

# Generalising from qualitative evaluation

Evaluation Journal of Australasia

2021, Vol. 21(1) 7–23

© The Author(s) 2021

Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/1035719X21993938

[journals.sagepub.com/home/evj](https://journals.sagepub.com/home/evj)**John Guenther** 

Batchelor Institute of Indigenous Tertiary Education, Australia

**Ian H Falk** 

Universitas Mahasaraswati Denpasar, Indonesia; Universitas Andi Djemma, Indonesia

## Abstract

Evaluations are often focused on assessing merit, value, outcome or some other feature of a programme, project, policy or some other object. Evaluation research is then more concerned with the particular rather than the general – even more so, when qualitative methods are used. But does this mean that evaluations should not be used to generalise? If it is possible to generalise from evaluations, under what circumstances can this be legitimately achieved? The authors of this article have previously argued for generalising from qualitative research (GQR), and in this article, they extrapolate the discussion to the field of evaluation. First, the article begins with a discussion of the definitions of generalisability in research, recapping briefly on our arguments for GQR. Second, the differentiation between research and evaluation is explored with consideration of what literature there is to justify generalisation from qualitative evaluation (GQE). Third, a typology derived from the literature is developed, to sort 54 evaluation projects. Fourth, material from a suite of evaluation projects is drawn from to demonstrate how the typology of generalisation applies in the context of evaluations conducted in several fields of study. Finally, we suggest a model for GQE.

## Keywords

evaluation outcomes, generalising from qualitative evaluation (GQE), generalising from qualitative research (GQR), evaluation, model for GQE, qualitative evaluation

---

## Corresponding author:

John Guenther, Batchelor Institute of Indigenous Tertiary Education, P.O. Box 2363, Parap, NT 0804 Darwin, NT, Australia.

Email: [john.guenther@batchelor.edu.au](mailto:john.guenther@batchelor.edu.au)

## Introduction

Generalising from qualitative research (GQR) has been an abiding interest of the authors throughout their careers (Falk & Guenther, 2006; Guenther & Falk, 2019a, 2019b). Of all the projects, 54 have been evaluations in a variety of contexts across Australia and Indonesia (Arnott et al., 2012; Falk et al., 2006; Guenther et al., 2009). Commissioners invariably want answers to questions, but a sub-text of their enquiries is whether these results can be applied in other places, situations or times for the purposes of upscaling or replicating programmes (see Mertens & Wilson, 2018, Chapter 8). The issues we have researched have spanned disciplines and topics, including social capital, community development, literacy education, education, biosecurity, domestic violence and child protection. In most cases, the concerns are about improving outcomes from programmes and assessing impact on programme participants.

This article brings together work from two main fields of interest: first, GQR; and second, exploring the purposes of evaluation, focusing on how generalising from qualitative evaluation (GQE) may be achieved, thereby furthering debate around GQE. Supporting data from a suite of policy-based evaluation projects is presented, which illustrate where generalisations from qualitative evaluations have been made. In conjunction with the literature, this provokes a discussion of a potential model for GQE and principles which emerge as relevant when considering GQE.

## Literature review

### *Is there a need to differentiate 'research' from 'evaluation'?*

It could be argued that there is little need to differentiate between research and evaluation because of the similarities between their associated practices, in terms of data collection methods, ethical conduct and analysis (Chen, 2018). Russ-Eft and Preskill (2009) suggest generalisability is not a major goal or concern for evaluation where it is for research, though this assertion does not mean that generalisability is not possible from evaluation. Patton (2015) argues that evaluation and research have different purposes: the former is about making judgements about programmes and policies, while the latter aims to 'test theory and contribute to knowledge' (p. 17). Nor does this assertion mean that new knowledge – for example, in the form of a new theory of change – cannot be used for generalisation. Mathison (2008) suggests that similarities exist: 'Because evaluation requires the investigation of what is, doing evaluation requires doing research' (p. 188). This leaves open the possibility that doing evaluation can create new generalisable knowledge that, for the purposes of this article, an understanding of GQR can be applied to GQE.

### *Definitions of and arguments for GQR*

Vogt (2005) defines generalisability as 'The extent to which you can come to conclusions about one thing (often, a \*population) based on information about another (often, a sample)' (p. 131). This definition disguises a contested understanding among research

methodologies, which tend to split along binarised qualitative/quantitative lines. Miller and Brewer (2003) define quantitative generalisation as ‘a process of first establishing the empirical reliability of facts and then using these facts to assess the validity of theory’ (p. 127). The argument for GQR stems also from other respected academics. For example, Patton (2015) summarises 12 approaches to qualitative generalisation depending on different inquiry perspectives. Eisenhart (2009) makes similar claims, identifying five main types of qualitative generalisation: theoretical, probabilistic, nomological, grounded and syntheses/meta-analysis. Lewis et al. (2013) argue for just three approaches – representational, inferential and theoretical generalisation, while Larsson (2009) identifies five ways that qualitative research can be employed for generalisation – two for falsification and three for confirmation. Tsang (2014), in an examination of generalisation from 25 case studies between 2008 and 2012, shows three types: theoretical, empirical and falsification. He concludes that ‘For cross-population generalization, there is simply no reason why case study results should be inherently less generalizable’ (Tsang, 2014, p. 379) compared with quantitative methods.

