CHAPTER 1: INTRODUCTION

This thesis uses a case study to examine how the theory underlying an adapted version of Guba and Lincoln's (1989) Fourth Generation Evaluation (4GE) model was enacted in the context of evaluating a secondary science teacher preparation program in one rural university in New South Wales, Australia.

In this introduction, I discuss the importance of program evaluation to Science Teacher Preparation Programs (STPPs) in Australia and address criticisms related to program evaluation in that specific context. I then outline how these criticisms extend to the program evaluation discipline in general and develop an argument for the need to undertake empirical investigations of the relationship between the theory and the practice of program evaluation models. In the remainder of the chapter, I present the aims of my research, the general research question, and I describe the significance, design and delimitations of the study. I close the chapter by outlining the structure of the thesis.

1.1 The importance of and challenges to the evaluation of Australian science teacher preparation programs

With globalization and increased competitiveness across various economic and social sectors, Australian universities are experiencing overwhelming pressure to demonstrate and provide evidence for the quality of education they are delivering (Anderson, Johnson, & Milligan, 2000; Shah, Lewis, & Fitzgerald, 2011). The quality improvement impetus is especially pronounced in the area of science teacher education, where the preparation of quality science teachers who can prepare students to meet the challenges of the 21st century has become a national priority, as is evident in the growing emphasis on the centrality of scientific literacy for the development of a knowledge-based economy (Australian Academy of Science (AAS), 2005; Chubb, 2014; Department of Education, 2002, 2003; Office of the Chief Scientist, 2012; Tytler, 2007; West, 2012).

Unfortunately, as the rising wave of criticisms about the quality of science teaching in Australian schools indicates (Dinham, 2013), science educators in Australia are, as yet, unable to provide rigorous evidence for the effectiveness of their STPPs in preparing quality science teachers. Korthagen, Loughran and Russell (2006, p. 1021) argue that teacher education programs (including STPPs) are seen by teachers to be mostly irrelevant to workplace needs and ill equipped to make long-lasting attitudinal or behavioral shifts in the

practices of teachers who, gradually, revert to the more traditional ways of teaching that they experienced in their days as school pupils. The Science, ICT and Mathematics Education for Regional and Rural Australia (SiMERR) National Survey (Lyons, Cooksey, Panizzon, Parnell, & Pegg, 2006) found that secondary science teachers in Australia (N = 561) reported that teacher education at university prepared them best for teaching their subject matter, but less well for other aspects of teaching (managing student behavior, teaching gifted and talented students, using ICT, etc.). On average, the respondents considered their pre-service education courses to have left them only "moderately prepared" for teaching science (Lyons et al., 2006, p. 77). More recently, an inquiry into the quality of senior science teaching and learning in Australian schools revealed that science is being taught in a "traditional way using the transmission model" (Goodrum, Druhan, & Abbs, 2012, p. ii) with 73 per cent of science students explaining that they spent most of the science lessons copying notes from the teacher. While it could be argued that the quality of school science instruction is not a definitive indicator of the quality of STPPs, it nevertheless suggests considerable reservations about the effectiveness of STPPs in preparing quality science teacher graduates.

In parallel with these criticisms, the Australian Federal government's increasing scrutiny of the quality of teacher education nation-wide has accentuated the stress placed on Australian universities to provide convincing evidence about the effectiveness of their teacher preparation programs (including STPPs). For instance, in March 2013, the Federal Government announced a new policy to bring about further improvements to the quality of teacher preparation through the development and implementation of more rigorous standards for teacher training programs (Australian Government, 2013). Similarly, developments such as the Tertiary Education Quality and Standards Agency (TEQSA), the Australian Government Quality Teaching Program (AGQTP) and the Australian Institute for Teaching and School Leadership (AITSL) reflect the government's strong focus on improving the quality of teaching and teacher education. Furthermore, national reports and inquiries into the quality of teaching have fuelled the political agenda advocating further quality monitoring and increased university accountability (see, for example, *Australia's Teachers Australia's Future: Advancing Innovation, Science, Technology and Mathematics* (Committee for the Review of Teaching and Teacher Education, 2003); *Top of the Class: Report on the Inquiry into Teacher Education* (Standing Committee on Educational and Vocational Training, 2007); *The Status and Quality of Year 11 and 12 Science in Australian Schools* (Goodrum et

al., 2012); *Great Teaching, Inspired Learning: A Blueprint for Action* (New South Wales Government, 2013)).

Monitoring the quality of academic programs has been a long-established tradition in most Australian universities. In the specific case of STPPs, program evaluations are typically overseen by a dedicated committee of the universities' Academic Boards (Rowlands, 2012; Winchester, 2007). Nevertheless, according to Rowlands (2013, p. 143), academic boards are "ill-equipped and ineffective" in assuring academic quality, and their role is perceived by many academics to be symbolic and figurative (Anderson, 2006; Blackmore, 2009; Rowlands, 2013). Other critics, such as Houston (2010), argue that the evaluation practices of many Academic Boards seem to have privileged the purpose of providing evidence of the universities' conformance to accountability requirements set by the Federal government at the expense of ensuring academic and program improvement. Similarly, Blackmore (2009, p. 858) explains that multiple and often contradictory discourses about academic quality have emerged within Australian universities. By drawing a useful distinction between evaluation for improvement and evaluation for accountability, she argues that the latter describes better the practices of academic boards:

> Evaluation for improvement focuses on identifying what worked, how, and why it worked, and how performance can be improved. Evaluation for accountability focuses on the processes and outcomes: the visible and the measurable, tracking the paper trails to predetermined outcomes. (p. 861)

On the matter of the implications of prevailing evaluation practices of some Academic Boards, Blackmore (2009) contends that through "mapping out paper trails of curriculum and assessment to provide the evidence required by quality assurance audits", Academic Boards have changed the institutional practices they are monitoring by conceptualizing quality:

> Quality is thus readily collapsed into an accountability exercise, operationalized as meeting predetermined benchmarks and standards, following processes and procedures laid out by a paper trail, rather than quality of teaching in terms of student needs and substantive pedagogical relationships. (pp. 862-863)

Houston (2010, p. 178) added that not only have these evaluation processes prompted ritual responses or performances of compliance but also "discouraged engagement with ideas and practices to improve quality".

The research examining the effectiveness of Academic Boards in improving and monitoring the quality of academic programs is scarce and inconclusive (Rowlands, 2013). Nevertheless, as the literature suggests, it seems that the existing evaluation practices carried out by these

boards are seldom done with reference to sound program evaluation theory. In light of the mounting pressure on universities to constantly improve their STPPs, it is timely that they revisit how program evaluation ought to be carried out.

Fitzpatrick, Sanders and Worthen (2004) argue that institutions should weigh up the benefits of evaluation models in order to adopt an institution-specific model that can satisfy the purposes of their evaluations. This necessitates that evaluators examine the various models and decide which one best suits their specific context. However, therein lies a challenge: evaluation models are not totally understood and how they are enacted in practice is still unclear (Mark, 2008; Miller, 2010; Smith, 2010). The challenge is to make an informed decision about which of the various evaluation models to choose and under what conditions the choice should be made. To overcome this challenge, evaluators need to understand and be able to compare the theories underlying the different program evaluation models. They also need to understand the peculiarities and challenges inherent to the implementation of the models in practice. It is to this challenge that I turn in the following section.

1.2 The problem of the theory-practice relationship in program evaluation

The need to examine the connection between the theory and practice of program evaluation models is not specific to Australian STPPs. On the contrary, it has been one of the most persistent challenges facing program evaluation scholarship. In the past two decades, program evaluation has emerged as a distinct scholarly discipline (Fitzpatrick et al., 2004), and several evaluation journals and professional international evaluation societies have emerged (e.g. the Australasian Evaluation Society, the American Evaluation Association, the Canadian Evaluation Society, the European Evaluation Society and the African Evaluation Association). All these developments have contributed to making evaluation – and to some extent program evaluation – a unique and distinctive discipline (Donaldson & Lipsey, 2006).

However, the shift in the status of professionalizing evaluation as a distinct discipline has not been accompanied by the required vigor in all related areas of research. For example, while the field of evaluation use and related concepts has prospered in the last few years (Germuth, 2010; Mark, 2008), empirical studies on how program evaluation theories and their practice are connected remain scarce (e.g. Datta, 2003; Henry & Mark, 2003b; King, 2003; Mark, 2001, 2003, 2008; Miller, 2010; Shadish, Cook, & Leviton, 1991; Smith, 1993; Smith, 2010).

A recurring question in program evaluation is “what works in program evaluation and under what conditions?” (Stufflebeam & Shinkfield, 2007). Answering this requires evaluators to

have at their disposal a wealth of empirical studies that provide insights into how different models are enacted across different contexts, as well as their individual assets and drawbacks. To illustrate the point, I borrow an example from Shadish (1998) who explained how psychotherapists use various therapeutic approaches to treat different cases. Shadish (1998, p. 10) points out that "when a new therapy appears, [psychotherapists] can classify it and quickly understand some of its likely strengths and weaknesses". This is one of the assets of the psychotherapy profession. Developing a similar knowledge base for program evaluators requires evaluation scholars to:

> move beyond the relatively weak connected theoretical musings and anecdotal reports of practice […] to add to the empirical knowledge base through carefully developed and executed studies that have the potential to extend our theories and guide our practices, studies that manifestly strengthen the link between theory and practice. (Cousins & Earl, 1999, p. 316)

The consequences of the gap in our understanding of the theory-practice relationship in the program evaluation scholarship are numerous. Without the necessary clarity about the connection between theory and practice, program evaluation models lack empirical grounding for defending their merit. Some program evaluation critics (e.g., Smith, 2008) argue that some of the existing models are conceptually ambiguous and that their defining features are not conceptually discernible. For example, one of the prominent issues in the debates on empowerment evaluation concerns the extent to which the model can be readily distinguished from other models that share with it an emphasis on collaborative processes and capacity development (Miller & Campbell, 2006). In this respect, the empirical investigation of evaluation would offer the requisite knowledge for researchers and practitioners to "[sort] through theories and [determine] their ultimate feasibility and merit" (Miller, 2010, p. 391). Furthermore, critics question whether program evaluation models actually provide enough guidance for their users, and whether such guidance can be and are actually implemented (Miller, 2010). These critics base their arguments on findings, such as those derived from the Christie (2003) study, which reported that around 70 per cent of evaluation practitioners (N=138) did not reportedly adhere to the prescriptions of the models they were using.

In sum, investigating the theory-practice relationship in relation to program evaluation models is important. It enables evaluation scholars to understand the various models and assess their relative strengths and weaknesses (Henry & Mark, 2003b). Furthermore, such investigations provide the evidence-base needed to give critical weight to evaluation findings, as well as increase their impacts and convince others – including potential clients of an

evaluation model – about the evaluation's contribution to program improvement (Henry & Mark, 2003b, pp. 72-73).

1.3 Rationale, aim and general research question

Despite repeated pleas for more empirical studies of the theory-practice relationship in program evaluation (Miller, 2010), little response is evident in the literature. In an attempt to contribute knowledge to this under-researched topic, I sought, in this study, to explore the fit between the theory and practice of an adapted version of Guba and Lincoln's (1989) 4GE model in the context of evaluating a STPP in a university in New South Wales, Australia.

The main purpose of the investigation was to provide insights into how well the theory underlying an adapted version of the 4GE corresponded with the processes and outcomes of its implementation across a defined context of practice. The general research question that guided the design and implementation of this study is:

> **How congruent are the underlying theory and practice of an adapted version of the 4GE model in the context of evaluating a secondary science teacher preparation program in an Australian university?**

Through investigating the relationship between the theory and practice of the adapted version of the 4GE, this study also contributes insights into a number of factors that strengthen or weaken this relationship. Using these insights and the lessons learned from the practice of the model, I was also able to discuss the challenges of using the adapted version of the 4GE.

1.4 The significance of the study

This study is significant for three reasons: first, it contributes knowledge to an underexplored area of research in program evaluation; second, it focuses on an original model of evaluation that is adapted from Guba and Lincoln's 4GE; and, third, it is situated in a context where the need for evaluation is particularly pronounced.

The significance of this study lies first and foremost in its contribution to a greater understanding of whether and how the theory and real-world practice of the 4GE align. This is an original contribution in light of the dearth of research in this area. From this perspective, the scholarly benefits of this study include providing evidence-based recommendations for the development and improvement of the 4GE, and offering a protocol for studying the theory-practice relationship in program evaluation.

The second reason this study is significant is because it uses an adapted version of the 4GE as a model for evaluation, which increases the usability of the model for the modern context. While Guba and Lincoln (1989) have argued persuasively for the merit of the 4GE, the model has not been extensively used in evaluation studies and only sporadically so in educational evaluation research, mainly due to its time-consuming processes and interpretive nature (Fishman, 1992; Lay & Papadopoulos, 2007). In this study, I have modernized the approach, making use of technological advances to enhance the transparency of the model, while at the same time decreasing the time frame and increasing the cost-effectiveness of its implementation. The adapted version of the 4GE model may be, thus, potentially more useful for evaluators.

The third reason the study is significant is that it takes place in the specific context of evaluating a STPP. Given the importance of program evaluation in science teacher preparation in Australia, the application of the 4GE is particularly interesting. It brings together the perspectives of various stakeholders in a discourse about the strengths and weaknesses of a STPP and allows the generation of context-specific improvement suggestions. As such, the study offers an alternative model for use by university faculties and Academic Boards who, as has been explained, face criticism about their current practices. The 4GE has been implemented successfully in many domains, such as nursing (e.g. Kosh, 1996; Kosh, 2000), health science (e.g. King & Appleton, 1999) and social studies (e.g. Huebner & Betts, 1999; Lay & Papadopoulos, 2007). Its use has revealed issues related to the programs evaluated and generated viable suggestions for improvement. The success of the 4GE in these contexts makes it a worthwhile candidate for use in an educational context, where evaluative information is both scarce and needed.

1.5 Overview of the study

This study comprises two components. The first is a case study evaluation of a STPP using an adapted version of Guba and Lincoln's 4GE. The second is an investigation of the application of the adapted version of the 4GE in the case study and an exploration of the relationship between the model's theory and its practice.

With regard to the first component, I used an adapted version of the 4GE to evaluate the Graduate Diploma in Education for science teaching at a university in New South Wales, Australia. This first component was framed as an attempt to model how the evaluation of an academic program might be carried out by Academic Boards and university faculties, or

schools. For the second component, I used an interpretive case study approach to investigate the fit between the theory and the practice of the adapted version of the 4GE employed in the first component.

Although the first component of the study is an evaluation and the second is formal academic research of that evaluation, the first component informs the second and I used the same tools to gather data for the two components. The data collection consisted of two interview rounds and one online discussion forum. Intensive individual interviews with 15 educational stakeholders (university lecturers and recently graduated science teachers) were used, along with analysis of artifacts, field notes and journals. An interactive online discussion forum encompassing all willing participants was used to collect further data. Following the principles of interpretive inquiries, data were analyzed concurrently with the data collection. The development of the data collection tools was informed by these analyses in a recursive fashion. The overall analytical process was inductive and based on the premises of the constant comparison (Charmaz, 2011) and the negative case analysis (Robinson, 1951) approaches.

1.6 Delimitations of the study

The focus of the study was to examine the congruence between theory and practice of an adapted version of the 4GE. Therefore, I will only discuss the delimitations related to the second and focal component of the study: the empirical investigation of the application of an adapted version of the 4GE.

One delimitation of the study derives from the research objectives. In this study, I endeavored to examine how the theory underlying an adapted version of the 4GE mapped into practice and whether, in fact, it did. This is not to be confused with the program theory evaluation tradition, which contemplates a program's theory (not an evaluation model's theory) and investigates how it unfolds in practice. While a program's theory explains why a program is expected to work and how its assumptions relate to its outcomes (Bickman, 1987), a program evaluation model's theory describes how the evaluation of a program ought to be done and what its outcomes are. Also, the current research is not to be confused with meta-evaluation (Scriven, 1969), which is the evaluation against pertinent standards. For this reason, the literature review presented in chapter two will not tap into these two areas of research.

A second delimitation of the study is the selection of a single application of the adapted 4GE. Multiple evaluation cases could have yielded better grounds for comparison and analysis of

data, but time and resource constraints caused me to limit the scope of this research to a single evaluation case. While the resulting findings cannot be generalized beyond the boundaries of the single case, there is certainly merit in doing the research, because it provides a new approach to the investigation of program evaluation models that might be more broadly applied.

A third delimitation of the study is the number of individuals who were invited to participate in the study. Since the evaluation focused on a STPP, only science teachers were invited to participate, even though other types of teachers also participated in many of the program's units. As a result, the pool of potential participants was reduced to the relatively few who held a science degree and had graduated to become science teachers. Furthermore, among the various potential stakeholders, only lecturers and teachers were recruited to participate in this research. Other stakeholders, such as accreditation agencies (e.g., the Australian Institute for Teaching and School Leadership) or quality assurance agencies (e.g., the Tertiary Education Quality Standards Agency) were excluded on the basis of their remote acquaintance with this specific program. I reasoned further that, since these organizations publicized their monitoring mechanisms online, there was no added value in including them in this study.

A fourth delimitation for the study was the inclusion of participants who graduated between the years 2006 and 2011. The decision to have a five-year time span was intended to mimic academic boards in their five-yearly evaluation cycles. However, during the five-year period, there were important changes to the program. Most importantly, there was some restructuring of the program based on the accreditation requirements set by the New South Wales Institute of Teachers in 2008 and the consequent extension of the program's duration from a year-long to one and a half year long program. There were also modifications to unit content brought about by turnover of academics teaching into the units.

1.7 Structure of the thesis

In this chapter, I have established the rationale for undertaking an empirical examination of the relationship between the theory and practice of an adapted version of the 4GE. I have also outlined the general research question and discussed the significance and delimitations of this study. In the next chapter, I review the literature on program evaluation and use it to develop a conceptual framework that enables the systematic investigation of the relationship between the theory and practice of evaluation models. In chapter three, I present a rationale for choosing the 4GE for the current study and present an adapted version of it. In chapter four, I

discuss the methodological strategies employed in the study and expand on the design and processes of data collection and analysis. Following that, the findings from the STPP evaluation are presented and discussed in chapter five. In chapter six, I present my findings and interpretations from the empirical investigation of the theory-practice relationship. Finally, in chapter seven, I discuss the contributions of my study, its limitations, and its implications for both theory and practice.

CHAPTER 2: LITERATURE REVIEW

In this chapter, I survey the program evaluation literature and discuss the complexities of definitions of programs and program evaluation. I also discuss the types and roles of theories of program evaluation. Additionally, I overview and critique the literature investigating the important issue of theory-practice relationship in relation to program evaluation models and present a conceptual framework for investigating this relationship. To this end, I develop and define a theoretical tool, the Program Evaluation Models' Essential Dimensions (PEMED), for defining program evaluation models. Using the PEMED in combination with Miller's (2010) criteria for empirically studying program evaluation models, I illustrate how the proposed conceptual framework can be used to investigate the relationship between the theory and practice of program evaluation models.

