	(no path)				0.10	0.03	0.00	0.24
Organisational resources	(no path)				0.06	0.01	0.00	0.11	0.06	0.01	0.00	0.11
Organisational systems and support	(no path)				0.03	0.00	0.00	0.09	0.03	0.00	0.00	0.09
Organisational leadership	0.15	0.03	0.00	0.29	0.04	0.00	0.00	0.08	0.19	0.03	0.00	0.37
Funding body expectations	(no path)				0.01	0.00	0.00	0.02	0.01	0.00	0.00	0.02
Political and funding environment	(no path)				0.09	0.02	0.00	0.18	0.09	0.02	0.00	0.18
Organisational resources												
Organisational systems and support	0.57	0.03	0.00	0.77	(no path)				0.57	0.03	0.00	0.77
Organisational leadership	0.24	0.03	0.00	0.23	0.48	0.03	0.00	0.48	0.72	0.05	0.00	0.72
Funding body expectations	-0.08	0.03	0.01	-0.08	0.27	0.03	0.00	0.29	0.19	0.04	0.00	0.21
Political and funding environment	0.07	0.03	0.02	0.07	0.43	0.05	0.00	0.44	0.50	0.05	0.00	0.51
Organisational systems and support												
Organisational leadership	0.81	0.06	0.00	0.60	0.05	0.03	0.06	0.04	0.87	0.06	0.00	0.64
Funding body expectations	0.48	0.05	0.00	0.38	(no path)				0.48	0.05	0.00	0.38
Political and funding environment	(no path)				0.68	0.07	0.00	0.51	0.68	0.07	0.00	0.51
Organisational leadership												
Political and funding environment	0.42	0.06	0.00	0.42	(no path)				0.42	0.06	0.00	0.42
Funding body expectations												
Organisational Leadership	0.11	0.06	0.06	0.10	(no path)				0.11	0.06	0.06	0.10
Political and funding environment	0.67	0.06	0.00	0.62	0.05	0.03	0.07	0.04	0.72	0.05	0.00	0.67
Reported impact evaluation												
Organisational resources	11.40	2.66	0.00	0.29	(no path)				11.40	2.66	0.00	0.29
Organisational systems and support	(no path)				6.48	0.31	0.00	0.22	6.48	0.31	0.00	0.22
Organisational leadership	(no path)				8.19	0.54	0.00	0.21	8.19	0.54	0.00	0.21
Funding body expectations	(no path)				2.22	0.48	0.00	0.06	2.22	0.48	0.00	0.06
Political and funding environment	11.26	2.64	0.00	0.29	5.72	1.47	0.00	0.15	16.98	2.37	0.00	0.44

β = unstandardized coefficient, SE = standard error, Std. β = standardised coefficient

Appendix D. Chapter Eight Supplementary Material

This appendix contains the quality appraisal component of the audit and appraisal tool used in the audit and appraisal of 392 evaluation reports as described in Chapter Eight. This appendix was submitted with the manuscript “An assessment of program evaluation methods and quality in Australian prevention agencies”.

D-1. Quality appraisal tool

Quality items	Scoring
1. Objectives and/or evaluation questions clearly stated	0 = No; 1 = Yes
2. Purpose of evaluation clearly stated	0 = No; 1 = Yes
3. Organisation and population context are clearly described	0 = No; 1 = Yes
4. Target group specified	0 = No; 1 = Yes
5. The logic model and/or model of action described clearly	0 = No; 1 = Yes
6. Intervention elements and procedures described clearly	0 = No; 1 = Yes
7. Evaluation design is clear and is defensible in relation to the purpose and questions.	0 = No; 1 = Yes
8. Participant inclusion/ exclusion criteria <i>or</i> purposive sampling characteristics are stated and defensible (qualitative)	0 = No; 1 = Yes; NA
9. Methods of sampling and recruiting participants have low risk of bias	0 = No; 1 = Yes (<i>minimum</i>); 2 = Yes (<i>high standard</i>); N/A
10. Sample size adequate for methods used <i>or</i> how well the sample met coverage of the topic (qualitative)	0 = No; 1 = Yes (<i>minimum</i>); 2 = Yes (<i>high standard</i>); N/A
11. Procedures for ethical conduct of the evaluation are described	0 = No; 1 = Yes
12. The questions included in data collection tools are stated	0 = No; 1 = Yes; N/A
13. Appropriate attention to reliability and validity of measures <i>or</i> rigour of qualitative data collection procedures	0 = No; 1 = Yes (<i>minimum</i>); 2 = Yes (<i>high standard</i>)
14. Analysis methods are stated	0 = No; 1 = Yes (<i>minimum</i>); 2 = Yes (<i>high standard</i>)
15. Steps taken to reduce confounding and/or bias in data	0 = No; 1 = Yes (<i>minimum</i>); 2 = Yes (<i>high standard</i>)
16. Response rates and participant characteristics given	0 = No; 1 = Yes; N/A
17. Strategy delivery reported	0 = No; 1 = Yes; N/A
18. Strategy reach reported	0 = No; 1 = Yes; N/A
19. Systematic and appropriate analysis and reporting of the data	0 = No; 1 = Yes (<i>minimum</i>); 2 = Yes (<i>high standard</i>)
20. Conclusions supported by data and relate to evaluation questions	0 = No; 1 = Yes (<i>minimum</i>); 2 = Yes (<i>high standard</i>)
21. Generalisability to other populations and settings/context explicitly discussed	0 = No; 1 = Yes
22. Gaps and limitations and/or unanticipated findings are discussed	0 = No; 1 = Yes (<i>minimum</i>); 2 = Yes (<i>high standard</i>)