In terms of GQEs, Yin (2013) advances the notion of analytic generalisation:

By analytic generalization is meant the extraction of a more abstract level of ideas from a set of case study findings – ideas that nevertheless can pertain to newer situations other than the case(s) in the original case study . . . (p. 325)

Elsewhere, Yin (2014) expands on this idea:

An analytic generalization consists of a carefully posed theoretical statement, theory, or theoretical proposition. The generalization can take the form of a lesson learned, working hypothesis, or other principle that is believed to be applicable to other situations (not just other ‘like cases’). (p. 68)

Based on the above, it is reasonable to accept that GQR is not only possible, but also practically doable. ‘Lessons learned,’ in Yin’s words, are often employed in evaluation reports to demonstrate what went wrong or what worked or what might work – potentially in the form of a theoretical proposition, or as we have previously termed it, a ‘Normative Truth Statement’ (Guenther & Falk, 2019b).

### *The evaluator’s role*

While it may seem somewhat tangential to the discussion of GQE, evaluators themselves play a significant role in the communication and use of evaluation results (King & Alkin, 2019). The Program Evaluation Standards suggest that ‘Without evaluator credibility, the utility of all facets of the evaluation is in jeopardy’ (Yarbrough et al., 2011, p. 15). While the Standards themselves barely mention generalisation, there is adequate justification for use of evaluation to generalise findings within and beyond a programme, as noted in Yin (2014), above. To some extent, credibility depends not only on credentials, but also on the trusted relationship between evaluators and commissioners (Wond, 2017). ‘Engaging in an effective evaluation that involves qualitative inquiry

without developing and maintaining trusting relationships is inconceivable' (Goodyear et al., 2014, p. 258).

### *A typology of GQE*

From a distillation of literature reviewed here, a typology of GQE is presented. Although there are overlaps in the proposed types, the intent is to test the products of our evaluation work to determine how the evaluation approaches fit the typologies presented. The synthesised typology now follows. The creation of a typology extrapolates and extends our previous theorising on the topic of GQR (Guenther & Falk, 2019a, 2019b).

*Generalisation through falsification.* Larsson (2009), discussing GQR, talks about challenging normative assumptions of truth – falsifying commonly held assumptions. Tsang (2014), discussing GQR in case studies, makes a similar point such that the result of falsification is a rejection of previous normative truth through a process of theory testing – 'a rejection of hypotheses based on case evidence' (Tsang 2014, p. 375). For example, if findings from an individual evaluation challenge causal logic, then a new truth statement can be developed. An example would be findings of a programme that previously had shown only positive results, now shows negative findings. Flyvbjerg (2006) points to the power of a single observation which can dismiss a previously held truth:

Falsification is one of the most rigorous tests to which a scientific proposition can be subjected: If just one observation does not fit with the proposition, it is considered not valid generally and must therefore be either revised or rejected. (p. 228)

*Generalisation through patterns.* Larsson (2009) provides justification for generalisation from patterns: 'Generalization by recognizing a pattern can happen even if the context-to-be-understood is different from the original study, as long as the pattern is recognizable' (p. 35). He suggests that this is highly interpretive and depends on audiences drawing conclusions and works best when the research is about process – again, highly relevant to evaluation. Inductive arguments are important here. Patton (2015) suggests that 'Knowledge that can be applied to future action and derived from multiple sources of evidence (triangulation)' (p. 716) in part through considering the patterns observed across programmes and through cross-disciplinary patterns. In a similar vein, Knottnerus et al. (2020) describe a similar process as 'exemplary generalization' (p. 137), where similarities between individual cases can lead to increased likelihood of medical diagnoses. Importantly, the generalisations made are not 100% certain; rather, they increase the probability of similar outcomes based on observed patterns in a limited number of cases to more cases.

*Generalisation from single case to (multiple) case.* Chenail (2010) suggests that the onus is on the 'reader' to generalise from the examined to the unexamined: 'Case-to-case transfer occurs whenever a person in one setting considers adopting a program or idea from

another one' (p. 5). This is akin to Stake's (2000) description of 'naturalistic generalization' where a reader's experience concurs with a normative statement. Smaling (2003), warns that case-to-case generalisation is 'based on chance-capitalization, more a guessing game or a stroke of luck' (p. 53), although he points to the possibility of using 'analogical reasoning' – that is, inductive argumentation (p. 56) which leads to plausible conclusions. Knottnerus et al. (2020) describe this kind of generalisation as 'receptive generalization' (p. 138), which is dependent not on luck, but rather on 'provision of sufficient, accessible and comprehensible information by the researchers' (p. 138).

*Analytic generalisation or generalisation to (and from) theory.* Yin (2014) makes a clear distinction between experiments and case studies: 'Both kinds of studies are likely to strive for generalisable findings or lessons learned – that is, analytic generalizations . . .' (p. 40). Mertens and Wilson (2018) define lessons learned as 'generalizations of conclusions applicable for wider use' (Kindle Location 13024). Yin (2013) also discusses the use of theory in evaluations: Case study evaluations frequently use logic models, initially to express the theoretical causal relationships between an intervention and its outcomes . . . (p. 324). Theory-driven evaluations are designed for this purpose:

The general idea behind [Theory Driven Evaluation] is to sufficiently understand the program so that it can be implemented or generalized beyond a particular context. (Goodyear et al., 2014, p. 86)

Another way of achieving generalisation from empirical qualitative data is to construct theoretical 'types', drawing from the patterns observed in individual cases (Weis & Willems, 2017). This approach differs from the second type described above in that it relies on inferential induction of the case study data to a theoretical type (which are then used in subsequent cases) rather than a predication of generalised truth based on the patterns themselves.