2.1 A survey of the program evaluation landscape

2.1.1 Definition of programs

Before starting the discussion about program evaluation theories, it is important to explain what is meant by the term "program". Simply defined, a program is one way of enacting a policy. It is an interpretation of how a given policy is portrayed in terms of specific activities and resources delivered to a particular audience. This definition is concise but simplistic because it does not capture the complex elements of a program. In this dissertation, I have adopted another, more exhaustive definition from Yarbrough, Shulha, Hopson and Caruthers (2011):

> [A program is] a set of planned systematic activities, using managed resources, to achieve specified goals, related to specific needs, of specific, identified, participating human individuals or groups, in specific contexts, resulting in documentable outputs, outcomes, and impacts, following assumed (explicit or implicit) systems of beliefs (diagnostic, causal, intervention, and implementation theories about how the program works), with specific, investigable costs and benefits. (p. xxiv)

While it is important to articulate a definition of a program clearly and concisely, it does not follow that evaluations should (or possibly can) assess a program's entire set of components. Indeed, an evaluation of all of the components of a program could prove more costly than the program itself and an argument can, thus, be made for its rejection based on cost-effectiveness. According to Yarbrough et al. (2011, p. xxiv), program components that can be subject to evaluation include one or more of the following:

> contexts and how they interact with programs and program components, participants and other beneficiaries as well as those who encounter costs or loss of benefits, needs, problems, and policy spaces in programs and their contexts, goals and objectives, resources and costs of all kinds, including staff, facilities, materials, and opportunity costs, activities, procedures, plans, policies, and products, logic models, beliefs, assumptions, and implicit and explicit program theories explaining why and how programs should work, and/or outputs, results, benefits, outcomes, and impacts.

Program evaluation approaches are usually designed to evaluate one or more of these components. I have adopted the Yarbrough et al. definition of programs in this study to highlight the most relevant components of the program under consideration. I use the definition to clarify the evaluated program later in the methodology section.

2.1.2 Definitions of program evaluation

A survey of the literature reveals a multitude of opinions regarding what evaluators do and what professional evaluation entails. Perhaps the simplest yet most enduring definition of program evaluation is that which coined the term as the systematic investigation of the worth and merit of an evaluand (that is, the entity to be evaluated and which, in the particular case of program evaluation, is a program). This definition is often associated with the perspective held by one of the pioneers in the field, Scriven (1991), who defined evaluation as:

> "[…] the process of determining the merit, worth, or value of something, or the product of that process… The evaluation process normally involves some identification of relevant standards of merit, worth, or value; some investigation of the performance of the evaluand on these standards; and some integration or synthesis of the results to achieve an overall evaluation or set of associated evaluations" (p. 139).

Many scholars of evaluation have elaborated on this definition to include other core factors that reflect their focus of evaluation. Some researchers have elaborated on the "valuing" component and expanded the meaning of the term to include not only merit and worth but other functions; for example, Stufflebeam and Shinkfield's (2007, p. 698) definition of evaluation as "the systematic process of delineating, obtaining, reporting, and applying descriptive and judgmental information about some object's merit, worth, probity, feasibility, safety, significance, or equity". Other researchers added a component of purpose to explain the feedback functions of evaluation, such as assisting judgment, informing decision-making, building capacity or empowerment. For example, Patton (2008, p. 39) defined program evaluation as "the systematic collection of information about the activities, characteristics, and results of programs to make judgments about the program, improve the program or

further develop program effectiveness, inform decisions about future programming, and/or increase understanding".

The range of definitions is interesting and rich but can sometimes cause confusion. However, a close examination of the various definitions reveals two constant processes that invariably define the evaluation scholarship: systematic investigation and feedback function. The first process, systematic investigation, draws a distinction between evaluation as a professional activity and evaluation as an everyday exercise of judgment. While the two activities share the same conceptual umbrella, the scholarly discipline of evaluation – and program evaluation in particular – is based on the principles of systematic investigation. This means that the evaluative work involves rigorous planning for and collection of data, and assessment of the validity of the information and inferences derived from it. The second process, commonly found across the range of definitions, emphasizes the feedback function of evaluation. This function has often been translated as a statement of the purpose of evaluation. Therefore, the feedback function of program evaluation approaches includes determining worth and/or merit, helping decision-making and action-taking, providing judgments about what works in a program and/or its implementation, providing judgment about cost-effectiveness and cost-efficiency, and so forth.

One particular definition of program evaluation that guided this study is from the third edition of the *Program Evaluation Standards* (Yarbrough et al., 2011). The definition clearly and accurately outlines evaluation in terms of its two core processes yet does not specify the methodology, purposes and evaluation use:

> [Program evaluation is] the systematic investigation of the quality of programs, projects, subprograms, subprojects, and/or any of their components or elements, together or singly, for purposes of decision-making, judgments, conclusions, findings, new knowledge, organizational development, and capacity building in response to the needs of identified stakeholders, leading to improvement and/or accountability in the users' program and systems, ultimately contributing to organizational or social value. (p. xxv)

By highlighting that there is variation in definitions, I do not mean to imply that evaluation theorists and practitioners need to commit consensually to an absolute perspective; rather, the fact of variation means program evaluators ought to be accountable to and clear about their espoused definition of evaluation in every evaluation situation they come across (Patton, 2003). Indeed, when engaging in evaluation practice, professional evaluators are involved in a deliberate and complex act that requires informed thinking about the various possible

definitions existing in the evaluation landscape. The diversity in perspectives on and definitions of evaluation has been embraced as an essential asset of the discipline; particularly, evaluation scholars agree that program evaluation is, essentially, a needs-based responsive discipline, and selecting an evaluation approach and, therefore, a definition of evaluation is contingent on the particular context within which the evaluation is embedded (Contandriopoulos & Brousselle, 2012; Kundin, 2010; Smith, 2010; Tourmen, 2009).

2.1.3 Types of program evaluation theories

Although still a relatively new discipline, the program evaluation literature offers numerous theories to guide evaluators with their practice and understanding of evaluation. These theories have taken different shapes and assumed different roles. The literature is often confusing about how these theories ought to be used. Therefore, a clarification of the types, uses and roles of theories in program evaluation is necessary at this stage.

Scriven (2003) provides an account of what evaluation theories are and how they can be understood. He notes that an evaluation theory can be either normative or prescriptive. Normative evaluation theories provide accounts about what the norms and nature of evaluation practice ought to be. These are general theories about "what evaluation should do or be, or how it should be conceived or defined" (p. 15). Conceptual frameworks that make claims about the logic and principles of designing and implementing an evaluation would fall under this category. Prescriptive evaluation theories, on the other hand, are about "what theories there are, or what evaluations types there are … and what they in fact do, or have done, or why or how they did or do that" (p. 15). Empowerment evaluation (Fetterman & Wandersman, 2005) and responsive evaluation (Stake, 1983) would both fall under this category as they both aim to explain one particular way of doing evaluation. Scriven (2003) also uses the term program evaluation "metatheory" to refer to a theory about program evaluation theories. A metatheory generally aims to classify normative or prescriptive theories or analyze and explain them. An example of a metatheory is Owen's (2006) forms of evaluation (proactive, clarificative, participatory/interactive, monitoring and impact).

Smith (2010) uses a parallel yet different nomenclature to distinguish between the different types of program evaluation theories. Accordingly, he restricts the use of the term "theory" to what Scriven refers to as normative theories. Importantly, Smith (2010) draws a distinction between the use of the term "theory" in the sciences and its use in the domain of evaluation. He asserts that, while the term in science refers to a coherent body of knowledge with

predictive powers, in evaluation, the term simply refers to conceptual dispositions relating to the most fundamental questions about evaluation practice. Smith (2010) uses the term "model" to refer to Scriven's prescriptive theories, and the term "approach" to refer to Scriven's metatheories. According to Smith (2010, p. 384), a model is a set of "more or less coherent resolutions of theoretical issues that provide a compatible set of prescriptions for how to conduct an evaluation" and approaches are "even broader conceptual collections, often representing groupings of models sharing similar principles" which serve the purpose of categorizing similar ways of reasoning and undertaking evaluation.

In this dissertation, I use Smith's classification scheme of program evaluation theories to draw distinctions where required. In specific terms, I use the term "approach" synonymously with "metatheory" and the term "model" synonymously with prescriptive theories. Any other use of the expression program evaluation theory will be used in reference to normative theories of evaluation. Table 2.1 provides a summary of these lexical distinctions.

Table 2.1 – Program evaluation theories nomenclature

Nomenclature adopted in this dissertation	Synonyms	Definitions	Examples
Program evaluation theory (Smith, 2010)	Normative theory (Scriven, 2003)	Provide accounts about what the norms and nature of evaluation practice ought to be	Qualitative Program Evaluation (Greene, 1994)
Approach (Smith, 2010)	Typology (Hansen, 2005) Metatheory (Scriven, 2003)	Classifying, comparing and contrasting descriptive theories	Evaluation forms (Owen, 2006)
Model (Smith, 2010)	Prescriptive theory (Scriven, 2003)	Explain what the norms actually are and how they are enacted and why.	4GE (Guba & Lincoln, 1989)

2.1.4 Roles of program evaluation theories

The status and roles of theories in evaluation have often been contentious among scholars (Donaldson & Lipsey, 2006). Distinguished evaluators at either end of the debate have long expressed disagreement about whether evaluation theory does or should play a central role for practitioners. For example, field pioneers such as Scriven (1998), Chelimsky (1998) and Stufflebeam (2001) assert that the mere assumption that a theory can capture, *a priori*, the complexities involved in developing and implementing a program in the complicated real world context as well as provide the necessary knowledge to develop evaluation questions

should be discredited. These researchers assert that evaluation practice does not primarily draw on theory but on a different body of knowledge. Schwandt (2003) introduced the term "practical knowledge in evaluation" to describe this body of knowledge. Inherent to his argument is the understanding that the nature of practical knowledge extends well beyond being the mere application of some theoretical dispositions. According to Schwandt (2003), evaluation is practiced on rough ground since evaluators' decisions are essentially rooted in the contingencies, parameters and constraints of the context of practice, which are not necessarily captured in theoretical prescriptions and admonitions in the scholarly evaluation literature. As such, the domain of practice is more to theory than just a mere application; it is the locus where the actual expertise of evaluators is forged, refined and developed.

On the other hand, scholars from the other end of the debate argue that program evaluation theory should be the platform for advancing professional practice in the discipline (see for example Alkin, 2013; Christie, 2003; Donaldson, 2003; Fetterman, 2003; Mark, 2003). For these scholars, the relationship between theory and practice is conceptualized as a symbiotic partnership. Accordingly, the act of evaluation should be predicated by theoretical assumptions and informed by explicit knowledge of theories that guide practice. Conversely, the study of evaluation practice should result in the development of better theories about evaluation. A third group of scholars sits outside the debate continuum, claiming that the exact role and place of theory, and tacit practical knowledge in the evaluation profession is still to be uncovered (e.g. Tourmen, 2009).

Without delving further into the debate, in this dissertation I place myself within the second camp, echoing the position of Shadish (1998) who articulates the centrality of evaluation theory clearly in his address to the American Evaluation Association entitled "Evaluation theory is who we are":

> [Evaluation theory] is what we talk about more than anything else, it seems to give rise to our most trenchant debates, it gives us the language we use for talking to ourselves and others, and perhaps, most important, it is what makes us different from other professions … Every profession needs a knowledge base. For us, evaluation theory is that knowledge base. (p. 1)

Shadish (1998) explicates further how evaluation theory contributes to shaping and refining the identity of the profession. He identifies five crucial avenues where evaluation theory plays a central role: (1) through providing the language that evaluators use to talk about the discipline; (2) by encompassing the most salient aspects of the profession; (3) by defining themes of major evaluation conferences; (4) through delimiting the unique body of

knowledge peculiar to the evaluation discipline and distinguishing it from other disciplines; and (5) through presenting this identification about evaluators to people from outside the discipline so that they can appreciate the profession as an independent discipline.

In the same vein, Miller (2010) focuses on highlighting the direct contributions that evaluation theories offer to practice. She explains that evaluation theories:

> provide practitioners with ideological perspectives on evaluation, sensitizing concepts to guide practice and, to varying degrees, with specific guidance on matters such as defining the appropriate role of the evaluator in relationship to the evaluand and to individuals in the settings which house it, selecting evaluation questions and pairing these with methods, determining whose informational needs are to be met via the evaluation, selecting who may participate in shaping the direction of the evaluation and in what fashion, and identifying when, how, and to whom evaluation findings are to be disseminated with what purpose. (p. 390)

Additionally, Donaldson and Lipsey (2006, p. 62) explain that evaluation theories enable practitioners to educate potential clients and evaluation users, choose the most appropriate evaluation models for particular programs and contexts and most importantly, anticipate, understand and deal with reactions to their work and avoid being the victim of the "kill the messenger phenomenon".

2.1.5 Studies addressing the theory-practice relationship in relation to program evaluation models

Having discussed the centrality of evaluation theory to defining the identity of the discipline and offering guidelines for its practice, it is necessary to explore studies that have investigated the connection between the theory and practice of program evaluation models. In this section, I have grouped these studies under three categories reflecting three approaches to this type of research. For each category, I showcase whether and how the research approach enables program evaluators to understand program evaluation models singularly and comparatively so as to describe how each works best under which conditions.

The first approach used to study the theory-practice relationship aims at investigating a single model's theoretical assertions through analyzing either a single or multiple case applications of that model. For example, Sridharan and Nakaima (2012) studied how a case evaluation of a dance program for health promotion using the theory-driven evaluation model helped outline a number of challenges concerning the implementation of the model. The authors then analyzed the challenges involved in implementing this model and recast the challenges into questions to provoke further progress with the model's underlying theory and assumptions.

Sridharan and Nakaima's (2012) study is thus a contribution in clarifying, through empirical investigation, the theory underlying theory-driven evaluation. Similarly, Smits and Champagne (2008) focused on the practical participatory evaluation model and performed an assessment of its theoretical underpinnings by referring to empirical evidence from several documented cases. The authors developed an elaborate list of assertions about how the model is supposed to work in practice and used existing applications of the model to challenge those assertions. Their study revealed that the support for the model was primarily based on theoretical arguments rather than on empirical grounds. In a similar vein, Miller and Campbell (2006) examined 47 cases of empowerment evaluation and determined whether these applications aligned with the principles underlying the evaluation model. Their purpose was to clarify a number of key points in the scholarly debate about empowerment evaluation and to pinpoint aspects that define the model. Therefore, the authors examined how evaluators in the 47 cases implemented the model, what the recurrent features were and how they compared to ten tenets of the empowerment evaluation model. Miller and Campbell's findings suggest that the evaluation cases frequently did not embody the core principles that underpin empowerment evaluation. Miller and Campbell's (2006) study clarifies important aspects relating to the theoretical debates about empowerment evaluation. Coryn, Noakes, Westine and Schröter (2011) used the same approach to examine 45 empirical applications that used the Theory-Driven Evaluation Model. Their study presented evidence to repudiate some aspects of the model and substantiate other aspects.

All of the aforementioned studies have examined the assumptions of a single evaluation model on empirical grounds. That is, the studies share a common goal: clarifying the tenets of a single evaluation model. As such, since the theoretical dimensions investigated in each study are peculiar to the specific model examined, it is not easy to compare the findings from one model to another and to use those comparisons to understand what works best in different program evaluation models and under what conditions. Therefore, I argue that adopting a similar approach to studying the theory-practice relationship in relation to program evaluation models is beneficial but not sufficient.

The second approach used to investigate the theory-practice relationship aims to develop the logic models underpinning evaluation models and using them to make comparisons between the various program evaluation models and their practice. Logic models are visual tools that represent how a program evaluation model is intended to work in the field (Hansen, Alkin, & Wallace, 2013). These visualizations portray the connections between the various

components of evaluation models: assumptions, evaluation contexts, evaluation activities, evaluation activities and external factors. Logic models are still new to the field of program evaluation and only few studies reportedly used these models to investigate the theory-practice relationship (see for example Dillman, 2013; Luskin & Ho, 2013; Vo, 2013). While the uses of logic models have potential to mediate knowledge transfer from practice to theory and vice-versa, they are not yet used to their full potential. For example, Dillman (2013) used logic models to compare the practices in three program evaluation models: Practical Participatory Evaluation, Values-Engaged Evaluation and Emergent Realist Evaluation, but did not use empirical data to support the claims from the comparisons. It might be argued that empirical data could be used to refine the development of the logic models, and thus improve the theoretical underpinnings of an evaluation model based on instances of practice. However, because this approach is still new, the extent to which logic models can portray the complex connections between the various components of an evaluation model is, as yet, not well established.

The third approach to studying the theory-practice relationship in relation to program evaluation models consists of using theorists' and practitioners' self-reported practices to examine whether and how practice is connected to theory. Christie (2003) initiated this research tradition and collaborated with eight prominent evaluation theorists in developing the theory-to-practice instrument. The instrument is made up of 38 items related to: the methods advocated by evaluation models; values embedded within evaluation models; and uses of the evaluation results. The 38 items are generic and not specific to any single model. The instrument was administered to the eight theorists as well as to 138 evaluators. Participants were asked to indicate their position concerning each item on an eleven-point Likert scale, with responses ranging from: "This statement is very similar to how I conduct evaluation" to "This statement is very dissimilar to how I conduct evaluation". The responses indicated how these evaluation theorists and practitioners attempt to undertake evaluation using their preferred model of program evaluation. Christie's (2003) study aimed to investigate and compare the reported practices of different evaluation theorists and practitioners. Her finding revealed that "only 36 % of the evaluators were within meaningful proximity of a theorist, indicating that most do not use frameworks aligned with a specific theoretical model" (p. 33).

Similar to the Christie (2003) study, Fitzpatrick (2004) interviewed prominent program evaluation theorists to investigate how they implemented their respective models. Fitzpatrick

(2004) engaged in prolonged reflections with each of the theorists about a single case application of their evaluation model. Although the cases were not the same for each theorist, the questions asked by Fitzpatrick were similar across interviews. The questions were meant to provide insights into the decisions made by the various evaluators and how these decisions were influenced by the context. The purpose of that exercise was twofold: first, to compare the evaluation practices of various theorists along pre-defined dimensions; and second, to link them to the respective theories and examine the connection between the theorists' theories and their practice of their respective models.

Another example of where self-reported practices are used to investigate the association between theory and practice in relation to program evaluation models is presented in the 2005 special issue of *New Direction for Evaluation*. The editors, Alkin and Christie (2005), presented a hypothetical evaluation scenario to four prominent theorists (see Donaldson, 2005; Greene, 2005b; Henry, 2005; King, 2005). They then asked each theorist to describe how he/she would ideally design and conduct that evaluation case. The editors analyzed the theorists' proposed evaluations and developed themes related to the influence of theory and context on practice. In this exercise, the editors used self-reported practice to examine how, hypothetically, theories were enacted in practice and to draw comparisons between different models of evaluation.

The studies by Christie (2003), Fitzpatrick (2004), and Alkin and Christie (2005) are useful for examining if the theory and practice of different evaluation models are aligned. Moreover, unlike the studies grouped under the first approach, they potentially offer the additional insight about how different models can be compared from both theoretical and practice-based perspectives. Nevertheless, the three studies are deficient in that they are based on reported rather than actual practices of evaluators, and the studies do not capture the extent to which the reported practices are congruent with actual practices. In that respect, Alkin and Christie (2005) critiqued their own work and explained that, since evaluation is situational, it is affected by the context of implementation which offers "its own constituency, set of values, programmatic elements, bureaucratic hurdles and other variables" (p. 3) and, therefore, the study of reported practices could be quite different from that of actual practices.

2.1.6 Miller's framework for investigating the theory-practice relationship

As discussed in the previous section, studies that examine how the theory and practice of program evaluation models are connected often do not address the goal of providing insights

into evaluation models both singularly and comparatively. This observation could be explained by noting the absence of a framework that enables researchers to define and empirically examine models both singularly and comparatively. Miller's (2010) framework is a step towards overcoming this obstacle.

In her attempt to develop a set of standards to empirically examine the theories underlying program evaluation models, Miller (2010) postulated a framework consisting of five criteria: operational specificity, range of application, feasibility in practice, discernible impact and reproducibility. These criteria will be explored at length later in this chapter.

Miller's framework is a valuable development in the field as it offers a guideline for considering how theory might be examined through practice. My interpretation of her framework is illustrated in Figure 2.1.

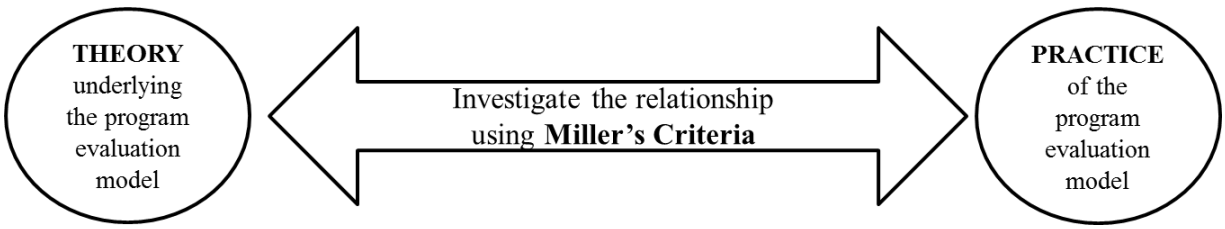


Figure 2.1 – Interpretation of Miller's (2010) framework

As illustrated in Figure 2.1, Miller's (2010) criteria represent a lens through which the researcher can interpret and give meaning to the alignment or misalignment between any evaluation model and its practice. That is, instead of merely reporting the areas where convergence and divergence occur between the theory and practice of an evaluation model, the researcher can use Miller's (2010) criteria to describe the extent to which each of the five criteria is met. The researcher can thus make claims about whether the program evaluation theory provides enough guidance to the evaluator, or whether the theory provided enough information about how and when the model is best applied, and so forth. Importantly, Miller's (2010) criteria can be used to study different models using the same standards, thus enabling comparative analyses. Nevertheless, what would enhance and facilitate Miller's formulation is a tool for defining the theory underlying a program evaluation model and informing how data collection from the context of practice ought to be carried out. Such a tool could provide researchers with a comprehensive and comparable description of theoretical aspects of a model, to which Miller's (2010) criteria are applied. In the following

section, I introduce my formulation of a tool to use for this purpose, in the context of a conceptual framework for investigating the theory-practice relationship relevant to program evaluation models.

2.2 A conceptual framework for investigating the theory-practice relationship

Expanding on the work of Miller (2010), I have synthesized from the literature a theoretical tool that I believe helps clarify the theory underlying a program evaluation model and organize data collection about its practice. I have integrated this tool – which I call the Program Evaluation Models' Essential Dimensions (PEMED) – within a conceptual framework for investigating the theory-practice relationship relevant to program evaluation models. The conceptual framework that I propose is illustrated in Figure 2.2.

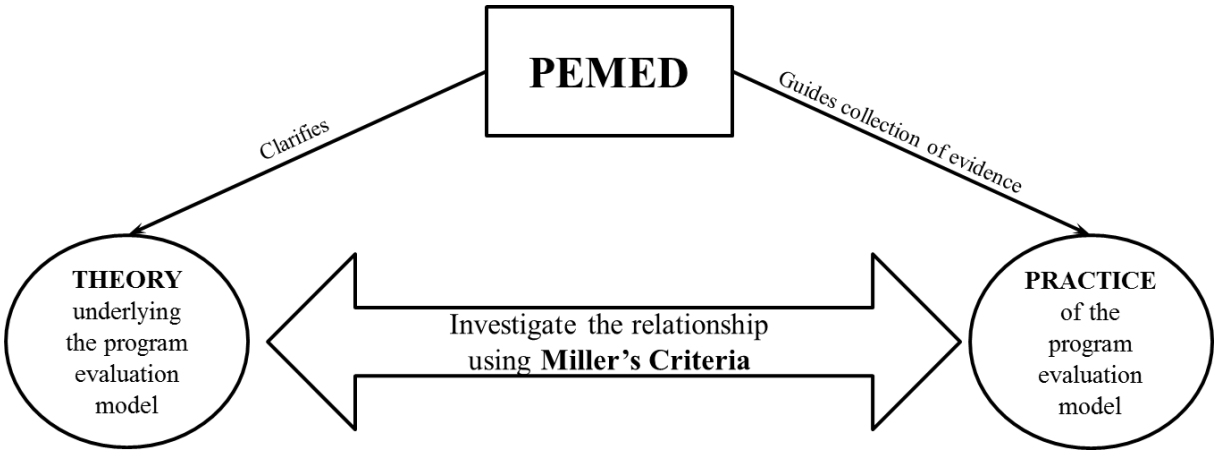


Figure 2.2 – Conceptual framework for investigating the theory-practice relationship

As is demonstrated in Figure 2.2, the conceptual framework comprises two elements: the PEMED and Miller's (2010) criteria. The PEMED serves two purposes: first, it provides a structure that helps to clarify the theoretical propositions of an evaluation model; second, it provides guidance for the organization and execution of data collection about the practice of the model. Miller's (2010) criteria are used to describe and give meaning to the relationship between the theoretical propositions and the evidence derived from the practice of the model. In the following section, I describe each of the two elements of this conceptual framework in depth.

2.2.1 Element 1: The Program Evaluation Models' Essential Dimensions (PEMED)

I developed the PEMED as a theoretical tool to help define a program evaluation model's underlying theory. Program evaluation models can differ based on the philosophy of evaluation they sustain, the role of the evaluator they prescribe, the questions they address, the methods they advocate, the audiences they serve, the needs they attend to, the degree of involvement of stakeholders they propose, the methods of dissemination of results they suggest and so forth. It is, therefore, worthwhile exploring what dimensions are essentially needed to systematically define models. In constructing the PEMED, I explored the literature and developed a tentative list of dimensions that define program evaluation models systematically and comprehensively. The PEMED comprises six dimensions, which are presented and defined in Table 2.2.

Importantly, while the PEMED was conceived as a tool to explicate program evaluation models comprehensively, not all models are necessarily conceptually well defined (Miller, 2010; Smith, 2008). Therefore, not all models can be articulated clearly across all the dimensions of the PEMED. In that respect, the possible uses of the PEMED can be extended from helping researchers define their program evaluation models to identify where and how these models need further clarification.

Table 2.2 – Program Evaluation Models' Essential Dimensions (PEMED)

Dimension	Definition
1. Views about social programs	Propositions about how social programs develop, improve and/or change
2. Paradigm of evaluation	Propositions about the act of evaluation, particularly regarding its ontology, epistemology and methodology
3. Views about utilization	Propositions related to how, when and by whom evaluations and their outcomes are used
4. Purpose of evaluation	Propositions about what purposes are best served using this model of program evaluation
5. Scope of the evaluation	Propositions about which aspects or parts of a program are best evaluated using this particular model
6. Operational procedures of the evaluation	Propositions about how to perform this evaluation and implement its strategies

The dimensions of the PEMED draw, in part, on the conceptual components of evaluation models identified by Shadish and his colleagues (1991). In discussing the lessons learned from evaluation practice and theory, the authors advanced five conceptual components that they deemed fundamental to the definition of any evaluation model. Their thesis was that

evaluators need to have a clearly identifiable perspective on: (1) “social programming” thus explaining how social programs change vis-à-vis social problems; (2) “theory of knowledge” and how evaluators construct value claims; (3) “theory of valuing” and what it means to engage in evaluation; (4) evaluation “knowledge use”; and (5) “evaluation practice”, including an understanding of the various strategies and tactical considerations that evaluators can use given the limitations they face (see Shadish et al., 1991, p. 32). These five components are integrated within the PEMED dimensions.

I have labeled the first dimension in the PEMED as “views about social programs”. This dimension contains the propositions of a model that describe how social programs develop, improve and/or change. This dimension is analogous to Shadish’s first conceptual component “social programming”. The second dimension, the “paradigm of evaluation”, includes two of the components discussed by Shadish et al. (1991) (i.e. theory of knowledge and theory of valuing) and is concerned with the ontology, epistemology and methodology of evaluation. Note that methodology is different from methods of evaluation: methods are the specific tasks and steps involved in an evaluation; methodology comprises assumptions which justify and rationalize the preference and use of particular methods. The third dimension, which runs parallel to Shadish’s knowledge use component, “views about evaluation utilization”, encompasses the assumptions related to how, when and by whom evaluations and their outcomes are used.

The fourth PEMED dimension, “purpose of the evaluation”, was postulated based on the analysis of four particularly important program evaluation metamodels: the chronological framework (Rossi, Lipsey, & Freeman, 2004), the alternative approaches framework (Fitzpatrick et al., 2004), Owen’s evaluation forms and approaches (2006), and Stufflebeam and Shinkfield’s classification (2007). These four metamodels posit that the choice of evaluation models depends, to a large extent, on the purpose of the intended evaluation, and that different evaluation models serve different purposes. This point is elaborated explicitly in Hansen’s (2005, p. 448) typology of program evaluation models, which is premised on the idea that models should “stipulate the question that a given type of evaluation seeks to answer”.

The fifth dimension of the PEMED is rooted in the definition of programs put forth by Yarbrough et al. (2011). The authors argue that a program evaluation does not necessarily need to evaluate all of a program’s components, and that several subcomponents could be

evaluated either separately or collectively. Congruently, with this insight, the fifth dimension of the PEMED is called “scope of the evaluation”.

All the dimensions outlined above are concerned with what Smith (2010) referred to as ideologies; that is, theoretical propositions about the nature of evaluation, its values and its practice. He stated that all program evaluation models can be conceived of as part ideology, part methodology and part intervention. The last and sixth dimension of the PEMED, “operational procedures”, reflects Smith’s insight that models are also procedural interventions, and parallels Shadish’s (1991) fifth component, “evaluation practice”.

2.2.1.1 Defining the dimensions of the PEMED

I postulate that six dimensions of the PEMED represent pillars for defining the theories underlying program evaluation models. These dimensions are not to be considered as independent from one another; in fact, some of them are locked into resonance with others. For example, the model’s assumptions about the paradigm of evaluation are expected to strongly influence the operational strategies that the model specifies. Though important, the nature of these resonance locks is a separate research agenda and transcends the scope of this dissertation.

In what follows, I present a comprehensive definition of each dimension and outline how each contributes to a better understanding of a program evaluation model. Where possible, I have presented a list of questions that evaluators can use to clarify their evaluation model’s underlying assumptions along the PEMED dimensions.

Dimension 1 – Views about social programs

Evaluation models make assumptions about the social function of a program and how the program is supposed to assume these functions. More often than not, the assumptions are tacit, implicit and blurred. Nonetheless, according to Shadish et al. (1991), the assumptions can and do impact evaluators’ practices and, therefore, their nature and content must be examined and made as explicit as possible before an evaluator undertakes an evaluation. For example, if the assumption is that a social program is a vehicle for social improvement and betterment, the implication is that evaluation practices that emphasize empowering stakeholders and building their capacity are to be considered; this is the case with empowerment evaluation (Fetterman & Wandersman, 2005). On the contrary, if social programs are viewed as tools for achieving uniformity and conformity, more standards-driven evaluation approaches are likely to be in place.

Shadish et al. (1991) surveyed the literature extensively to summarize the important aspects of a program evaluation model that evaluators need to clarify prior to engaging in evaluation. The authors suggest that evaluation models should indicate how social programs address problems, whether incrementally or radically. The models should also specify to what extent programs are to be examined in relation to their political and organizational contexts.

Dimension 2 – Paradigm of evaluation

Guba (1990, p. 18) defines the term "paradigm" as an assortment of three related sets of beliefs: ontology, epistemology and methodology: the ontological assumptions explain the nature of the knowable; the epistemological assumptions explain the relationship between the knower and the known; while the methodological assumptions explain how the knower inquires about what can be known. Taking the definition of Guba (1990) on board, the paradigm of evaluation is, thus, a coherent set of responses to the following questions:

1. What does the evaluation model imply about the nature of the evaluation act?
2. What does the evaluation model imply about the relationship between the evaluator and the act of evaluating?
3. What does the evaluation model imply about the best method(s) to engage in evaluation?

Shadish et al. (1991) addressed the ontological question by labeling it "theory of valuing" (p. 455) and the epistemological question by labeling it "theory of knowledge" (p. 463). Importantly, they identified relevant sets of questions that help evaluators explain the paradigm of evaluation underlying their program evaluation model of choice. These questions are presented in Table 2.2. I have expanded on these questions to include additional considerations regarding the paradigm of evaluation underlying program evaluation models. For example, under the methodological category, I have grouped assumptions about the role of the evaluator and other stakeholders in the evaluation processes. The rationale for this grouping is that the roles of the evaluator and other stakeholders essentially impact which methodologies are favored. Furthermore, in relation to the role of the stakeholders, I have used the three components proposed by Cousins and Whitmore (1998) that aim to identify how the idea of collaboration is framed in an evaluation model. These components are: "control of the evaluation process, stakeholder selection, and depth of participation" (Cousins & Whitmore, 1998, p. 10). Thus, the "paradigm of valuing" dimension aims to clarify how the evaluation model suggests allocating the responsibilities of the various stakeholders with regards to how much decision-making they can exert while voicing their ideas about

technical matters related to an evaluation. Additionally, this dimension clarifies the model's propositions as to the range of stakeholder groups that need to be involved in the evaluation, as well as the degree of involvement of the stakeholders, which could range from mere consultative sources of information to legitimate partners in all aspects of the evaluation design. The resulting set of questions evaluators could ask regarding an evaluation model in order to clarify the model's underlying assumptions along the dimension Paradigms of valuing is presented in Table 2.3.

Table 2.3 – Questions relating to the dimension "Paradigm of evaluation"

Assumption		**Related questions**
Ontological assumptions		What does the evaluation model imply about the nature of the evaluation act?
Epistemological assumptions		What kind of knowledge is privileged in this model? Can the evaluator produce this kind of knowledge?
Methodological assumptions	Methods for data processing	Which methods are advocated in the model as most suitable for the conduct of the evaluation?
	Role of the evaluator	What does the evaluation model imply about the role of the evaluator?
	Role of the stakeholders	What does the evaluation model imply about how stakeholders are selected? What does the evaluation model imply about the extent to which stakeholders have control over the evaluation process? What does the evaluation model imply about the depth of involvement of stakeholders in the evaluation?

Dimension 3 – Views about utilization

Christie (2007) argues that evaluation utilization is by far the most substantially researched area in the program evaluation literature. The two terms, "evaluation use" and "evaluation utilization" are used synonymously in the literature to refer to the ways in which evaluation processes and findings are used (Alkin & Taut, 2003).

Evaluation use has long been restricted by evaluators to reflect a measurement of impact. That is, it was used as a judgment of how the evaluation outcomes were used to inform decision-making and action taking. This concept is referred to in the literature as instrumental use (Greene, 1988; Shulha & Cousins, 1997). Increasingly, more scholars have come to recognize other kinds of use, though "impact" as part of the concept of "use" continues to be

acknowledged. Patton (1998) was the first to introduce the term “process use” to emphasize the “individual changes in thinking and behavior and program or organizational changes in procedures and culture that occur among those involved in evaluation as a result of the learning that occurs during the evaluation process”. Alkin and Taut (2003) built on Patton’s definition to draw a distinction between process use and use of evaluation findings, the latter including instrumental and symbolic use. While instrumental use refers to the direct use of evaluation knowledge and findings, symbolic use occurs when people make claims or arguments based entirely on the mere existence of an evaluation.

Shadish et al. (1991) introduced the idea of communication of evaluation information with the various stakeholders as a subcomponent of the theories of use that are supported by program evaluation models. The authors advised that explaining the ways through which evaluation information is shared by an evaluator, and the frequency and form of these communications constitutes a part of the espoused theory of use that every model adopts. In a subsequent publication, Shadish (1998) suggested additions to the definition of the dimension, “use”, to enrich its content, making a distinction between long-term versus short-term uses.

Table 2.4 summarizes the questions that evaluators could ask regarding an evaluation model to clarify underlying assumptions along the dimension “views about utilization”.

Table 2.4 – Questions relating to the dimension "Views about utilization"

Assumption	Related questions
Intended uses	What does the evaluation model imply about how evaluation findings are to be used?
Processes to foster use	What strategies are suggested in the evaluation model to foster the use of evaluation findings and/or processes?

The concept of “evaluation use” should not be confused with that of “evaluation influence”. The latter is explained by Kirkhart (2000, p. 7), as “the capacity or power of persons or things to produce effects on others by intangible or indirect means”. While both concepts of use and influence are potential consequences of evaluation practice, the first includes ways in which evaluation information and processes intentionally or unintentionally impact the evaluand while the second refers to the impacts that evaluators may not be even aware of (Kirkhart, 2000).

Dimension 4 – Purpose of the evaluation

Pertinent literature suggests that not all approaches to program evaluation share a common purpose. In fact, various scholars have developed classifications of program evaluation models based entirely on a distinction of purposes that these models serve. For example, Hansen (2005, p. 458) identifies six types of models that differ based on the questions they address and the purposes they serve: "results models, process models, system models, economic models, actor models and programme theory models".

In order to fully appreciate how the dimension of purpose defines program evaluation models, I argue that one must examine a model's propositions regarding: (1) who sets the purposes of evaluation; (2) how these purposes are set; and (3) what these purposes are and, consequently, what major questions the model attempts to address.

Table 2.5 summarizes the questions that evaluators could ask regarding an evaluation model to clarify underlying assumptions along the dimension "purpose of evaluation".

Table 2.5 – Questions relating to the dimension "Purpose of evaluation"

Assumption	Related questions
Defining the purposes of the evaluation	What are the purposes of the evaluation model?
	Who sets the purposes of evaluation in the evaluation model?
	How are these purposes set in the evaluation model?
Outcomes of the evaluation	What are the projected outcomes of the evaluation model?
Timing the evaluation	What does the model suggest as to when the evaluation can take place in relation to the state of the program?

I have also added a consideration of the different kinds of outcomes that could be expected out of an evaluation. The information needs of those who commission and evaluation and those involved in an evaluation are imperative to defining the purpose of that evaluation. According to Owen (2006), the outcomes of an evaluation, which are conceptually different from the uses of an evaluation, encompass one or more of the following four constructs. First, outcomes include evidence, which is the data collected during the evaluation. This is a part of all types of evaluation because evaluation by itself is a systematic inquiry devised to generate evidence about something. Second, outcomes include conclusions that are developed as a synthesis of the collected data and information. Third, outcomes can include judgment. However, not all evaluations develop judgments. Judgments are the type of outcomes that

result from “placing value on conclusions thus stating that the program is ‘good’ or ‘bad’, or that the results are ‘positive’, ‘in the direction desired’, or ‘below expectations’” (Owen, 2006, p. 20). Fourth, outcomes can include recommendation; though, not all evaluators develop recommendations about “suggested courses of action, advice to policy-makers, program managers or providers about what to do in the light of the evidence and conclusions” (Owen, 2006, p. 20). It is, therefore, imperative to clarify what the outcomes of the evaluation are and to make sure not to expect more – or less – from any selected model. It is also important to communicate this information to everyone involved in the evaluation to avoid disappointments and mistrust.