The methodology used for the development of this article now follows.

## Method

Jackson and Mazzei (2018) describe the 'thinking with theory' approach adopted in this article as one which embraces ' . . . the practice of putting theory to work in a move that begins to create a language and way of thinking methodologically and philosophically together that is up to the task' (p. 719). Mazzei (2020) discusses this as 'process methodology as happening in the middle of things, in the threshold', 'always in the process of becoming' (p. 1). She says this 'improvisational inquiry is what I have encountered in my own work thinking concepts and problems together' (p. 1).

The theory in the work described here is concerned with generalisability from qualitative methods. The task is the synthesised understanding of 25 years of evaluative work. The materials are the artefacts or products of our work (reports and publications). We are not as concerned with individual research/evaluation projects, but with

their impact, consequences and unintended outcomes. We, as the authors, have the best chance of knowing that, which can be referenced under insider knowledge. This knowledge holds 'special advantage' for the insider because of the 'enmeshed' nature of research which is situated in sites that 'hold their work and lives together' (Kemmis et al., 2014, p. 5). The problem we are 'in the middle of' is the problem of how to account for the generalisations emerging from numerous qualitative evaluation projects. We acknowledge that this methodology is not the kind of orthodox research methodology used in a single case study or even in a large-scale multi-site evaluation, but it is fit for the task described in this article.

The research question posed for this task is as follows:

How are outcomes from qualitative evaluations generalised?

Richardson and St Pierre (2018) describe 'creative analytic processes' where writing itself is the method of inquiry. The researchers have adopted this process as a means of objectifying data which could be seen as being 'too close to us'. GQR was first described in 2006 at a conference on research for vocational education (Falk & Guenther, 2006). Ten years on, with new evidence, a new analytic and 'thinking with theory' process was resumed, first for a *Handbook of Vocational Education* (Guenther & Falk, 2019b) and then for a journal on qualitative research (Guenther & Falk, 2019a). This latest expression of creative analytical process builds on previous work by extending the thinking with theory from the field of the researcher to the field of the evaluator, retaining the capacity to use insider information regarding uses and impacts through a rigorous and justifiable methodology.

### *Procedure and analysis*

The typology of GQE posited earlier is used as a lens for critical assessment of evaluation projects – the data for the purposes of this article. The 'data' analysed are evaluation reports and papers we have produced. In each report, we looked for evidence of application of (or at least possibility for application of) typologies. The method applied to this process was a series of dialogic discussions around a whiteboard. These were somewhat messy dumps of ideas designed to creatively analyse the body of work up for discussion (see Figure 1). Next, we considered how we used reporting language to support generalisation. We also considered the application of GQR principles we have previously described in terms of Normative Truth Statements (NTSs) and Generalisability Cycles. Finally, we considered what we as evaluators brought to the evaluation project's capacity for generalisation.

### **Findings**

Here, the typologies applied to each of 54 qualitative evaluation projects presented in Table 1 are considered. To give depth to the analysis, a suite of policy-based evaluations conducted between 2005 and 2010 are examined.

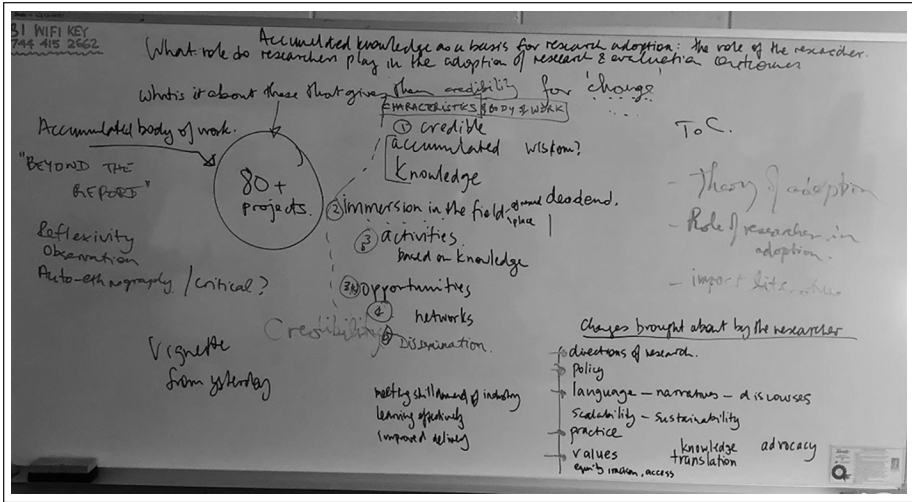


Figure 1. Example of one of the whiteboard discussions used for the purpose of this article.

Table 1. List of qualitative evaluation projects conducted by the authors from 1995 to 2020.