Additionally, I have added a consideration of the timing of the evaluation in relation to the program’s life and development. Timing affects (and is affected by) the purpose of the evaluation. Owen (2006) presents two useful attributes for defining programs that deserve particular attention from evaluators: a program’s “level” (p. 27) and its “state of existence” (p. 51). According to Owen (2006), a program can exist in one of three comparatively defined levels. At the broadest level, the mega or corporate level, a program is set by the top level decision makers in an organization, for example at the level of the department of education in a country. At this level, Owen (2006) explains, the design of programs is more likely to be presented in overall economic terms or in relation to the social impact it is expected to make. At the second level, the macro level, the program design reflects the input by divisions or regional groups or departments within an organization. At the third level, the micro level, a program is restricted to the work of small units or even individuals within an institution or department. Owen (2006) argues that the level at which a program is conceived, developed and implemented has direct implications on the purpose of evaluation models because different organizational levels have different priorities, modes of operation and information needs.

Moreover, the life span of a program, or its “state of existence” as Owen (2006) calls it, also has implications for the purpose of an evaluation. For example, if a program is still underway, the information needs of its creators differ from those of the creators of a recently terminated program. While the former group could be asking for a cost-effectiveness analysis, the latter is more likely to ask for information about the impact of the program. Therefore, since different evaluation information is needed for different states of existence, it is imperative to understand what the evaluation specifies with regard to the state and level of a program model because this will shape the purposes of the evaluation.

Dimension 5 – Scope of the evaluation

According to Yarbrough et al. (2011), program evaluations can vary in complexity depending on which aspects of a program they target, which can include one or a combination of the following:

> contexts and how they interact with programs and program components, participants and other beneficiaries as well as those who encounter costs or loss of benefits, needs, problems, and policy spaces in programs and their contexts, goals and objectives, resources and costs of all kinds, including staff, facilities, materials, and opportunity costs, activities, procedures, plans, policies, and products, logic models, beliefs, assumptions, and implicit and explicit program theories explaining why and how programs should work, and/or outputs, results, benefits, outcomes, and impacts. (p. xxiv)

Shadish et al. (1991) also contend that, while changes at the level of an entire program often produce more impact than changing subcomponents of the program, the former is less likely to occur. It is important then that a program evaluation model specifies which aspect(s) of a program are included in the evaluation.

Dimension 6 – Operational procedures of the evaluation

This dimension is included in the PEMED because evaluation models need to provide specific guidelines as to how they are best implemented. Accordingly, the models need to have advanced organizers that indicate what type of data is to be collected. They should also include: descriptions of the steps taken by the evaluator to implement the evaluation model; the strategies employed for collecting evidence and other information; and the strategies used for the data analysis, interpretation and reporting.

2.2.2 Element 2: Miller's criteria

While the PEMED helps to clarify the model's theoretical propositions and collect data along the six dimensions, Miller's (2010) criteria are used to interpret the relationship between the theory and practice of a program evaluation model, and make assertions about the model's characteristics, such as its operational specificity, its range of application, its feasibility, and its impact. A summary of Miller's (2010) criteria used in this study is presented in Table 2.6.

Table 2.6 – Miller's criteria for examining a program evaluation model's theory in practice

Miller's (2010) criteria	Operational definitions
Operational Specificity	The extent to which theory provides clear guidance for practitioners
Range of Application	The extent to which theory describes under what contextual circumstances the model is applicable and feasible.
Feasibility in practice	The extent to which theory can be applied in practice
Discernible impact	The extent to which theoretically conceptualized impacts are in fact happening in practice
Reproducibility	The extent to which the use of the model and its impacts can be reproduced over time

The first criterion, "operational specificity", invites researchers to examine whether the program evaluation model in question offers enough guidance for evaluators. Miller (2010) explains that such guidance includes recommendations and admonitions regarding:

> when, how, and what evaluation questions are identified and prioritized; who participates in each stage of the evaluation process; what role the evaluator assumes; what methods are ideal; how values underlying the theory are best enacted; and how plans for using the evaluation process and its results are considered. (pp. 391-392)

The second criterion, the "range of application", concerns an examination of the "limits of the theory's application" (Miller, 2010, p. 393). Miller argues that the study of evaluation models needs to unveil under what contextual circumstances a program evaluation model can be applied so as to achieve its purposes. The third criterion, the "feasibility in practice" refers to an examination of how doable and easy it is to apply the operations advocated by the model's underlying theory. Miller (2010, p. 394) argues that some models may be difficult, if at all possible, to implement because of the "role demands placed on evaluators". Smith (2010) argues that investigators need to examine other aspects – such as the design quality, effectiveness, efficiency and robustness, in addition to side effects – to gain a better appreciation of a model's feasibility. Moreover, Smith (2010) argues that these aspects enable researchers to appreciate the ways in which a model is superior to available alternatives. The fourth criterion in Millers' (2010) framework is "discernible impact". Since models are designed to achieve planned outcomes and uses, a crucial aspect of the empirical assessment of these models is the examination of whether and to what extent the model actually fulfills what it promises to.

A fifth criterion in Miller's (2010) framework is reproducibility. Miller (2010) explains that an important aspect of the empirical study of program evaluation models is to determine whether the use of the model and its impacts can be reproduced over time. Therefore, to be able to investigate reproducibility, a researcher must be able to compare the application of the evaluation model across various contexts. The fifth criterion in Miller's (2010) framework extends beyond the boundaries of this study. Indeed, since this study comprised a single application of the evaluation model, I did not consider it in this research.

2.3 Conclusion

In this chapter, I have explored the program evaluation landscape, clarifying important concepts and identifying appropriate terminology and definitions. Also, I have built on existing studies to develop a conceptual framework for investigating the relationship between the theory and practice of program evaluation models. In the following chapter, I establish the appropriateness of my choice of the 4GE model of evaluation to the study context using the PEMED. I also present my interpretation of the 4GE model and adaptations made to it.

CHAPTER 3: ESTABLISHING THE APPROPRIATENESS OF AND DEFINING THE A4GE

The role of the evaluator starts well before carrying out a particular evaluative inquiry (Skolits, Morrow, & Burr, 2009). Evaluation work starts with selecting and rationalizing the choice of a program evaluation model, a process that requires sound background knowledge of available theories, approaches and models of evaluation, as well as a keen understanding of the context of the evaluand. Only after this stage can implementation take place, accompanied by rigorous monitoring of the evaluation activities.

In this study, I decided to use and adapt Guba and Lincoln's 4GE to evaluate a Science Teacher Preparation Program. In this chapter, I present my approach to checking the appropriateness of the 4GE for the study context and discussing the adaptations I made to the model. To this end, I first discuss the literature relating to how evaluators can define the context of an evaluation and use it to check the appropriateness of their selected program evaluation model. I then use my situational analysis of the study context to justify the appropriateness of the 4GE. Additionally, harnessing the lessons learned from previous empirical studies on the application of the 4GE, I present my adaptations of the model and use the PEMED to describe and outline its theoretical underpinnings.

3.1 Establishing the appropriateness of a particular program evaluation model to the context of implementation

Educational program evaluation is a multi-modal enterprise characterized by a wealth of models and approaches that serve a variety of purposes. One area of consensus among program evaluation scholars is that models are not in competition for superiority but are different means to different ends. From this perspective, diversity is an asset to the discipline. Nonetheless, diversity also poses many challenges to evaluators in choosing a model that best suits a particular evaluation situation.

Making a decision about an evaluation model can be carried out in one of two ways. Evaluators can either analyze the context of the evaluation then decide on a model, or they can choose a model and then check its appropriateness to the evaluation context. While the first way seems a more reasonable method, it is more difficult to implement given that few evaluation theorists have developed frameworks to guide and rationalize this kind of decision-making in evaluation practice (Kundin, 2010). The lack of comparative research about the different evaluation models (Henry & Mark, 2003b; Smith, 2008) adds further

challenges to this approach of choosing a program evaluation model. Alternatively, evaluators can decide on a model based on its characteristics and potential. They can then check the appropriateness – or inappropriateness – of the model in light of their analysis of the context of implementation. This analysis tells them whether they should proceed with the model or replace it with a different one. This way of deciding on an evaluation model is particularly useful for evaluators, like me, who are interested in carrying out research on the model. One pitfall of following this approach, however, is that evaluators might be biased to use their chosen evaluation model irrespective of a misfit between the context and the model.

That the context of practice determines the appropriateness of an evaluation model is widely accepted among members of the evaluation communities (Fitzpatrick, 2012). Mark, Greene and Shaw (2006) explain that, while it is only natural to have different evaluation practices across different contexts given the diversity in the field, these practices ought to fit and be congruent with the contexts of practice. Nevertheless, the challenge facing evaluators is to understand how the context of the evaluation is defined, captured and interpreted, and how it can be used to rationalize the decision-making involved in evaluation practice. Greene (2005a, p. 84) argues that scholars need to develop "a more sophisticated conceptualization and study of just how context matters in evaluation". In the following, I have surveyed the literature about how the context influences decision-making in evaluation practice and have used that survey to portray my attempt to capture the evaluation context in the present evaluation in order to establish the appropriateness of the 4GE.

3.1.1 Understanding the context of implementation of an evaluation

The literature on the role of context in evaluation practice dates back as far as the beginnings of program evaluation itself (Fitzpatrick, 2012). Perhaps the most comprehensive account of context was presented in the seminal work of Patton (2008), who argued that evaluators need to learn to be situationally responsive. Patton presents fundamental questions that guide evaluator's situational analysis. These questions are grouped under four categories:

1. Understanding the program: what are the program's primary goals? What are the strategies for attaining these goals? Who are the intended beneficiaries of the program's intervention? How has the program changed over time? What led to those changes?
2. Identifying primary stakeholders and their interests: what is the political context for the evaluation? Who will be the primary intended users of the evaluation?

3. Understanding the program's evaluation history: what prior experiences, if any, has the program had with evaluation? What are current monitoring and evaluation approaches, if any? How are monitoring and evaluation data currently used, if at all? What factors affect current uses?
4. Understanding the decision and action context: what decisions, if any, is the program facing? What are the timelines for such decisions?

Similarly, in discussing how context matters to evaluation practice, Greene (2005a) suggests that evaluation contexts can be defined in relation to five intertwined categories: (1) the descriptive and demographic aspects, (2) the material and economic features related to the quantity and quality of its resources, (3) the general organizational atmosphere that governs the processes of the program, (4) the types and natures of interpersonal interactions that frame relationships, and (5) the politics in a setting, and identification of power and/or influence hubs.

Another discussion of the context and how it is defined in the practice of program evaluation comes from Tourmen's (2009) empirical examination of how experienced and less experienced evaluators make decisions about evaluation design. Tourmen's study revealed that evaluators must consider the characteristics of the program, stakeholders' views about evaluation, as well as the resources available for the evaluation to make decisions about the choice of a program evaluation model.

The perspectives of Patton (2009), Greene (2005a) and Tourmen (2009) are complementary in many ways. For example, all three scholars noted the need to understand the physical setting of the evaluand. This is evident in the first category in each of Patton's and Greene's lists, although Patton's questions suggest an added element that relates to the history of the program and its development. Also, Tourmen's findings suggest that experienced evaluators have defined this dimension in relation to two core concepts: the characteristics and the state of the program. From analysis of the similarities and differences across the three works, I generated a comprehensive list of all the factors that the evaluators need to examine as part of their situational analysis of the context that embeds the program evaluation. These factors are outlined in Table 3.1.

Table 3.1 – Factors guiding the situational analysis of the context

Factor	Description
Context of the evaluand	***1.*** *Description of the evaluand:* ***a.*** What are the evaluand's goals? ***b.*** What are the strategies for attaining these goals? ***c.*** Who are the intended beneficiaries of the evaluand's intervention? ***2.*** *The state of the evaluand:* ***a.*** How has the evaluand changed over time and what led to those changes? ***b.*** What decisions, if any, is the evaluand facing? What are the timelines for any such decisions? ***3.*** *Organizational setting of the evaluand:* ***a.*** How can the organization that houses the evaluand be described? ***b.*** Who are the stakeholders involved in the development, delivery and monitoring of this evaluand?
Context of the evaluation	***1.*** *Evaluation background:* ***a.*** Why is this evaluation commissioned? By whom? ***b.*** What are the existing monitoring and evaluation approaches, if any? How are they currently being used? ***2.*** *Evaluation resources:* ***c.*** What resources are available for the evaluation? ***3.*** *Evaluation leverage:* ***d.*** Can the evaluation stimulate change? At what levels? ***e.*** Who will be the primary intended users of the evaluation?
Context of the evaluator	***1.*** *Is the evaluator commissioned? By whom?* ***2.*** *Is the evaluator an insider or an outsider to the host institution or program?* ***3.*** *What resources are at the evaluator's disposition and what constraints are imposed on the evaluator?*

3.1.2 Using situational analysis to establish the appropriateness of a program evaluation model to the context of implementation

The situational analysis of the context of implementation of an evaluation allows the evaluator to explain HOW the contextual factors rationalize the choice of a program evaluation model. To this end, the PEMED proves to be particularly useful. Using the PEMED as a theoretical organizer, the evaluator is able to look through the various theoretical propositions of an evaluation model and analyze whether these propositions align with the conditions set by the contextual factors. This type of analysis provides a rationale for either using the program evaluation model or replacing it by another that better suits the context.

To illustrate the point, take, for example, the PEMED dimension “views about utilization”. This dimension is used by the evaluator to clarify how evaluation findings are meant to be used by a specific model. To be able to justify the selection of that model, the evaluator would, at least, consider the organizational setting of the evaluand (that is, how are decisions made about the evaluand? What decisions, if any, is the evaluand facing? What are the timelines for such decisions?), and the context of the stakeholders (who are the stakeholders involved in the development, delivery and monitoring of this evaluand? who will be the primary intended users of the evaluation?). These factors have implications about how evaluation can and will be used. For example, if the institution that hosts the evaluand makes decisions based on statistical reports only, it would be unwise to choose an interpretive model of program evaluation because it is unlikely that its findings will be used. Similarly, if the primary intended users of the evaluation are science educators, then the evaluator must choose a model that provides them with some kind of use (e.g. instrumental or process use).

3.2 Establishing the appropriateness of the adapted version of the 4GE to the study context

In this section, I argue for the appropriateness of an adapted version of the 4GE for evaluating the STPP in this study. To this end, I first present a description of the 4GE using the PEMED and highlight the adaptations that I made to it. Then, following the approach described in the previous section, I check the appropriateness of the adapted version of the 4GE (which will be referred to as the A4GE) with reference to the situational analysis of the context of implementation in this study.

3.2.1 Describing and defining the adapted fourth generation evaluation

Guba and Lincoln (1989) named their approach the Fourth Generation Evaluation in contrast with three predecessor generations, which adopted positivism and post-positivism as their foundational paradigms. They identified those generations through their focus on measurement, description and judgment, respectively, and argued that the fourth generation evaluation, being based on the interpretive paradigm, is an advance on those, in that it seeks to acknowledge the values and experiences of the people involved in an evaluation. Guba and Lincoln (1989) present their view of evaluation, the 4GE, as being organized around the principles of interpretive inquiry. The authors present a thorough treatise on their evaluation model in a seminal book entitled, *Fourth Generation Evaluation* (Guba & Lincoln, 1989).

The adaptations I made to the 4GE are based on my attempt to (1) overcome the challenges imposed by a face-to-face negotiation setting, and (2) modernize the approach by taking advantage of the technological developments and by enhancing some of its processes, such as increasing the degree of participation of stakeholders and enhancing the transparency of the evaluation processes. In particular, I argue for the need to operationalize the model differently to enhance some of its features, such as the transparency of the process and the depth of stakeholder participation. The major modifications to the 4GE include improvements to the dynamics of information sharing and dissemination, enhancement of the context of the negotiation and its implementation, and clarification of the type of information sought from the evaluation. I argue that these enhancements do not comprise a new model because the core values underlying the 4GE are the same. In fact, the A4GE still holds constructivism as its guiding paradigm, adopts hermeneutics as its methodology, and focuses on the claims, concerns and issues of stakeholders as its advanced organizers. I call it adapted because it has been modernized and improved.

In this section, I present the A4GE's theoretical propositions along the various PEMED dimensions. While my purpose here is to use the theoretical propositions together with my analysis of the context to rationalize the appropriateness of the A4GE for the study context, these theoretical propositions will also be used extensively in Chapter 6 where I investigate the relationship between the theory and practice of the A4GE. Because I adapted the 4GE model, I will present, where applicable, the rationale for these adaptations as well as their implications for the model's theoretical propositions.

3.2.1.1. Views about social programs

The propositions of the 4GE along the dimension "views about social programs" have not been modified. Accordingly, all the propositions entertained by the 4GE are maintained in the A4GE.

The 4GE and therefore the A4GE hence make several propositions about social programs and change that deviate significantly from previous evaluation generations. Perhaps the most fundamental difference is its embrace of the role of values at the center of program design and implementation, as well as evaluation. Indeed, resting on the assumption that every modern society is value-pluralistic and that people's interpretations are inextricably linked to the particular contexts within which they operate, the A4GE posits that programs are shaped by their context of implementation and, therefore, they can only be understood and evaluated

within the constraints of their cultural, physical and psychological contexts. In that respect, the 4GE authors hold the position that programs are not static and that "when they are introduced into a particular context they will be at least as much affected (changed) by that context in as much as they are likely to affect the context" (Guba & Lincoln, 1989, p. 45).

Moreover, the 4GE authors posit that program development is a continuous process that cannot be engineered: "it is a nonlinear process that involves the infusion of new information and increased sophistication in its use into the constructions of involved human constructors" (Guba & Lincoln, 1989, p. 109). Accordingly, social programs develop based on the reconstruction and redefinition of a program's goals and processes in ways that are sensitive to the stakeholders' concerns, experiences and values. Importantly, the A4GE implies that information and education are strong and powerful determinants of change.

Finally, the interpretive nature of the A4GE implies that social programs are never ideal. Rather, they are at best the most sophisticated interpretation of how to enact a particular policy in a specific contextual and temporal frame. To the extent that these frames evolve, the program needs to be revised accordingly.

3.2.1.2 Paradigm of evaluation

In common with the 4GE, the A4GE is rooted in the interpretive paradigm, adopting a stance that ideas are constructions of people and depend largely on the value systems of their holders. It holds relativism as its ontological assumption, which posits that human sense making is an act of interpretation, independent of any foundational reality (Guba & Lincoln, 1989). As such, the A4GE relies heavily on the realities created by people in ways that are largely influenced by the contexts in which those people are operating as well as their backgrounds and values. The epistemological assumption underlying the A4GE is that of transactional subjectivism. Therefore, assertions about reality and truth depend solely on the meaning sets (information) and degree of sophistication available to the individuals and audiences engaged in forming those assertions (Guba & Lincoln, 1989). The methodology of the A4GE is hermeneutic-dialecticism. To this end, the constructions (interpretations) entertained by the involved individuals and groups (stakeholders) are first uncovered and plumbed for meaning and then confronted, compared and contrasted in encounter situations (Guba & Lincoln, 1989).