Year	Project name/evaluation of . . .	Principal field of study	Typology(ies) employed <sup>a</sup>
1995	Literacy and numeracy practice	Policy	2, 4
1997	NSW literacy tests	Adult learning	4
1998	Meander Valley Weed Strategy	Agriculture/NRM	2
1998	North East Education and Training project	Adult learning	2
1999	Distance Materials	Adult learning	2
1999	Literacy and Numeracy	Adult learning	2
1999	Deloraine Online	Adult learning	2
1999	User Choice	Policy	2
2000	George Town Municipality Skills Audit	Adult learning	2
2000	Community Development Projects	Community and cultural development	2,3
2000	Queensland Community Training	Adult learning	2, 4
2001	Circular Head Skills Audit	Adult learning	2
2001	Training Needs Analysis (AFFA)	Adult learning	2
2003	DPIWE skills audit	Adult learning	2
2004	Break O'Day Volunteer Resource Centre	Community and cultural development	2
2004	Kigaruk Men's Indigenous Leadership Development Program	Community and cultural development	2, 3
2004	Information Technology Training Project	Adult learning	2

(Continued)



**Table 1.** (Continued)

Year	Project name/evaluation of . . .	Principal field of study	Typology(ies) employed <sup>a</sup>
2004	Secondary Education in the NT	Policy	2
2005	Oakleigh House Redevelopment Project	Schools, families and social support	2
2005	Learnline at Charles Darwin University	Adult learning	2
2005*	NT Government's Family and Domestic Violence Strategy	Policy	2
2005	East Arnhem Communities for Children	Schools, families and social support	2, 4
2005	Palmerston/Tiwi Communities for Children	Schools, families and social support	2, 4
2005	South-East Tasmania Communities for Children	Schools, families and social support	2, 4
2005	Burnie Communities for Children	Schools, families and social support	2, 4
2005	Child Nutrition Program Invest to Grow	Schools, families and social support	2, 4
2006*	NT 'bilateral' Domestic and Family Violence Initiatives	Policy	2, 4
2008	Centacare Tasmania programmes	Schools, families and social support	2
2008*	Northern Territory Differential Response Framework (DRF) Pilot	Policy	3, 4
2008	Safe from the Start	Schools, families and social support	2
2008	Parents and Learning at Napranum	Adult learning	3
2008	FAST NT programmes	Schools, families and social support	1, 2, 4
2009	Department of Education Flexible Provision Program, North-West Tasmania	Schools, families and social support	2, 4
2009	Nungalinga College Bilingual Family and Community Studies Project	Adult learning	3
2009	Akeyulerre Healing Centre	Community and cultural development	1, 3, 4
2009	TrainingPlus	Adult learning	1
2009	Communities for Children Palmerston Tiwi Islands	Schools, families and social support	2, 4
2009	NTCS Sex Offending	Schools, families and social support	2, 4
2010	Centacare's (NW Tasmania) Weathering Family Separation Program	Schools, families and social support	2, 4
2010	Burnie Communities for Children	Schools, families and social support	2, 4

(Continued)



**Table 1.** (Continued)

Year	Project name/evaluation of . . .	Principal field of study	Typology(ies) employed <sup>a</sup>
2010*	Northern Territory Families and Children Targeted Family Support Service (TFSS)	Policy	1, 2, 4
2010	Cognitive Coaching Program for the Tasmanian Principals Association North-West Tasmania	Adult learning	2
2010	Teacher Accommodation in North-West Tasmania	Policy	2
2011	Aspire – A Pathway to Mental Health Inc.	Schools, families and social support	2, 4
2011	Burnie High School Newstart Program	Schools, families and social support	2, 4
2011	TVET System performativity	Adult learning	2
2011	AUSVEG	Agriculture/NRM	2
2014	Wraparound Services	Community and cultural development	1, 2
2016	Warlpiri Education and Training Trust	Adult learning	2, 4
2017	Suicide Story	Schools, families and social support	2, 4
2017	Opportunities and Benefits for collaboration on Indonesian Plant Biosecurity: a bilateral approach	Agriculture/NRM	2, 4
2018	inDigiMOB	Adult learning	2, 4
2019	Nawarddeken Academy	Schools, families and social support	2, 4
2020	Codes4Life	Community and cultural development	2, 4

DRF: differential response framework; NRM: natural resource management.

<sup>a</sup>(1) Falsification; (2) patterns; (3) (single) case to (many) case(s) generalisation through inductive reasoning; and (4) analytic generalisation to (and from) theory (lessons learned).

*Projects and typologies*

Table 1 lists all the qualitative evaluation projects the authors have conducted in the period 1995 to 2020. For each project we have identified one of five principal fields of study as policy, adult learning, agriculture/natural resource management (NRM), community and cultural development, and schools, families and social support. These fields of study represent our fields of interest in research and evaluation. In each case, our assessment determined that some form of GQE resulted. In the last column, we identify the typologies we have employed for each evaluation project. Most projects fit a ‘patterns’ or ‘analytic’ typology, although there is a small number that employ case to case and falsification typologies. There is also a shift in typologies that becomes evident from about 2005 where analytic typologies become more prominent. This came about

primarily because commissioners were demanding use of logic models or theories of change. This is not to suggest that there is an increased probability of generalisability where more than one typology is observed. The point of the table is simply to demonstrate that there is a fit to one or more type of generalisation, similar to an approach adopted by Tsang (2014) in relation to case study research.

### *Case study of policy-based evaluations*

To provide more detail, and for the purpose of illustration, we have (for convenience) selected one strand of policy-based evaluations that we worked on in the period from 2005 to 2010 simply to demonstrate how the typologies may be applied and to show how these typologies relate to different forms of generalisation. The selected cases are marked with asterisks in the first column of Table 1. The first two related to domestic violence (2005 and 2006) and the latter two to child protection (2008 and 2010). All projects were based in the Northern Territory of Australia, and all relate to social policies relevant to families and children.

Each project was auspiced through different government departments. Three of the projects could be described as multi-site projects where a key part of the methodology was identifying patterns across the data set (2005 Family and Domestic Violence and 2010; Family and Domestic Violence, 2006 Targeted Family Support). Collectively, all four projects drew on 251 qualitative participant interviews. Patterns in the data were therefore very important for establishing normative truths that could apply to recommendations. Three of the four projects applied a theoretical analytic process to assess and revise a theory of change (2006 Family and Domestic Violence, 2008 DRF and 2010 TFSS). One project (2008 DRF) was built around a single case study and used several data sources. One project, 2010 TFSS, used falsification by challenging assumptions embedded within the logic of the programme.

In each case, findings led to generalised conclusions. The 2005 Domestic Violence project identified generalisable 'factors contributing to effectiveness of the Strategies' (Learning Research Group and Department of Chief Minister, 2007) and associated recommendations. These learnings were carried over into the second project, the 2006 Family and Domestic Violence evaluation. One of the recommendations from the 2005 project was to use logic models for future assessment, and this was carried into the 2006 evaluation, which drew from 10 separate trials across the Northern Territory. The final report (Arnott et al., 2009) drew on a methodology that included a logic model for each activity and generalised questions, such as 'What are the common principles and factors that contribute to family violence programs working well?' The answer to these questions was in effect a set of common principles and factors, leading to a set of recommendations and an accompanying set of 'criteria for sustainable development', all of which were general in nature, some of which we have drawn on for later studies.

The commissioners then funded a related project which sought to build a differential response framework to address child protection issues in the Northern Territory. The project revolved around a single trial site as a case study, relying on a detailed logic model. This was used to test assumptions of the programme. Although most of

the findings were focused on the case study site, the recommendations were future-oriented as the relevant department was considering extending the trial to a territory-wide service. This was later done and we were invited to conduct an evaluation of what became known as the Targeted Family Support Service (TFSS). The final report for this evaluation (Arnott et al., 2012) again made generalisable future-oriented recommendations that responded in part to one of the evaluation questions: ‘What are the implications that emerge for the future of the DRF in the Northern Territory?’ The final report also explicitly revised the theory of change based on the findings and, importantly, falsified some assumptions built into the logic model. The extent to which government applied these generalisable recommendations is questionable – one reason being that a change in government in 2012 shifted the focus of child protection to a more neo-liberal conservative policy.

In both the domestic violence and child protection examples, there was tacit evidence of an iterative ‘generalisability cycle’ (Guenther & Falk, 2019b) at play. These cycles were not identified at the time. However, with hindsight, the iterative processes of testing and building theory through logic models, confirming and challenging assumptions based on empirical evidence and adding new NTSs as a result were evident over time and across all four projects. The important point to note here is that over time it is possible to track the development of programmes beyond the evaluation and retrospectively observe how they connect and build on each other. We explore this in some detail in our work on GQR (Guenther & Falk, 2019a).

The principles, lessons learned and recommendations generated are effectively ‘Normative Truth Statements’ which apply both within the cases – for example, the 2005 Domestic Violence project had generalisable implications for whole of government, cross silo strategies within the Northern Territory Government – and beyond the case, where, for example, the DRF project made recommendations for 2010 upscaled TFSS project. The common thread across all for cases was the identification of patterns within practice and strategy, which allowed us to draw generalisable conclusions and make recommendations accordingly.

## **Discussion**

### *Thinking with theory: Strengths and limitations*

The approach taken for this article, drawing on a ‘thinking with theory’ coupled with ‘creative analytic’ methodology has allowed us as retrospectively consider the artefacts of our work as assemblages examined through a lens of generalisability theory to generate a typology for GQE. We acknowledge the highly subjective nature of this work, which could be seen as a limitation. However, within the frame of interpretive qualitative paradigms there is adequate justification for this approach, where ‘insider knowledge’ provides capacity for deep critical reflection (Kemmis et al., 2014). Thinking with theory ‘relies on a willingness to borrow and reconfigure concepts, invent approaches, and create new assemblages that demonstrate a range of analytic practices of thought, creativity, and intervention’ (Jackson & Mazzei, 2018, p. 717). The development of typology

for GQE is creative work, but we argue is also a legitimate product of rigorous academic critical reflection. And while not all qualitative evaluations will or should be used for generalisation, the typology can be used by other evaluators to justify utility of their work beyond the specific context of the evaluand.

### *Reporting language for GQE*

Any form of generalisation, quantitative, qualitative or combination, is carried out through language. Language is used with the intent to consider future applications, where there is a desire to know how probable something will be on the basis of what has already happened. The language we use to express this element of prediction could be numerical (based on statistics), or ‘recommendations’ or ‘principles’ often arising from observation of ‘patterns’ in the data, and ‘lessons learned’ often expressed as an ‘analytic generalisation’ towards a theoretical position for future testing. These forms of language are different ways of saying that there is sufficient evidence to evaluate the success or otherwise of something that has happened – something in the past – and imply a level of confidence of what might happen in the future (see Mertens & Wilson, 2018, Chapter 13). For example, an evaluation report may end with a set of recommendations. An example is drawn from the DRF report (Guenther et al., 2009, p. 41):

Recommendation 15: It is recommended that a case management approach be continued.

Here, it can be seen that the recommendation suggests something should be done, based on the findings – in this case, an example of applying the ‘single case to many’ type. The language structure directly implicates future events and outcomes for the upscaled programme to be implemented across multiple sites.