Like the 4GE, the A4GE is classified as a responsive evaluation in recognition of the way it focuses the evaluation, determines what questions are to be asked and what information is to

be pursued. Accordingly, the outcomes of the A4GE are constructed through an interactive, negotiated process involving stakeholders.

Both the 4GE and the A4GE frame evaluation as a "joint collaborative process" (Guba & Lincoln, 1989, p. 253). Therefore, the A4GE defines the role of the evaluator as "orchestrator of a negotiation process that attempts to culminate in consensus on better informed and more sophisticated constructions" (Guba & Lincoln, 1989, p. 45). Furthermore, Guba and Lincoln (1989, p. 264) argue that, while the evaluator maintains much of the traditional roles, he or she is still "a technician but as a human instrument and data analyst; describer, but as a historian and illuminator; and a judge, but as an orchestrator of the judgmental process". Additionally, the evaluator is responsible for developing a partnership with other stakeholders so that they jointly create a construction of the evaluand. Guba and Lincoln (1989, p. 11) argue that an essential role of the evaluator is to interact with stakeholders in a manner respecting their dignity, integrity and privacy to the "level of full participative involvement in which the stakeholders and others who may be drawn into the evaluation are welcomed as partners in every aspect of the design, implementation, interpretation and resulting action of an evaluation – that is, they are accorded a full measure of political parity and control".

Finally, concerning the theoretical propositions of the 4GE about the selection of stakeholders, the model posits that stakeholders consist of three groups that should all be present in the evaluation: The "agents" (Guba & Lincoln, 1989, p. 40) who are responsible for producing and implementing the evaluand; the "beneficiaries" (p. 40) who profit from using the evaluand; and the "victims" (p. 40) who are negatively affected by the use of the evaluand. Nevertheless, in the A4GE, I avoid the use of the term "victims" and replace it with "disadvantaged". I believe that the term "victims" bears a connotation that some stakeholder groups are intentionally put at a disadvantage or that they are victimized – most likely by the "agents". Acknowledging that educators (among other program providers) are professionals who ideally strive to offer the best they can to the majority of student teachers (or other program recipients), I think it is unfair to use terminologies that suggest that they are the agents of a victimization process.

3.2.1.3 Views about utilization

The theoretical propositions of the A4GE along the dimension "views about utilization" overlap with those of the 4GE. The A4GE proposes that evaluation should achieve education as its purpose, as is explicitly stated in one of its authenticity criteria, the "educative

authenticity" (Guba & Lincoln, 1989, p. 248). Accordingly, one use of the A4GE is process use. Indeed, the model assumes that stakeholders involved in the evaluation are part of a teaching and learning process, and that, as a result of their involvement, their understanding and appreciation for the various constructions about the program should be enhanced.

The A4GE is also designed to help decision-making about what courses of action should be taken as a result of the development of the joint construction(s). This means that the A4GE assumes that stakeholders involved in the evaluation will make instrumental use of its outcomes. Guba and Lincoln (1989) argue that both catalytic authenticity (the extent to which the evaluation stimulates and facilitates action) and tactical authenticity (the degree to which stakeholders and participants in the evaluation feel empowered to act) should be achieved. Guba and Lincoln (1989, p. 10) argue that the 4GE must have "an action orientation that defines a course to be followed, stimulates involved stakeholders to follow it, and generates and preserves their commitment to do so".

To increase the likelihood of both kinds of use, the A4GE model encourages dissemination of evaluation reports at regular intervals throughout the evaluation. This is to keep stakeholders informed about the progress of the evaluation but also to invite them to review and analyze the constructions of others who are involved in the evaluation. Consequently, the A4GE evaluator should produce regular reports about interim findings and share them with the various stakeholders throughout the course of the evaluation.

3.2.1.4 Purpose of evaluation

In relation to the theoretical propositions along the dimension "purpose of evaluation", there are no adaptations that I have made to the 4GE and, consequently, there is total overlap between the propositions of the 4GE and those of the A4GE.

Accordingly, the purpose of the A4GE is to develop participants' constructions about the evaluand. These constructions are the interpretations of the various stakeholders about the claims and concerns, and resolved and unresolved Issues (CCIs) about the evaluand. Guba and Lincoln (1989) define CCIs as follows:

> a claim is any assertion that a stakeholder may introduce that is favourable to the evaluand; a concern is any assertion that a stakeholder may introduce that is unfavourable to the evaluand, and an issue is any state of affairs about which reasonable persons may disagree. (p. 40)

Essentially, the A4GE aims to create an avenue for stakeholders to understand, discuss and possibly resolve these CCIs in ways that are sensitive to the values and needs of the various

parties involved. As such, another purpose of the A4GE is to provide education about the evaluand so as to stimulate appropriate action. As Guba and Lincoln assert, in essence, "evaluation is a teaching/learning process" (Guba & Lincoln, 1989, p. 254).

According to the A4GE, social problems are only defined within a context through the interpretations of the people who inhabit this context. Consequently, the specific purposes of an evaluation are only set through a negotiation between the various stakeholders –including the evaluator – involved in an evaluation. Indeed, the 4GE authors maintain that evaluators are partners with stakeholders in the decision-making processes about the evaluation.

With respect to the projected outcomes of the A4GE, Guba and Lincoln (1989, p. 255) argue that the 4GE is "a process with unpredictable outcomes" and that "it is not possible to say with confidence what the outcomes of the evaluation will be – what we will know when it is completed or what value position we will take". They explain that, since the process does not guarantee the emergence of a single joint construction, there is no guarantee as to whether the evidence derived from the evaluation would yield a judgment of worth and merit, and about what. For example, it could be that two constructions emerge from the interpretation of the data and that while one leads to the belief that a particular aspect of the evaluand is, say effective, the other could be pointing in the exact opposite direction. So, essentially, the A4GE does not necessarily produce data in the form of conclusions, recommendations and judgments about worth and merit in the conventional sense of the word; rather the output is a compilation of CCIs.

Finally, with regard to the timing of the evaluation in relation to the program's development, Guba and Lincoln (1989) do not explicitly advocate a particular time in which the 4GE and, hence, the A4GE, is best undertaken. Nonetheless, due to its focus on identifying issues based on the experiences of stakeholders and their interaction with the program, the A4GE is preferably used during or after the program has been implemented.

3.2.1.5 Scope of the evaluation

Like the 4GE, the A4GE does not predetermine the focus and scope of an evaluation. Instead, these elements are determined through a negotiation between the various stakeholders involved in the evaluation. Consequently, any part of the program could be evaluated using the A4GE.

3.2.1.6 Operational procedures of the evaluation

In this section, I only briefly describe the operational procedures of the A4GE since these will be discussed in greater detail in Chapter 4. Like the 4GE, the A4GE data collection process consists of two interview rounds and a negotiation session. The data collection is centered on the CCIs of various stakeholders. During the two interview rounds, the stakeholders raise issues; these are then discussed in a negotiation session that is mediated by the evaluator in an attempt to achieve consensus and resolution.

There are few records in the literature of studies employing the 4GE in practice (Fishman, 1992; Lay & Papadopoulos, 2007). While those studies acknowledge the model's strengths in bringing about a fresh perspective about the CCIs related to the programs evaluated, they also point out several limitations to the negotiation stage prescribed in its methodology. Therefore, in what follows, I explore these limitations and present my adaptations to the 4GE operational procedures.

Huebner and Betts (1999) used the 4GE to assess a youth development project. They noted that a perfect consensus is at best an unrealistic objective of the model. The authors attribute the difficulty of reaching consensus to the differences that stakeholders bring to the negotiating table and the difficulty of finding people from all stakeholder groups who are articulate enough to represent their views. These findings resonate with the conclusions of Lay and Papadopoulos (2007) who used the 4GE to evaluate a government-funded initiative to fight child poverty and social exclusion in England. These authors argue that the negotiation requires highly skilled moderators to appease the tensions created by uneven sides during the negotiation. Likewise, in a reflection on three 4GE applications (in "Care of the elderly" wards in an acute care setting, a nursing program evaluation and an elderly care setting), Kosh (2000, p. 121) points out more difficulties related to the negotiation process of the 4GE. She notes that people were not so interested in "having a say" in the negotiation; they actually wanted their concerns to be heard but were reluctant to have a say because the disclosure of their identities made them vulnerable. Kosh (2000) argues that whereas the negotiation phase was initially conceived as a way for empowering certain groups, it seemed counter-productive especially where some groups were unwilling to be persuaded. In her analysis of the three cases, Kosh (2000) discusses the consequences of negotiating at an uneven table. She describes how differences in input, skills and knowledge, the domination of the process by certain groups, industrial relations and the unwillingness of some to listen to

others can all subvert the negotiation, thus dampening the strength and input of some stakeholders and reinforcing the constructions of the more “politically” powerful groups.

Additionally, in analyzing the dominance of some powerful stakeholders, Kosh (2000, p. 124) argues that, although the aim of the negotiation is to engender action planning and action taking, “action is not always the outcome of the negotiation process”. Laughlin and Broadbent (1996) had earlier attended to this weakness of the negotiation in 4GE while criticizing the underlying assumption of Guba and Lincoln, which states that “action connections are assumed to fall out naturally from the discursive process rather than being a further extension of the discourse”. They argue that this assumption is naïve as it confuses “consensus on understanding” with “consensus on praxis”. Enlightened understanding, the authors explain, does not automatically lead to action. Alternatively, the authors propose that action alternatives should be debated just as rigorously following similar discursive rules.

In sum, previous applications of the 4GE reported in the literature uncovered major limitations to the negotiation process. Nonetheless, I believe the negotiation concept by itself is not a limitation of the model as much as is the context within which it occurs. That is, the short time frame in face-to-face encounter, the mixing of theory and practice in the discourse, as well as inequities across the negotiation table, can all be framed as the context or conditions of the negotiation. Therefore, the adaptations I made to the 4GE with regard to its operational procedures were based on my attempt to overcome these contextual challenges.

In the A4GE, the first two rounds of the evaluation were the same as in the 4GE. Therefore, I identified the array of stakeholders relevant to the projected evaluation and elicited their CCIs using interviews. However, for the negotiation session, I employed a virtual medium instead of the face-to-face negotiation. I postulated that the use of this virtual forum might lessen the impact of power differentials caused by face-to-face interactions. The forum was also extended over longer periods of time, thus enabling stakeholders to better develop and articulate their arguments and make better and more informed decisions. Lastly, the forum was designed in such a way as to discriminate between the discourse on theory and the discourse on praxis. To this end, when coding and analyzing the data, and in addition to identifying the CCIs, I identified suggestions made by stakeholders about changing or improving some aspects of the program and later used those insights when designing the negotiation forum.

In addition to the changes introduced to the negotiation session, and to enhance stakeholder involvement while at the same time increasing the transparency of the processes embedded in the model, I argue that the A4GE should make use of digital online technologies to provide timely and comprehensive means for information sharing and dissemination. For this, I created a website that enables stakeholders to access the actual data from the evaluation as well as the interpretations that I made as the evaluator. The website can be accessed on this link: http://www.adaptedfourthgenerationevaluation.com/. Appendix A provides an extensive description of the website and its functions. The website was formatted to allow participants to interactively browse through the findings that emerged from the data analysis. This interactive way of overviewing the data was developed based on the premise that it would make the model more convenient to the participants who volunteered their time to take part in the evaluation process. Instead of having to read through hard copies of interim reports, stakeholders gain relatively faster and easier acquaintance with the evaluation data and findings. The website interface enables participants to explore the data by themes or by looking at data from particular stakeholder groups. Therefore, this adaptation to the 4GE potentially allows the stakeholders to explore the data in ways that are meaningful to them. Accordingly, it could potentially increase the transparency of the evaluation process and deepen the degree of participation of involved stakeholders.

3.2.1.7 Summary

A summary of the various theoretical propositions of the A4GE along the various PEMED dimensions is presented in Table 3.2.

Table 3.2 – Theoretical propositions of the A4GE along the PEMED dimensions

PEMED dimensions	A4GE theoretical propositions
Views about social programs	Program development is an evolving process that is predicated on stakeholders' experiences with the program
	Programs are never ideal; they are at best the most sophisticated interpretation of how to enact a particular policy within the constraints of the cultural, physical and psychological contexts
Paradigm of evaluation	The A4GE is interpretive and responsive
	The A4GE stakeholders are partners during the whole evaluative process
	The A4GE evaluator is an orchestrator of the evaluation processes
Views about utilization	Process use: stakeholders involved in the evaluation learn about the evaluand and the evaluation processes
	Instrumental use: The A4GE must have an action orientation
Purpose of the evaluation	The A4GE aims at developing the constructions of the stakeholders about the program
	The A4GE is best carried out during or after a program has been implemented
Scope of the evaluation	The A4GE does not predetermine the focus and scope of an evaluation
Operational procedures	Data collection in the A4GE is hermeneutic and dialectical
	Data analysis in the A4GE is based on constant-comparison and yields the claims, concerns, issues and program improvement suggestions of stakeholders
	Issues are negotiated in a virtual asynchronous negotiation forum and are either resolved or redefined
	The A4GE evaluator is responsible for using interactive technologies to disseminate evaluation information
	In the A4GE, discussions about theory and praxis are carried out separately

3.2.2 Using situational analysis to establish the appropriateness of the A4GE to the study context

Having described the A4GE in relation to the six PEMED dimensions, in this section, I rationalize my choice of this model for the conduct of this particular evaluation. To this end, I present a situational analysis of the context of the current evaluation guided by the three previously discussed factors: the context of the evaluand, the context of the evaluation and the context of the evaluator.

3.2.2.1 Context of the evaluand

Description of the evaluand

The evaluand in this study is the Graduate Diploma in Education for Science Teaching (hereafter referred to as GDE(ST)). The GDE(ST) is only one stream of the more general Graduate Diploma in Education (the GDE). However, in light of recurring calls to improve the quality of science teachers in Australia (Chubb, 2014; Goodrum et al., 2012), and considering the difficulty of evaluating all the other streams, I decided to focus solely on the science teaching stream, given my background experiences with science teaching and science teacher education.

The GDE(ST) offers graduate science student teachers the credentials needed for teaching accreditation in Australia. Although not explicitly stated on the university's website, the aims of the GDE(ST) can be inferred from the learning outcomes listed under the descriptions of the program's constituent units (see this link https://my.une.edu.au/courses/2011/courses/GDED/program-of-study.html for a list of the constituent units). These aims include, but are not limited to, developing the candidates' appreciation and understanding of: (1) different approaches to teaching and learning and classroom management, (2) the relevance and application of key scientific concepts in the schools' curricula and the skills involved in teaching them, (3) safety issues and requirements associated with science teaching, (4) nature and forms of assessment, (5) innovative technologies and their application to the teaching and learning context, and (6) aboriginal, NESB and special education. Additionally, candidates gain awareness of the accreditation processes and the range of professional development opportunities available to them in addition to legislative requirements of science teachers and the implications these have on teaching.

To be eligible for admission into the GDE(ST) program, a candidate must hold a relevant university science degree. The approved science majors are biology, chemistry, earth and environmental science, information technology, mathematics, or physics. Alternatively, the candidate must hold a three-year diploma acceptable to the University's School of Education. A candidate for the GDE(ST) can be Australian or International and there is no age level constraint for admission into the program.

The structure of the GDE(ST) (as it was when I began my data collection in 2011) aligned with the New South Wales Institute of Teachers (NSWIT) guidelines. Upon enrolment, candidates are expected to complete a program of study consisting of eight compulsory

teaching units, 50 days of professional experience and four science curriculum units. As such, the period of candidature is one-and-a-half years for fulltime candidates and up to six years for part-time candidates.

At the time of the evaluation, the compulsory teaching units for prospective secondary science teachers of the GDE(ST) were: Aboriginal Education, Curriculum and the Social Context of School, Literacies in Context, ICT in Education, Planning for Effective Learning, Inclusive and Special Education, Classroom , Management and Teaching for Cultural Diversity-NESB students. While the GDE(ST) is designed for secondary science teachers, many of these units are common to primary science teacher education programs. As such, these units are designed to accommodate both secondary and primary science teachers. The four science education units are only taken by the secondary science teachers and consist of: Science Education 7-10: Foundation for Teaching, Science Education 7-10: Teaching and Learning, Science Education 11-12: Advanced Pedagogy, and Science Education 11-12: Plan, Assess and Report. A description of these units, taken from the official university website for the academic year 2011, is found in Appendix B. In addition to these units, candidates are expected to take the professional experience sequence of units (hereafter referred to as PREX) concurrently with other specified units. Importantly, in 2011, the PREX units were not allocated any credit points, as was the case with other units. Upon successful completion of all of the requirements above, the student teacher would have taken a total of 72 credit points which qualifies them for the GDE(ST). Candidates who successfully complete the GDE(ST) are able to get registration with the NSWIT, which, at the time, was the accrediting body for teachers and are, therefore, eligible to work in any of the country's independent or government schools.

The GDE(ST) is offered in on-campus and off-campus modes. The off-campus tradition at the host university was, in the past, conducted by correspondence, and is now offered online. Generally, off-campus student teachers undertake their PREX units in approved schools of their choice while on-campus student teachers are placed by the professional experience office into one of the nearby schools.

The state of the evaluand

Since the current evaluation was framed as a protocol for a university review of the program where an interpretive framework adopted for reviewing academic programs, I describe here the changes that affected the evaluand over a five-year period. This is reminiscent of the time

frame adopted by the host university's school of education for conducting a program evaluation.

From 2006 to 2011, several changes to the structure and content of the GDE(ST) occurred as required by NSWIT, the accreditation authority. First, the period of candidature changed once: from 2006 to 2008, the GDE(ST) was one year for fulltime candidates. In 2009, it became one-and-a-half years. Second, the number of days in the professional experience changed three times: in 2006 and 2007, it was 40 days (2x20 days practicums); in 2008, it was increased to 60 days (3x20 days practicums); in 2009 and 2010, it was changed to become 45 days long (1x5 days practicum and 2x20 days practicums); then in 2011 the number of days was set to 50 (1x10 days practicum and 2x20 days practicums). The number of units that were required in the GDE(ST) changed once: from 2006 to 2008, candidates were required to complete 48 credit points, which consisted of four core units and four science education units. The professional experience units were not allocated any credit points. After 2009, the duration of the GDE(ST) increased and, with it, four additional core units that needed to be taken by student teacher candidates.

The repeated changes to the structure of the GDE(ST) indicates the impact of external forces at play. Hence, the make-up of the program is seldom ideal but rather a compromise to comply with these pressures. The A4GE's propositions about social programs (Dimension 1 of the PEMED) suggest that programs cannot be evaluated in isolation of their cultural, physical and psychological contexts. Additionally, programs can never reach an ideal state and they are, at best, the most sophisticated interpretation of how to enact a particular policy. As such, evaluators using the A4GE are keen to consider the effect of these compromises, which are often driven by a university's need to meet the requirements set by external accreditation and quality monitoring agencies (such as AITSL and AUQA), Academic Boards and competition with other universities for students. It follows from this argument that the A4GE's views about social programs are compatible with the context of this study.