Falsification is an important type of generalisation – and while not observed as often as other types, in the cases described earlier it did play a role. For example, in the 2012 evaluation report for TFSS (Arnott et al., 2012), a recommendation emerged as follows:

Recommendation 3: It is recommended that in any future programs the key assumptions of that program are established.

This recommendation emerged because the programme assumptions developed by the funder were found to be false. One assumption, for example, was that ‘partnerships’ would facilitate better client outcomes. This turned out not to be true for this case and therefore draws into question the future use of partnerships in a continued version of the programme model.

The language of a report allows a prediction about future programmes, policies or practice to be made, even if that is not the intent of the evaluation. Researchers may apply methodologies and make more of the generalisability potential but will nevertheless use similar language structures used by evaluators. This clarifies the point cited earlier by the Joint Committee on Standards for Educational Evaluation Program

Evaluation Standards in its reference to generalisation: ‘direct evaluation impact interferes with generalisation that research aspires to’ (Yarbrough et al., 2011, p. xxvi) in that we can now see the actual point of that ‘interference’: the actual language of the outcomes is the point of interference as it is the point of generalisability, if we are talking about GQE, and the same words are the unavoidable means of reporting any kind of evaluation, and evaluators cannot avoid using that language.

### *Evaluator credibility based on ‘trust’*

There is some literature on the impact of trust on research and evaluation processes (see section on Evaluator Credibility in the work of Yarbrough et al., 2011). There is little research on the impact of trust on how the commissioner of the research or evaluation is influenced by their own developing trust of the evaluator’s track record over time (Wond, 2017 being one exception). So, we rely at this stage on inductions based on our evaluative work.

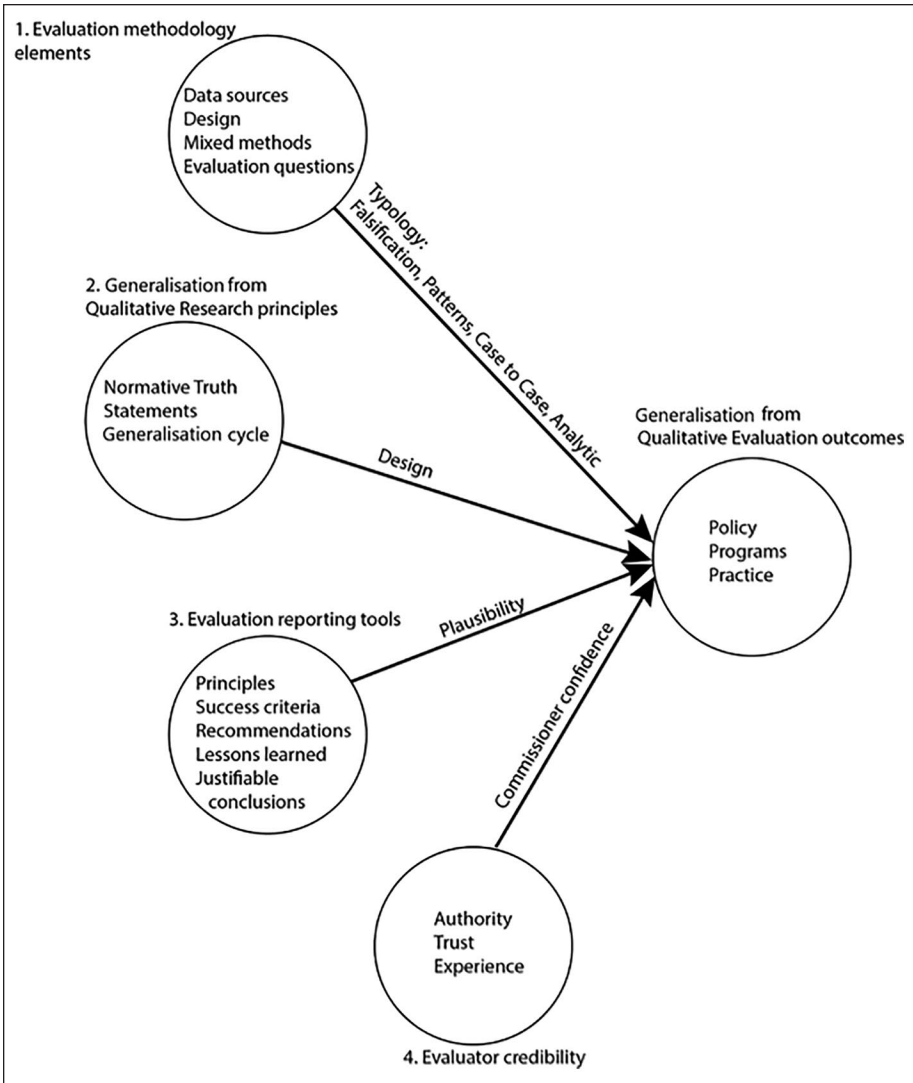
What we can say with confidence is this: Increased trust of the evaluator by the commissioner increases the commissioner’s likelihood of adopting evaluation results, assuring utility, and of implementing principles, recommendations, and lessons learned in ways we classify as ‘generalisation’. For now, we can only suggest that further research on the topic is needed to test our assumptions.

### *Towards a model of GQE application*

The key focus of this article has been the development of a typology based on a review of the literature. We have applied this typology to 54 qualitative evaluation projects to test whether or not the typology fits. In the discussion, particularly on reporting language for GQE, we have unpacked what this might mean for evaluation utility towards generalisation.