With respect to the decisions awaiting the GDE(ST), in 2011 the university decided to phase out the GDE(ST) and replace it with a two year Masters of Teaching program. This step was taken in response to the AITSL requirement to make the teacher preparation courses two years fulltime. Therefore, 2013 was the last year the GDE(ST) was offered.

Since the GDE(ST) was going to be "replaced" with another program, it was opportune for lecturers involved in the development of the new program to reap the lessons learned from

the GDE(ST). After all, the new program was intended to serve the same goals as the GDE(ST) and, therefore, its development can benefit from experiences with the GDE(ST). The A4GE's underlying paradigm of evaluation (Dimension 2 of the PEMED), which encourages the interpretive hermeneutic approach, presents itself, thus, as a viable option to achieve this particular outcome, because it adopts a "discovery" and not a "verificative" position (Guba & Lincoln, 1989). Particularly, since the A4GE's purpose (Dimension 4 of the PEMED) is to solicit the CCIs of the various stakeholder groups, it is useful in pointing out the assets and liabilities of the GDE(ST). Furthermore, stakeholders' suggestions on how to improve the GDE(ST), and which are also elicited in the A4GE, will help the developers of the new program gain some evidence and experience-based insights to enhance the quality of the Master of Teaching (Secondary science).

Organizational setting of the evaluand

The university where the evaluation took place is divided into two large faculties (Faculty of Arts and Sciences and Faculty of the Professions) that jointly host 10 schools. A Head of School manages each school. The Head of School in the School of Education has two deputies: the Deputy Head of School Teaching and Learning, and the Deputy Head of School Research. As the name implies, the Deputy Head of School Teaching and Learning is concerned with the teaching and learning activities in the school, which includes program monitoring and evaluation.

Lecturers teaching into the GDE(ST) do not need to hold a PhD to be eligible to coordinate a teaching unit and, while prior experience in teaching is desirable, it is not mandatory either. The units are often taught by more than one lecturer but are coordinated by one lecturer. Similarly, casual markers, who are managed by the unit coordinator, can do the marking of the assignments. Lecturers in the GDE(ST) work within discipline-based teams. So, for example, there is the science education team, the teaching and learning team, and so forth. While these teams meet regularly, they do not engage in discussions or decision-making about programs such as the GDE(ST). These are the responsibility of the Deputy Head of School Teaching and Learning and the course coordinators.

Since many stakeholders are involved in administering and monitoring the GDE(ST), the use of the A4GE is warranted by its participatory nature (Dimension 2 of the PEMED). Indeed, the A4GE brings together various stakeholders in a negotiation about the problems and assets of the GDE(ST). Through engaging in dialogue, these stakeholders expose their differing beliefs and values and exchange information and evidence about the program. Furthermore,

the hierarchy in the school lacks an established structure or process for effectively engaging lecturers with other stakeholders (such as school teachers) in dialogue about the quality of programs. Therefore, the use of the A4GE can be justified in relation to its processes (the hermeneutic cycles) which offer channels for lecturers and other stakeholders to exchange perspectives and debate opinions concerning the GDE(ST).

3.2.2.2 Context of the evaluation

Evaluation background

The current evaluation was not commissioned by the university, or anyone else for that matter. For this reason, not many lecturers and teachers were willing to commit their time and effort to it because it was simply not part of their workload. On a positive note, the fact that the evaluation was not commissioned gave me complete freedom as to which evaluation model to use. Furthermore, I was not bound by political agendas as to the scope and purposes of the evaluation. Therefore, the A4GE was appropriate since its formulation does not impose a narrow scope on the evaluation (Dimension 5 of the PEMED) but rather promotes the use of negotiations to determine its goals and uses stakeholders' input to determine its scope.

As outlined in chapter one, program evaluation is a regular process in the host university. The quality of academic programs is monitored by the Academic Board. To this end, the Academic Board is assisted by two committees: the Teaching and Learning Committee and the Academic Programs Committee. The Teaching and Learning Committee ensures compliance of programs with the Tertiary Education Quality Standards Agency (TEQSA) and the Australian Qualifications Framework (follow this link to see a complete description of the roles of this committee: http://www.une.edu.au/__data/assets/pdf_file/0006/25368/abhandbook2013.pdf). Similarly, the Academic Program's Committee provides advice to the Academic Board on the academic merit of new and amended programs as well as on the withdrawal of programs, and assists faculties and schools in the development of programs (follow this link to see a complete description of the roles of this committee: http://www.une.edu.au/about-une/leadership/governance/academic-board/handbook/academic-board-handbook/academic-board-committees/academic-board-committees/academic-program-committee/?a=52525).

The quality of academic programs is also monitored indirectly by the Corporate Intelligence Unit (CIU) that administers regular and compulsory student evaluations of units that constitute these programs. Nevertheless, the CIU does not analyze programs per se, only

individual units. The items from the instrument used in conducting student evaluations of the units are presented in Table 3.3.

Table 3.3 – Items used for conducting student evaluations of units*

Type of questions	Actual items used in the student evaluations of units
Quantitative questions using a five point Likert scale	– The learning outcomes of this unit were made clear to me – The unit enabled me to achieve the learning outcomes – The unit was intellectually stimulating – I found the resources provided for the unit (e.g. online, print) to be helpful – I received constructive feedback on my work – The feedback I received was provided in time to help me improve – The overall amount of work required of me for this unit was appropriate – Overall, I was satisfied with the quality of this unit
Qualitative questions	– What were the best aspects of this unit? – What aspects of the unit are most in need of improvement?

*Retrieved from: http://planning.une.edu.au/Evaluations/Process.htm

While student evaluation of units is highly regarded in the university (for monitoring and staff promotion), as can be seen in Table 3.3, their content can only give a rough estimate about the quality of the programs hosting those units.

In addition to the Academic Board Committees and the CIU, the quality of academic programs is monitored and reviewed by the school at the host university. The GDE(ST) coordinator, under the direction of the Deputy Head of School Teaching and Learning, oversees program monitoring, which involves analysis of the following key performance indicators: enrolment number of commencing students, progress rate of students, retention rate of students, student feedback, student outcomes and the program's financial income (University of New England, 2010). With regard to program reviews, the process is only conducted periodically (once every three or five years) and is usually replaced by another process when accrediting bodies demand program reviews from the university. When the GDE(ST) is under review, the Deputy Head of School Teaching and Learning chairs a committee to perform the review, which involves evaluating the program's structure, the alignment of learning objectives to learning outcomes and assessment, the currency of the curriculum, the quality of teaching and learning including assessment, student perceptions

and feedback, and feedback from the relevant professional, accrediting and employer groups (University of New England, 2009).

While it is apparent that different evaluation activities occur at the university level, these processes are characterized as being mostly based on measuring performance indicators and quality of performance. Against this backdrop, the use of the A4GE is warranted in that it presents itself in this context as a viable alternative and complementary option for conducting regular faculty reviews. With its focus on uncovering the CCIs of various stakeholders (Dimension 3 of the PEMED), the A4GE presents a meaningful and collaborative approach, focused on the needs and challenges faced by those involved in the program. Instead of reporting on key performance indicators, the A4GE unveils the assets and liabilities of a program through an analysis of the CCIs of various stakeholder groups who have direct experience with the program. As such, it is more focused on promoting program improvement rather than providing accountability evidence.

Evaluation resources

The human, financial and time resources available to the conduct of this evaluation were limited. Being the only evaluator meant that I had to conduct the evaluation activities (collection and analysis of data) unassisted by university infrastructure. Nevertheless, because the A4GE advocates depth rather than breadth of stakeholder participation (Dimension 2 of the PEMED), I could execute the data collection and analysis processes of the A4GE with relative comfort since I was managing a relatively small number of data sources. In this respect, the use of NVivo 10 alleviated much of the burden often associated with interpretive data analysis (such as coding).

Evaluation leverage

Since the current evaluation was not commissioned and was framed as part of a PhD study, its potential to stimulate action at the school level was relatively low. Nevertheless, the use of the A4GE is rationalized because the model offers at least two ways through which it can benefit its users: through instrumental use and through process use (Dimension 3 of the PEMED). University lecturers, program coordinators, the Deputy Head of School Teaching and Learning and the Head of School are the primary intended users of this evaluation. Through their involvement in the evaluation, they can benefit from knowing the issues that they, their colleagues and their student teachers are raising about the GDE(ST). Through exposure to what others say about the GDE(ST), they have opportunities to reflect on their own arguments as well as to analyze the arguments put forth by others. The result of this

process is more informed and sophisticated pedagogical perspectives as well as suggestions for improving the program. Similarly, the Academic Board, which is charged with overseeing the GDE(ST), can gain insights from this evaluation about the quality and processes of an innovative participatory approach to evaluation.

3.2.2.3 Context of the evaluator

Because the evaluation was carried out as part of my PhD study and was not commissioned by anyone, I had considerable freedom to think about the appropriateness of the evaluation model to the context. Being a student at the host university meant that I was familiar with the GDE(ST) and its requirements. Additionally, two of my supervisors were teaching units into that program and were readily available to answer my concerns and queries about the program. Furthermore, having substantial experience in science teaching and science teaching degrees, I was familiar with the challenges that teachers face in schools and was able to relate to them well. Also, my previous research experience in my Masters' degree exposed to me the practices of program evaluation and the contingencies that impinge on its practice. All of these factors provided me with the needed confidence and requisite knowledge and skills to administer the A4GE, which relies heavily on the involvement of the evaluator with the various stakeholders to collaboratively develop and debate constructions about the evaluand (Dimension 2 of the PEMED).

3.3 Conclusion

In this chapter, I have argued for the appropriateness of the A4GE for the study context by using the PEMED and my situational analysis of the context. Throughout this process, I defined the A4GE and presented its theoretical propositions along the PEMED dimensions. In the next chapter, I outline the methodology used to carry out the evaluation of the GDE(ST) using the A4GE as well as the empirical investigation of the A4GE using the conceptual framework discussed in Chapter 2.

CHAPTER 4: METHODOLOGY

In this chapter, I explain the methodological design and procedures that I used to carry out the two research components of this study: the evaluation of the GDE(ST) using the A4GE and the empirical investigation of the application of the A4GE. After outlining the research component and their objectives, I discuss the theoretical assumptions underpinning the approach used for this research and present my rationale for choosing the interpretive case study design. Based on that, I present an overview of the research design and discuss the processes of participant selection, data collection and data analysis. Additionally, I discuss the measures taken to ensure the rigor of the study. I close the chapter with a discussion of important ethical considerations. The description of the processes of data collection and analyses of the A4GE featured in this chapter is elaborate, as it constitutes a significant part of the data used in Chapter 6 to examine the congruence between the theory and implementation of the A4GE.

4.1 Research components

The current investigation comprises two components:

1. Evaluation of the GDE(ST) using the A4GE, and
2. Empirical investigation of the application of the A4GE.

The aims of the first component are to construct the CCIs of each stakeholder group, explain how the CCIs of each group were similar and how they are different, negotiate and potentially resolve the identified issues and, when possible, develop suggestions for improving the GDE(ST).

For the second component, I make use of the conceptual framework built around Miller's (2010) criteria to answer the general research question posited in Chapter 1:

> How congruent are the underlying theory and practice of an adapted version of the fourth generation evaluation model in the context of evaluating a secondary science teacher preparation program in an Australian university?

The data collection process (see Table 4.1) for the two components occurred simultaneously through a succession of two interview rounds followed by the administration of a virtual negotiation session. In addition to intensive interviews and the forum, further information was obtained from the analysis of my field notes and research journal.

Table 4.1 – Research components, data sources and approaches to data collection

Rounds of data collection	Research components	Data sources	Approach to data collection
Interviews Round 1	(1) Evaluation of the GDE(ST)	Interviews	A4GE
	(2) Investigation of the application of the A4GE	Interviews Field Notes Journal logs	Interpretive Case Study
Interviews Round 2	(1) Evaluation of the GDE(ST)	Interviews	A4GE
	(2) Investigation of the application of the A4GE	Interviews Field Notes Journal logs	Interpretive Case Study
Negotiation Session	(1) Evaluation of the GDE(ST)	The virtual forum	A4GE
	(2) Investigation of the application of the A4GE	Online forum Field Notes Journal logs	Interpretive Case Study

For the first component of the investigation, the evaluation of the GDE(ST), the methodology of the A4GE provided guidance for the data collection, analysis and interpretation. This first component is not the main focus of this study but is essential for understanding the results of the empirical investigation of the application of the A4GE, which is the second and more central component. Nevertheless, since the data collection for both components was carried out simultaneously, I outline the methodology used for the two components in this chapter.

4.2 Approach to research: The interpretive case study

The choice of an approach to research depends on a variety of considerations, including the researcher's assumptions and worldviews, the nature of the problem under consideration, as well as the researcher's personal experiences (Creswell, 2007). While the approach for carrying out the first component of the study, that is the evaluation of the GDE(ST), followed the premises of the A4GE and its underlying methodology, the research design selected for the second component, that is the empirical investigation of the application of the A4GE, was an interpretive case study. The choice of the A4GE as an approach to doing the evaluation has already been rationalized in Chapter 3. In this section, I outline the rationale underlying

my choice of the interpretive case study as an approach for the investigation of the application of the A4GE.

4.2.1 The interpretive research paradigm: Foundations, rationale and implications

Several paradigms exist in educational research, of which the most prominent are post-positivist, interpretive and critical theory (Cohen, Manion, & Morrison, 2007; Lincoln, Lynham, & Guba, 2011; Rossman & Rallis, 2012). The selection of any of those paradigms represents a commitment to a certain way of making and understanding knowledge claims. According to Creswell (2007, p. 15), the choice of a particular paradigm reflects the researcher's ontological, epistemological, axiological, rhetorical and methodological assumptions. There is extensive literature about paradigms in educational research (see for example Cohen et al., 2007; Lincoln et al., 2011) and, therefore, these paradigms will not be reviewed here.

In this study, I embraced the principles and methods advocated in the interpretive paradigm because I sought in-depth analysis of the fit between the theory and practice of the A4GE in the specific context of implementing it to evaluate the GDE(ST). I summarize the assumptions underlying the interpretive paradigm against each of Creswell's five attributes in Table 4.2.

Table 4.2 – Assumptions entertained by the interpretive research paradigm

Attribute	Definition	Assumptions of the Interpretive Paradigm
Ontology	Stance towards the nature of reality	Relativist: Realities are subjective and multiple; realities are constructed by participants, they do not exist outside a participant's interpretation
Epistemology	Stance towards the nature of knowledge and how it can be known.	Transactional/subjective: strong relationship between the researcher and what is being researched
Axiology	Stance towards the role of values	Researcher's and participants' values are part of the data; researcher acknowledges biases and their role in shaping interpretations
Rhetoric	Stance towards the language used in the research	Literary/informal style: researcher writes in a personal voice and uses qualitative terms while avoiding matching interpretations to "standard" scholarly definitions
Methodology	Stance towards the process of research	Researcher follows an inductive logic paying particular attention to context and follows an emerging design

As opposed to its positivist counterpart, wherein researchers seek to establish correlations and test hypotheses using the methods of inquiry of the natural sciences (Lincoln & Guba, 1985), the interpretive paradigm is concerned with the meanings and interpretations made by the individuals, thus allowing the researcher to develop subjective descriptions of phenomena (Lincoln et al., 2011). This is commensurate with the purposes of this study in which I investigated the theoretical assumptions of the A4GE in light of the experiences of stakeholders involved in its application. Although the critical theory paradigm also allows for similar analysis, its focus is on social change and challenging the social realities imposed by existing social structures (Rossman & Rallis, 2012). This is not aligned with the purposes of this study. Hence, this research was conducted using the interpretive paradigm.

Acknowledging that different paradigmatic bases produce different standards for carrying out and judging investigations, the selection of the interpretive paradigm presented several noteworthy implications for the conduct of the study. First and foremost, the interpretive inquiry is particularly attentive to the influential role of the context of the study, which takes place in a natural uncontrolled setting (Marshall & Rossman, 2011). Furthermore, when using interpretive traditions, researchers need to deploy multiple methods for data collection that are both "interactive and humanistic" (Marshall & Rossman, 2011, p. 3). Indeed, the primary concern in this kind of research is to derive meanings from the lived experiences of the participants (Miles & Huberman, 1994). Additionally, interpretive researchers must also acknowledge the personal biases that impinge on the participants' interpretations and systematically engage in iterative reflections about how their personal biographies and cultural backgrounds shape the investigation (Creswell, 2007; Marshall & Rossman, 2011). Lastly, accepting the assumptions of the interpretive paradigm also means that researchers need to adhere to standards of research rigor appropriate to interpretive paradigm (Lincoln & Guba, 2007).

Taking into account these implications, I devoted particular attention, in this study, to the context where the A4GE was examined and investigated. Indeed the context presented a wealth of information that I drew on to make sense of the various insights derived from the implementation of the A4GE. For example, knowing about previous evaluations and how they were sometimes perceived as unfair by some lecturers, made me appreciate why some of the informants were reluctant to speak freely about the program, particularly about their teaching units. Also, understanding the organizational context allowed me to gain further insights about why some participants encouraged or discouraged the use of the A4GE as a

tool for evaluation. Moreover, to provide rich data, I used interviews, a negotiation forum, observation notes and my personal reflective journal to collect data for the study. Furthermore, to keep track of my own biases and interpretations, I kept a journal log where I recorded my personal responses throughout the data collection and analyses. I also included my personal speculations and questions, and dated them. Finally, to ensure rigor in carrying out the interpretive research, I took several steps to enhance the trustworthiness of the study and to address authenticity criteria.

4.2.2 The interpretive case study: Foundations, rationale and implications

Several researchers have discussed alternative approaches to carrying out interpretive research. Merriam (1998) defines five qualitative streams when carrying out interpretive research: basic/generic qualitative research, ethnography, phenomenology, grounded theory and case studies. Likewise, Creswell (2007) presents an elaborate analysis of how researchers should choose among five dominant approaches to qualitative research: narrative research, phenomenology, grounded theory, ethnography and case study. According to Creswell (2007, pp. 78-79), the selection of any one of these approaches should be rationalized in relation to the focus of the study, the unit of analysis, the data collection forms, the data analysis strategies, and the outcome or written report. The use of narrative research is preferred when the focus of the study is exploring the life of a particular individual. Also, phenomenology is best used for understanding the experiences of individuals; ethnography is most compatible with a study seeking to describe and interpret a cultural group, while grounded theory is aimed at developing a theory grounded in data.

A case study is, therefore, deemed appropriate when: the focus of a study is on developing an in-depth understanding and analysis of a case/phenomenon; the unit of analysis is an event, activity or a program; data are collected through multiple sources, such as interviews, observations and documents; the analysis of data is carried out through the description of the themes that emerge from the case; and, finally, when the outcome of the investigation is a detailed analysis of that particular case. Additionally, Merriam (1998) explains that case studies are best used when the researcher is investigating a phenomenon in its real-life context, particularly when the boundaries between the phenomenon studied and its context are blurred. Furthermore, she adds that the case study approach brings about an in-depth understanding of a phenomenon about which little is known. Based on these reasons, I have adopted a case study approach (Merriam, 1998; Stake, 1995; Yin, 2009) to empirically

investigate the application of the A4GE and to investigate the connection between its theory and practice.

Adopting the interpretive case study approach has two important implications for carrying out the research protocols and for understanding the findings derived from its conduct. First, the case study researcher should focus on providing a thick description of the context in which the phenomenon is studied (Merriam, 1998). Indeed, the term, “thick description”, is often associated with the case study approach and is commonly used by authors of case study writings (Creswell, 2007; Lincoln et al., 2011; Marshall & Rossman, 2011; Patton, 2002; Stake, 1995). Denzin (1989, p. 83) suggests that thick descriptions extend beyond facts and superficial appearances to present in detail the “context, emotion, and the webs of social relationships that join persons to one another”. Denzin (1989, p. 83) adds that thick descriptions should also reflect the “voices, feelings, and meanings of interacting individuals”. Second, the use of the interpretive case study means that the focus of the research is limited to theoretical generalization and not statistical generalization, as is the case with experimental and other post-positivistic designs. Stake (1978) uses the attributes context-specific and naturalistic to define the nature of the generalization derived from case studies. In that respect, the generalization of case study findings is bounded to the case itself and, at best, extended to a type of cases.

In this study, thick descriptions were crucial for understanding whether and how the underlying theoretical propositions of the A4GE and its practice were connected. These descriptions clarified how the theory of the model was enacted in a particular context, what aspects enhanced its use and what others presented barriers to its application. Additionally, acknowledging that the findings of this study could not be statistically generalized and applied to other contexts, I have presented a detailed contextual analysis to show how the study findings are nested in the study context and how particular findings could be meaningful for their own context.

4.3 Research design

I developed the study design, shown in Figure 4.1, to achieve the purposes of this study. I undertook a preparation stage by developing the PEMED from the literature and then defining the 4GE and adapting it. Subsequently, I explained and outlined the theoretical propositions of the A4GE using the PEMED. Additionally, and using situational analysis of

the context of implementation of the A4GE, I checked and justified the appropriateness of the A4GE for the study context. Finally, I used the PEMED to plan the data collection.

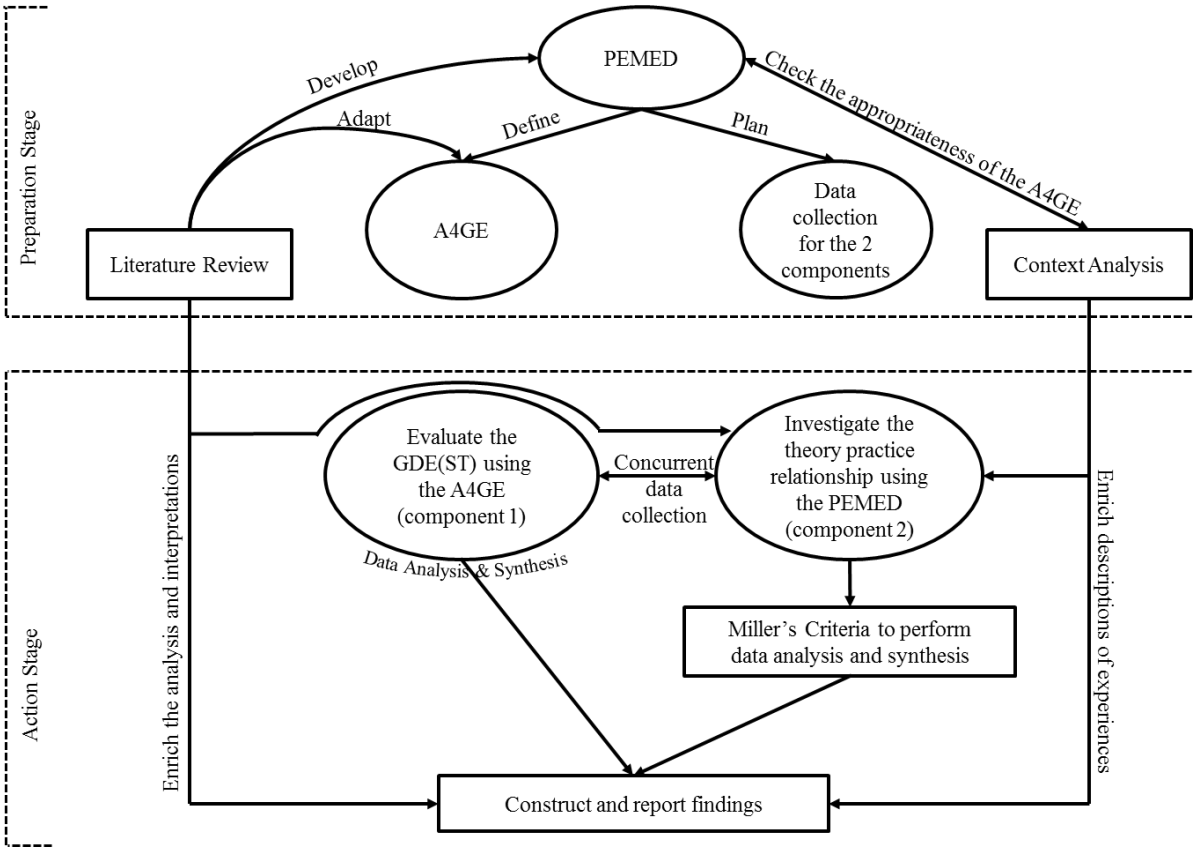


Figure 4.1 – The research design

Once the preparation stage was complete, I proceeded to the action stage of data collection and analysis. As shown in Figure 4.1, I collected data about the practice of the A4GE (component two of the research) at the same time that I was carrying out the evaluation of the GDE(ST) (component one of the research).

Throughout the data collection stage, I used constant comparison (Charmaz, 2011) to construct the findings from the evaluation of the GDE(ST). Once the evaluation of the GDE(ST) and the data collection for the investigation of the theory-practice relationship were complete (component 2), I used constant comparison and negative case analysis (Robinson, 1951) to analyze the data for the second component.

Throughout the action stage, I consulted the literature to gain further guidance into the data collection, and to enrich the analyses and interpretations. Similarly, I kept notes of the contextual factors that influenced stakeholders' interpretations wherever appropriate so as to

situate these interpretations contextually. I also used member checking throughout the data collection to enhance the trustworthiness of the study.

Once the data collection stage was complete, I developed the evaluation report for the first component of the research and then constructed the findings for the empirical investigation of the A4GE (component two of the research) using Miller's (2010) criteria.

4.4 Selection of participants

The sample size in this study was consistent with what would be expected of a program evaluation carried out as part of a periodic faculty or school review at a university. Accordingly, only one STPP was considered for the evaluation.

Following Guba and Lincoln (1989), the participants for the A4GE consisted of three types of stakeholder groups. The "agents" formed the first type of groups and included those involved in the production, use and implementation of the evaluand, which, in this case, comprised university lecturers and the Head of School. The "beneficiaries" formed the second type of stakeholder groups and included those who profited in some way from the evaluand. The 'disadvantaged' formed the third type of groups; these were those who did not profit from the evaluand as much as they could potentially have. Pre-service science student teacher graduates and experienced science teachers were recruited as belonging to either one of the latter two groups, which depended upon their experiences with preparation and professional practice. Details of the stakeholder samples and the number of interviews are summarized in Table 4.3.

Two stakeholder groups were thus involved in this study: university lecturers (including the Head of School) and science teachers. The groups not only participated in the evaluation but also provided their point of view about the usefulness and relevance of the evaluation model. University lecturers were selected on the basis of their experience in and knowledge about the different key areas of science teacher preparation (e.g., Science Teaching, Classroom Behavior Management, Curriculum Studies, ICT, Aboriginal Education). The teachers included recently graduated secondary science teachers and experienced secondary science teachers who were all graduates from the selected university and program. They were selected to ensure representation from private and public schools, and from regional, rural and urban settings.

Table 4.3 – Data sources, number of sources and number of interviews.

Data Source	Number of sources	Number of interviews		Cumulative number of contributions
		Round 1	Round 2	
Lecturers (agents)	8 (out of 19 invited)	8	5	13
Head of school of education (agent)	1	0	1	24
Science teachers (disadvantaged and/or beneficiaries)	6 (out of 30 invited)	6	4	23
Negotiation forum	6	N/A	N/A	N/A
Other	Researcher's reflective journal Field notes Previous evaluation documents/records	N/A		N/A

Thirty invitations were sent to teachers who were recent graduates of the program. Out of the teachers who responded to these invitations, six signed and returned the consent form and indicated their willingness to take part in the study. Similarly, eight (out of 19 invited) lecturers signed the consent form and took part in the study. Importantly, the small response rates portrayed here are reminiscent of those often reported in faculty reviews of programs carried out across universities and do not undermine the merit of the A4GE as a model for evaluation.

I considered including other stakeholders in the study such as representative members from the Australian Science Teachers Association, the Board of Studies, and the Australian Institute for Teaching and School Leadership. However, I decided to exclude them from this study because they could not be expected to be sufficiently familiar with the specific program being studied and their input would be only tangentially relevant to the peculiarities of the course. Furthermore, their positions on programs in general are already on the public record.

4.5 Data collection rounds

Figure 4.2 shows the data collection rounds used in the two components of the study. As shown, data collection consisted of two interview rounds and culminated in a negotiation session. The arrows in the figure indicate the progression of interviews, which started with the first stakeholder group (university lecturers) then the second stakeholder group (science

teachers) during the first interview round. The arrows then spiral inwards, signaling the start of the second round of interviews, which was more focused around the CCIs identified during the first round. As in the first round, the second round of interviews was first with the lecturers group then the science teachers. The negotiation session followed these two interview rounds. The inwards spiraling of the arrows in Figure 4.2 reflects the focusing process inherent to the A4GE methodology, which proceeds from discussing as many CCIs as are available, to negotiating a selected few.

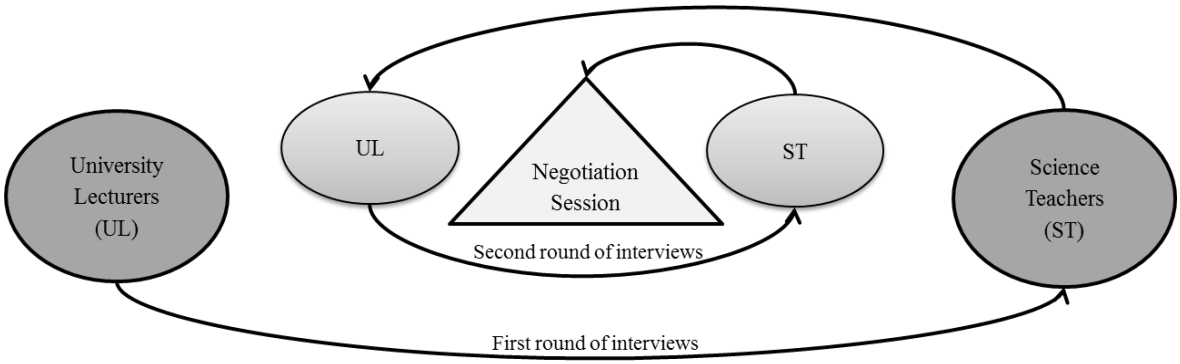


Figure 4.2 – The data collection rounds

4.5.1 Round 1 of interviews

The first round of interviews was entirely based on eliciting the stakeholders' initial CCIs about the GDE(ST). In this round, open-ended interviews were used to conduct the hermeneutic cycle proposed by Guba and Lincoln (1989). The hermeneutic cycle (see Figure 4.3) refers to the probing of individual and collective CCIs of the different stakeholders.

The number "1" in the circles in Figure 4.3 refers to the first stakeholder group and the letter next to the number refers to the individuals within the group. Hence, the circle entitled "1a" refers to the first interviewee within the first group and the circle entitled "1z" refers to the last interviewee within that same group. Each participant in each stakeholder group was asked about his/her experiences with the evaluand and, in this way, the participant's interpretations – or constructions (using Guba and Lincoln's terminology) – about the program were developed and noted. Additionally, participants were asked about their conceptions of the evaluative approach and whether they thought it was valuable and worthy of their participation and in what ways. Further probing questions were used to unveil their

conceptions as to whether they believed the A4GE being undertaken was fair and effective in providing judgment about the quality of the GDE(ST).

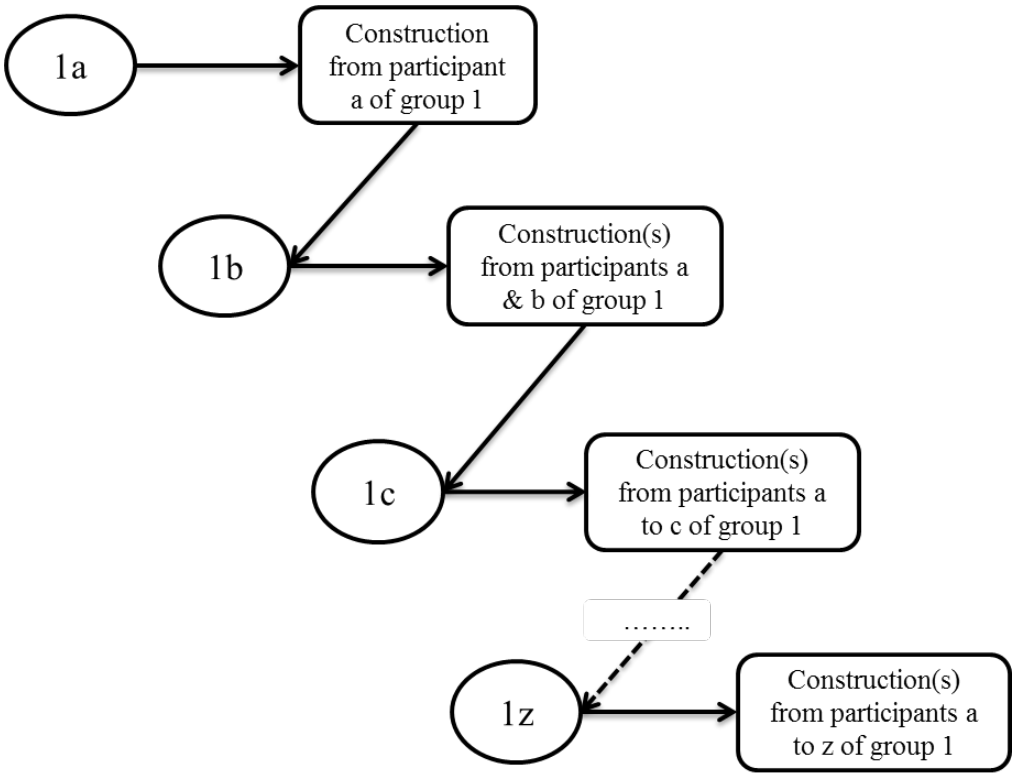


Figure 4.3 – The hermeneutic cycle for each stakeholder group

After the first interview with the first participant, 1a, the interview was transcribed, member-checked and analyzed, and the Claims and Concerns expressed by 1a were noted. During the second interview, in addition to asking the same questions asked of 1a, the second participant, 1b, was exposed to the Claims and Concerns made by 1a and asked to comment on those. Therefore, the resulting constructions were more sophisticated, building on the ideas of 1a and 1b. Continuing in this fashion, I derived the constructions from all participants in group 1.

The constructions that resulted from the hermeneutic cycle with the first group of stakeholders (i.e., lecturers) were then exposed to the second group of stakeholders (teachers) in a similar fashion (see Figure 4.4). Therefore, at the end of the first round of interviews with the two groups, I had an extensive list of all the CCIs as part of the constructions I developed from all the participants from both stakeholder groups.

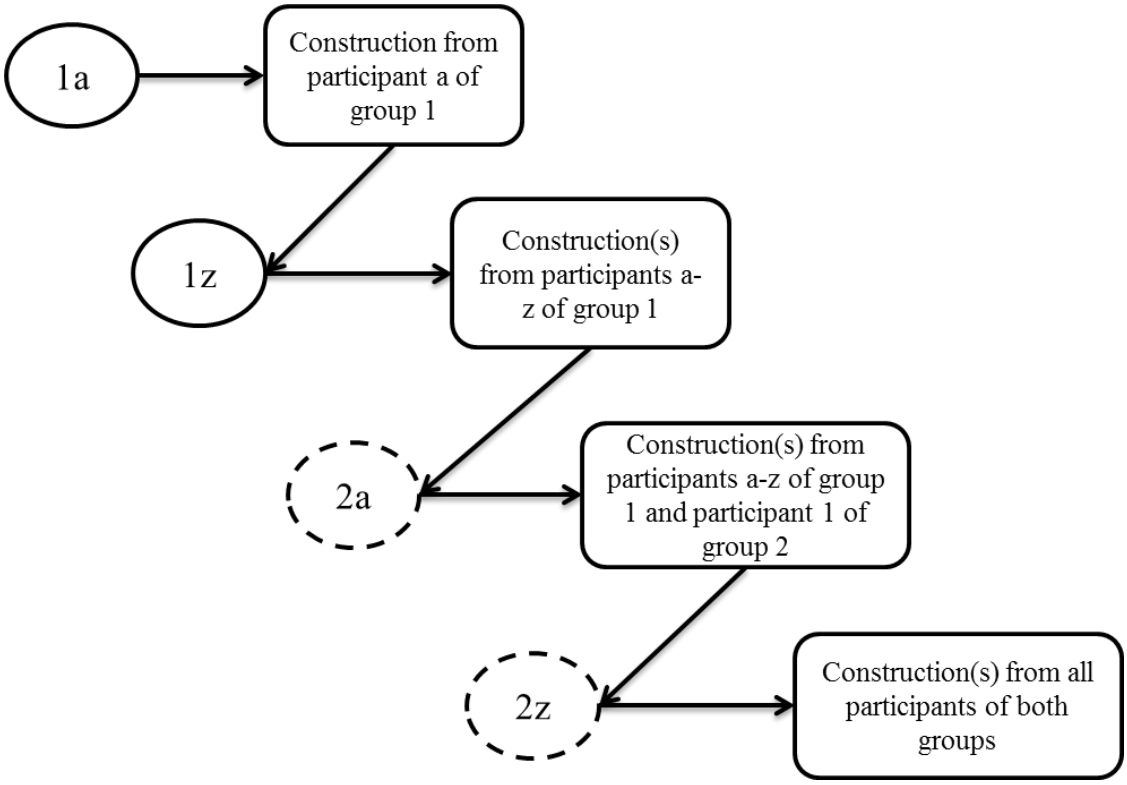


Figure 4.4 – The hermeneutic cycle for all stakeholder groups

Once the first round of interviews was completed, I compared and contrasted the various Claims and Concerns and developed a list of Issues for each stakeholder group. Based on the analysis of existing issues, I searched for additional information from other sources (documents or the literature) to help participants back up their claims and provide evidence when and where necessary. Furthermore, I generated a report that contained the CCIs of each group of stakeholders. The content of the report was then transferred onto a password-protected website that could be accessed by the participants and that presented the data and summaries of the data in an interactive online format. The website can be accessed via the following link: http://www.adaptedfourthgenerationevaluation.com/. The design of the website was piloted prior to its publication. A trial case was set up using "dummy" data, and three lecturers and two teachers (not involved in the study) were invited to interact with the site and provide feedback. The feedback was incorporated into a second version of the website. Snapshots of the site are presented in Appendix A.