In synthesising the discussion and findings up to this point in the article, a tentative model of GQE is posited, with four assumptions: First, there is a perhaps yet still unresolved assumption about the intersection of research and evaluation, clarified somewhat under the subheading above ‘Reporting language for GQE’. Second is the assumption relating to why we would want to generalise from evaluations, noting that not all evaluations can or should be used for this end. Third, we are assuming that the principles for GQR are now well established from our previous work and this can be used as a platform for extrapolation to GQE. Finally, although we acknowledge that this has not been fully explored in this article, we propose that evaluator credibility increases the confidence commissioners have in their ability to generalise findings from evaluations.

Figure 2, developed by the authors to represent the processes of GQE for this article, sets out four elements of qualitative evaluations, aligned in circles down the left side, brought together in relation to GQE outcomes of policy, programmes and practice (right side circle). The four elements on the left are, first, the methodological elements and, second, the capacity to use NTSs and the generalisability cycle – part of the evaluation design for GQE (see Guenther & Falk, 2019b). Third is



**Figure 2.** A model for potential GQE.

important and newly identified in the generalisability potential – the evaluation reporting tools. This consists of the ‘reporting language’ explained earlier, whereby the four types of GQE are embedded in plausible recommendations, lessons learned and principles. These are tools used by evaluators to report outcomes and provide the pivot for potential generalisability from qualitative evaluation. Fourth and finally, there is the importance of authority, trust and credibility as factors in supporting generalisability, which, over the long term, provides significant impetus to

both the uptake of outcomes and generalisability of the evaluation(s). This final point requires validation with further research.

The proposed model is not designed to explain everything about GQE. For example, although we have identified four types of GQE, the extent to which they support generalisability may depend on multiple types being present to support arguments for GQE. Similarly, the influences of external factors such as politics and ideology are not considered in this model.

## Conclusion

We have developed a tentative model of GQE which shows four elements brought together in relation to generalising from the evaluative outcomes of policy, programmes and practice. The four elements are as follows:

1. The data sources on which the evaluation and therefore potential generalisation are based;
2. The use of NTSs and the generalisability;
3. The evaluation reporting tools or reporting language used to express evaluative outcomes;
4. The influence of authority, trust and credibility.

In arriving at this model, the article reviewed the research on GQR and then made the case for extending this to GQE through the articulation of a typology, tested against a set of 54 qualitative evaluation cases. Drawing from a selection of the cases, we have pointed to the utility of evaluations for GQE using the language of reporting where recommendations, principles and lessons learned can be used to create plausible arguments for generalisability.

## Funding

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

## ORCID iDs

John Guenther  <https://orcid.org/0000-0002-0080-1698>

Ian H Falk  <https://orcid.org/0000-0001-9266-7387>

## References

- Arnott, A., Guenther, J., & Cummings, E. (2012). *Evaluation of the Northern Territory targeted family support service*. The Northern Institute, Charles Darwin University.
- Arnott, A., Guenther, J., & Williams, E. (2009). *Towards an investment framework to reduce family violence in the Northern Territory* (Vol. 1). Social Partnerships in Learning Consortium, Charles Darwin University. [https://www.researchgate.net/publication/321874149\\_Towards\\_an\\_investment\\_framework\\_to\\_reduce\\_family\\_violence\\_in\\_the\\_Northern\\_Territory\\_Evaluation\\_report\\_Volume\\_I\\_prepared\\_by\\_the\\_Charles\\_Darwin\\_University\\_Social\\_Partnerships\\_in\\_Learning\\_SPIL\\_Consortium](https://www.researchgate.net/publication/321874149_Towards_an_investment_framework_to_reduce_family_violence_in_the_Northern_Territory_Evaluation_report_Volume_I_prepared_by_the_Charles_Darwin_University_Social_Partnerships_in_Learning_SPIL_Consortium)



- Chen, H. (2018). Evaluation versus research. In B. Frey (Ed.), *The SAGE encyclopedia of educational research, measurement, and evaluation* (pp. 630–632). SAGE. <https://doi.org/10.4135/9781506326139>
- Chenail, R. (2010). Getting specific about qualitative research generalizability. *Journal of Ethnographic & Qualitative Research*, 5, 1–11.
- Eisenhart, M. (2009). Generalization from qualitative inquiry. In K. Ercikan & W. M. Roth (Eds.), *Generalizing from educational research: Beyond qualitative and quantitative polarization* (pp. 51–66). Routledge.
- Falk, I., & Guenther, J. (2006, July 12–14). *Generalising from qualitative research: Case studies from VET in contexts* [Paper presentation]. Th15 NCVET Conference, Mooloolaba, Queensland, Australia.
- Falk, I., Guenther, J., Lambert, T., & Johnstone, K. (2006, September 4–7). *Role of evaluation in assessing and developing communication and governance processes in an evidence-based policy development/implementation environment*. [Paper presentation]. AES International Conference, Darwin, Northern Territory, Australia. <https://espace.cdu.edu.au/view/cdu:2445>
- Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry*, 12(2), 219–245. <https://doi.org/10.1177/1077800405284363>
- Goodyear, P., Barela, E., Jewiss, J., & Usinger, J., (Eds.). (2014). *Qualitative inquiry in evaluation: From theory to practice*. Jossey-Bass.
- Guenther, J., Arnott, A., & Williams, E. (2009). *Evaluation of the Northern Territory differential response pilot project* [Social Partnerships in Learning (SPiL) Consortium]. Charles Darwin University.
- Guenther, J., & Falk, I. (2019a). Generalizing from qualitative research: A reconceptualization based on vocational learning examples. In S. McGrath, M. Mulder, J. Papier, & R. Stuart (Eds.), *Handbook of vocational education and training: Developments in the changing world of work* (pp. 1–19). Springer. [https://doi.org/10.1007/978-3-319-49789-1\\_91-1](https://doi.org/10.1007/978-3-319-49789-1_91-1)
- Guenther, J., & Falk, I. (2019b). Generalising from qualitative research (GQR): A new old approach. *The Qualitative Report*, 24(5), 1012–1033. <https://nsuworks.nova.edu/tqr/vol24/iss5/6/>
- Jackson, A., & Mazzei, L. (2018). Thinking with theory: A new analytic for qualitative inquiry. In N. Denzin & Y. Lincoln (Eds.), *The SAGE handbook of qualitative research* (5th ed., pp. 717–737). SAGE.
- Kemmis, S., McTaggart, R., & Nixon, R. (2014). *The action research planner: Doing critical participatory action research*. Springer.
- King, J. A., & Alkin, M. C. (2019). The centrality of use: Theories of evaluation use and influence and thoughts on the first 50 years of use research. *American Journal of Evaluation*, 40(3), 431–458. <https://doi.org/10.1177/1098214018796328>
- Knottnerus, B. J., Bertels, L. S., & Willems, D. L. (2020). Qualitative approaches can strengthen generalization and application of clinical research. *Journal of Clinical Epidemiology*, 119, 136–139. <https://doi.org/10.1016/j.jclinepi.2019.11.002>
- Larsson, S. (2009). A pluralist view of generalization in qualitative research. *International Journal of Research & Method in Education*, 32(1), 25–38. <https://doi.org/10.1080/17437270902759931>
- Learning Research Group and Department of Chief Minister. (2007). *Evaluation of the Northern Territory government domestic and aboriginal family violence strategies*. Learning Research Group, Charles Darwin University. [https://www.researchgate.net/publication/339549906\\_Evaluation\\_of\\_the\\_Northern\\_Territory\\_Government\\_Domestic\\_and\\_Aboriginal\\_Family\\_Violence\\_Strategies](https://www.researchgate.net/publication/339549906_Evaluation_of_the_Northern_Territory_Government_Domestic_and_Aboriginal_Family_Violence_Strategies)

- Lewis, J., Ritchie, J., Ormston, R., & Morrell, G. (2013). Generalising from qualitative research. In J. Ritchie, J. Lewis, P. Lewis, C. Nicholls, & R. Ormston (Eds.), *Qualitative research practice: A guide for social science students and researchers* (pp. 347–366). SAGE.
- Mathison, S. (2008). What is the difference between evaluation and research: And why do we care? In N. Smith & P. Brandon (Eds.), *Fundamental issues in evaluation* (pp. 183–196). Guilford Press.
- Mazzei, L. A. (2020). Postqualitative inquiry: Or the necessity of theory. *Qualitative Inquiry*, 27(2), 198–200. <https://doi.org/10.1177/1077800420932607>
- Mertens, D. M., & Wilson, A. T. (2018). *Program evaluation theory and practice* (2nd ed.). Guilford Press.
- Miller, R., & Brewer, J., (Eds.). (2003). *The A-Z of social research*. SAGE.
- Patton, M. (2015). *Qualitative research and evaluation methods: Integrating theory and practice* (4th ed.). SAGE.
- Richardson, L., & St Pierre, E.A. (2018). Writing: A method of inquiry. In N. Denzin & Y. Lincoln (Eds.), *The SAGE handbook of qualitative research* (5th ed., pp. 516–529). SAGE.
- Russ-Eft, D., & Preskill, H. (2009). *Evaluation in organizations: A systematic approach to enhancing learning*. Basic Books.
- Smaling, A. (2003). Inductive, analogical, and communicative generalization. *International Journal of Qualitative Methods*, 2(1), 52–67. <https://doi.org/10.1177/160940690300200105>
- Stake, R. (2000). Case studies. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 435–454). SAGE.
- Tsang, E. W. K. (2014). Generalizing from research findings: The merits of case studies. *International Journal of Management Reviews*, 16(4), 369–383. <https://doi.org/10.1111/ijmr.12024>
- Vogt, W. (2005). Generalizability. In W. Vogt (Ed.), *Dictionary of statistics and methodology* (3rd ed.). SAGE.
- Weis, D., & Willems, H. (2017). Aggregation, validation, and generalization of qualitative data: Methodological and practical research strategies illustrated by the research process of an empirically based typology. *Integrative Psychological and Behavioral Science*, 51(2), 223–243. <https://doi.org/10.1007/s12124-016-9372-4>
- Wond, T. (2017). Trust matters: Distrust in an external evaluation of a public sector program. *International Journal of Public Administration*, 40(5), 408–415. <https://doi.org/10.1080/01900692.2015.1126732>
- Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). SAGE.
- Yin, R. (2013). Validity and generalization in future case study evaluations. *Evaluation*, 19(3), 321–332. <https://doi.org/10.1177/1356389013497081>
- Yin, R. (2014). *Case study design and methods* (5th ed.). SAGE.