Through the interactive portal of the website, participants could explore the data by selecting a theme or a topic which interested them. They could then choose which stakeholder group – science teachers alone, the university lecturers alone, or all the stakeholders – they wanted to browse. Upon selecting a theme and a stakeholder group, participants would then be given

the choice of browsing through the Claims, the Concerns, the Issues or the Suggestions. Additionally, having decided on a theme, a stakeholder group and the nature of the information, participants could either skim through the raw data or browse summaries of the data that I had already developed from the data. This interactive approach to browsing through the findings enabled participants to make comparisons between the CCIs of the two stakeholder groups.

4.5.2 Round 2 of interviews

In the second round of data collection, participants were asked to elaborate their CCIs further and to comment on other participants' CCIs. Therefore, their discussions were more focused than was the case in the first round. Similar to the first round, though, the second round was designed around the principles of the hermeneutic circles shown in Figure 4.4. The purpose of the second round of interviews was to achieve a higher sophistication of responses from participants by exposing them (through the website) to each other's interpretations of the program processes and outcomes. Participants were also asked to identify five to ten major issues that were brought up in the first round of interviews that they wished to negotiate in the last stage of data collection. Additionally, participants were asked to comment about the processes and outcomes of the evaluation model. Furthermore, participants provided feedback about various aspects of the website.

In the second round, I used similar protocols for member-checking and for data analysis as I had in the first round. At the conclusion of this round of interviews, I updated the website with the insights gained and developed a virtual negotiation forum around the most prominent issues that stakeholders chose.

4.5.3 Round 3 of data collection: The negotiation session

During this last round, the negotiation session, participants were engaged in negotiations about the most prominent issues nominated in the second round. In preparation for this round, I compiled the arguments made by stakeholders during the previous rounds and identified reflective questions for each nominated issue. I also compiled the suggestions that were made about these issues and presented them separately as proposals worthy of discussion. Participants were then invited to comment on the questions and proposals.

Ideally, the negotiations should continue until all participants agree on the resolution or non-resolution to the issues discussed. However, as acknowledged by Guba and Lincoln (1989) and ascertained in published research studies (Huebner & Betts, 1999; Kosh, 2000; Lay &

Papadopoulos, 2007), this is often not a realistic goal. Therefore, the online forum remained active for six weeks only, during which time participants could access it an unlimited number of times. Once the negotiation session was completed, an overall report containing the CCIs as well as propositions and suggestions for improvement of the evaluand was developed and sent to all participants.

4.6 Data collection tools

As explained in the previous section, data collection was carried out using interviews, field notes and observations, and a virtual negotiation forum. These instruments are explained in detail in the following sections.

4.6.1 Interviews

Interviews are essential tools for accessing people's perspectives about phenomena, particularly when the information sought cannot be obtained through observation or other means (Patton, 1990). Interviews were the main instruments for collecting data throughout the study. They were each approximately one-hour long and were used for two purposes: the administration of the A4GE and the collection of data about its practice. For the administration of the A4GE, interviews were used to elicit the participants' CCIs, perceptions and interpretations about the GDE(ST). For the second purpose, interviews provided participants' opinions about the A4GE and about its strengths and limitations as a model for program evaluation.

All face-to-face interviews were recorded using two digital voice recorders (a second for backup). Skype-based interviews were also recorded using Pamela software as well as the digital voice recorders. Interviewees were briefed on the need for recording the interviews for the purposes of later transcription and analysis through the consent form information package sent to them at the time of recruitment. Their permission for recording was taken again prior to the commencement of the interviews.

The structure of an interview can vary from being highly structured or standardized to highly unstructured and informal (Merriam, 1998, p. 74). Different interview types are used for different purposes and each has its own strengths and limitations. With structured interviews, a researcher asks all respondents the same questions, in the same sequence and using the exact same wording (Cohen et al., 2007). As the interview loses more of this structure, the questions become less standardized and the researcher has more freedom in ordering the

sequence of the questions and even adding more questions in the form of probes. These interviews are referred to as semi-structured interviews since the researcher is still guided by a set of pre-formulated questions but is not restricted to them as in structured interviews. Non-structured interviews are the most flexible of all types of interviews and can even take the form of informal discussions (Merriam, 1998). In these interviews, the researcher is not guided by an interview protocol but rather comes with a clear purpose and one or two questions. The flow of the interview and the responses of the interviewee often guide the direction of the interview and provoke further questions and probes.

In this study, interviews were largely unstructured at the beginning of the first round of data collection and, as more interviews were being conducted, they became more focused and structured. This is in alignment with the methodology of the A4GE, which specifies that the first interview should be unstructured and aimed at unveiling the claims and concerns of the first interviewee. The following interview would be based on the first and the second interviewee is asked to not only represents his/her own claims and concerns but also to comment on the first interviewee's construction. In order to scaffold these comments, I analyzed the first interview and summarized the claims and concerns and presented them to the second interviewee. In this way, the interview protocol for the second interviewee gained more structure than the first. Using the same approach, the third interviewee built on the joint construction presented by the first two interviewees. Therefore, to scaffold for this process, I focused my interview protocol even further to probe the interviewee's comments on the joint construction and on the additional claims and concerns revealed by the second interviewee. Continuing in this manner, while interviewing all the participants, the interviews gained more and more structure and focus. Nonetheless, at no point were the interviews rigidly structured and they did not restrict interviewees from expressing their own constructions. They were flexible enough for me to ask follow-up questions as appropriate. A sample of an interview protocol is presented in Appendix C.

Skype was used extensively during the data collection. The use of Skype as an alternative to face-to-face interviewing offered major benefits: Skype is free software that can be downloaded and installed easily; the interface is user-friendly which allows the few (if any) unfamiliar participants to become acquainted quickly with the software; the software offers synchronous interaction between interviewer and interviewee, which is a hallmark of interpretive interviewing where the researcher seeks clarification from the interviewees throughout the interview (Kvale & Brinkmann, 2009); and it was cost effective as it allowed

a dispersed population to become involved in the study – indeed, use of Skype opened access to participants who would otherwise have been excluded due to their geographical location, a particularly important asset since most of the teachers were dispersed and the costs associated with traveling and related expenses would have made them inaccessible to the study.

One drawback of using Skype was potential technical failures, which could jeopardize the flow of the interviews and cause loss of valuable data. To overcome that hurdle, I stressed the importance of calling from computer to computer because the Internet connection using that medium is more reliable than exists when using mobile networks. Loss of nonverbal cues (hand gestures, body language) is another disadvantage associated with the use of audio communication through Skype. Nonetheless, this disadvantage was outweighed by the advantage of accessing geographically dispersed people, though I encouraged video conferencing via Skype when possible as a potential way to enhance rapport between the interviewees and myself.

The interviews for the first round of data collection were piloted using a sample of three participants: one university lecturer and two experienced secondary science teachers. The purpose of the pilot tests was to check the appropriateness of the questions I had formulated as well to test the media of communication, since I had planned on using Skype where face-to-face interviews were not possible. The pilot tests helped illuminate a number of issues that were subsequently accounted for in the interviews, for example drop out, delays and unclear sound from mobiles, especially when the video feature was used. Moreover, one pilot interviewee said that he needed to see a video image of me in order to feel that he was having a comfortable conversation with a real person.

At the completion of each interview, a soft copy of the recordings was sent to a transcription company. The interviews were typically transcribed in three working days. The transcripts were verbatim and included “stutters, mumblings and onomatopoeia” (Cooksey and McDonald, 2011, pp.485) so as to capture and transparently portray the extent to which the arguments and reasoning of the participants were clear and where there was ambiguity or uncertainty. I reasoned that incorporating these elements in the transcription would further enhance the trustworthiness of my study.

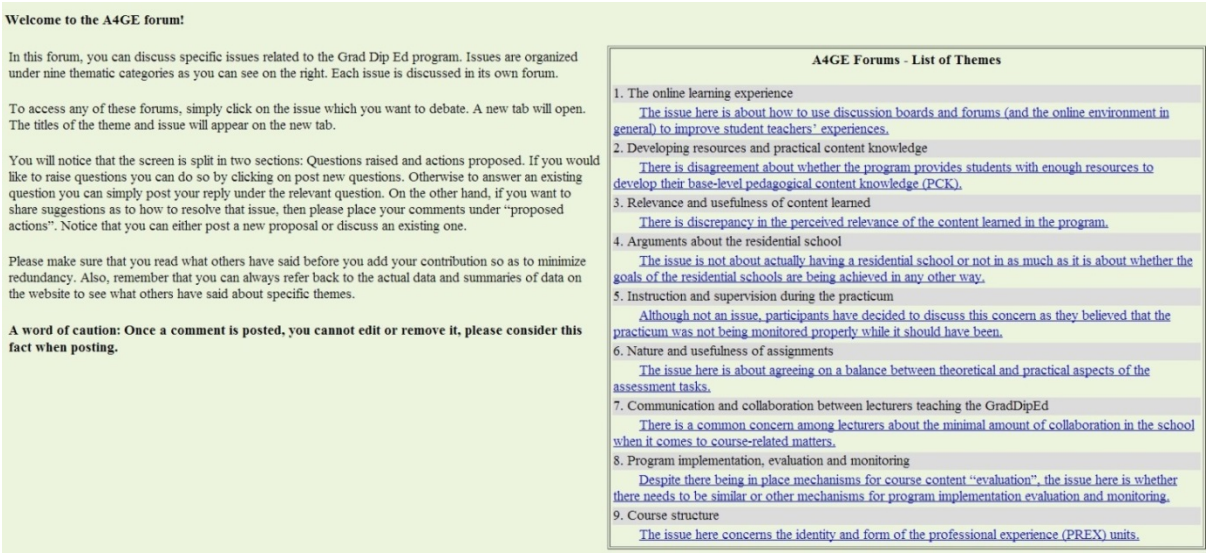
4.6.2 Online negotiation forum

The second instrument for data collection consisted of the virtual negotiation forum, which was administered after the two rounds of interviews. Conceived as an attempt to overcome

the hurdles often associated with face-to-face negotiation, the forum is an original contribution to the development and modernization of the 4GE. While I fully designed the forum, a paraprofessional web developer assisted me with the execution of the design. His expertise helped redefine some aspects of the tool to make it more user-friendly.

The online forum was embedded within the evaluation website. Therefore, when discussing the issues, participants could refer to the findings from the study to understand the issues further. To help participants become acquainted with the forum, I developed a tutorial and published it on YouTube: http://www.youtube.com/watch?v=zh5UlHBfp7k. The video and forum were both piloted with colleagues who checked whether they were easy to understand, use and navigate.

The main forum webpage (see Figure 4.5) contained nine issues that were suggested by participants in the second round of data collection.



Welcome to the A4GE forum!

In this forum, you can discuss specific issues related to the Grad Dip Ed program. Issues are organized under nine thematic categories as you can see on the right. Each issue is discussed in its own forum.

To access any of these forums, simply click on the issue which you want to debate. A new tab will open. The titles of the theme and issue will appear on the new tab.

You will notice that the screen is split in two sections: Questions raised and actions proposed. If you would like to raise questions you can do so by clicking on post new questions. Otherwise to answer an existing question you can simply post your reply under the relevant question. On the other hand, if you want to share suggestions as to how to resolve that issue, then please place your comments under "proposed actions". Notice that you can either post a new proposal or discuss an existing one.

Please make sure that you read what others have said before you add your contribution so as to minimize redundancy. Also, remember that you can always refer back to the actual data and summaries of data on the website to see what others have said about specific themes.

A word of caution: Once a comment is posted, you cannot edit or remove it, please consider this fact when posting.

A4GE Forums - List of Themes

1. The online learning experience
   The issue here is about how to use discussion boards and forums (and the online environment in general) to improve student teachers' experiences.
2. Developing resources and practical content knowledge
   There is disagreement about whether the program provides students with enough resources to develop their base-level pedagogical content knowledge (PCK).
3. Relevance and usefulness of content learned
   There is discrepancy in the perceived relevance of the content learned in the program.
4. Arguments about the residential school
   The issue is not about actually having a residential school or not in as much as it is about whether the goals of the residential schools are being achieved in any other way.
5. Instruction and supervision during the practicum
   Although not an issue, participants have decided to discuss this concern as they believed that the practicum was not being monitored properly while it should have been.
6. Nature and usefulness of assignments
   The issue here is about agreeing on a balance between theoretical and practical aspects of the assessment tasks.
7. Communication and collaboration between lecturers teaching the GradDipEd
   There is a common concern among lecturers about the minimal amount of collaboration in the school when it comes to course-related matters.
8. Program implementation, evaluation and monitoring
   Despite there being in place mechanisms for course content "evaluation", the issue here is whether there needs to be similar or other mechanisms for program implementation evaluation and monitoring.
9. Course structure
   The issue here concerns the identity and form of the professional experience (PREX) units.

Figure 4.5 – Homepage of the negotiation forum

As can be seen in Figure 4.5, the Home Page of the forum features a welcome note that guides participants on how to use the forum. The page also includes a list of the nine nominated issues. For each issue, I prepared a separate forum which could be accessed by clicking on the issue statement from the forum Home Page. Once participants click on an issue, a new tab opens that features the forum page for that particular issue. An example forum page is shown in Figure 4.6.

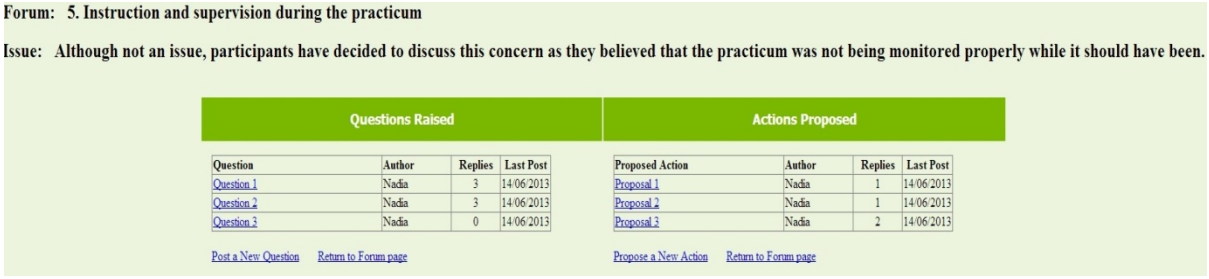


Figure 4.6 – Example forum page for an issue

As clarified in Figure 4.7, the forum for each issue contains two compartments, one for discussing the questions related to that issue and one for discussing the proposals and improvement suggestions that relate to that issue. To develop the content of the forums, I compiled, for each of the nine nominated issues, a list of reflective questions that I had constructed with participants during the second round of interviews. For each issue, I also developed a list of proposals that I derived from the two rounds of interviews. Using these forums, negotiators could click on any of the questions and proposals and insert their comments there. Other negotiators could then add to the comments or even ask new questions and make new proposals. These additions will then appear on the relevant issue forum page.

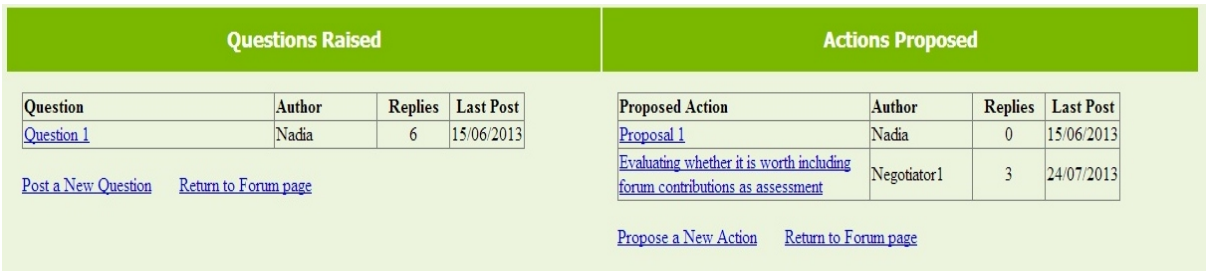
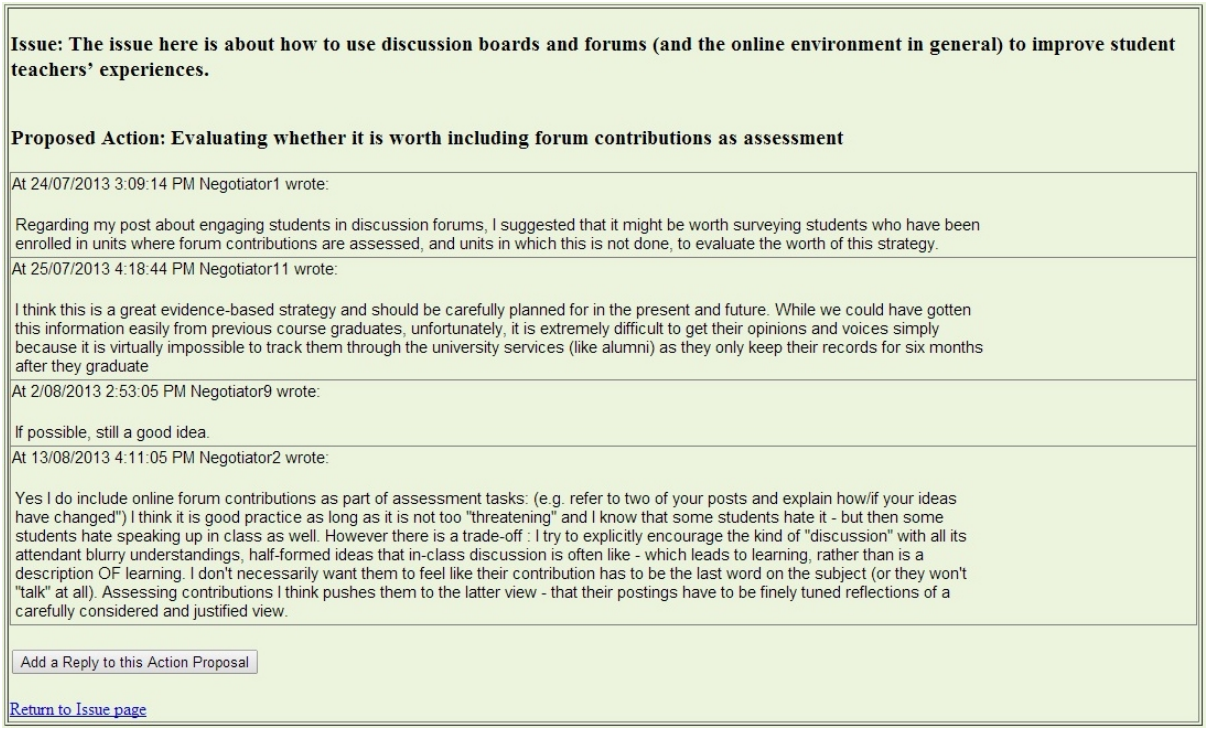


Figure 4.7 – Compartments of the negotiation forums

When participants click on a question or on a proposal, a new tab containing the discussions about the question or proposal opens. A sample discussion of a proposal is shown in Figure 4.8.



Issue: The issue here is about how to use discussion boards and forums (and the online environment in general) to improve student teachers' experiences.

Proposed Action: Evaluating whether it is worth including forum contributions as assessment

At 24/07/2013 3:09:14 PM Negotiator1 wrote:

Regarding my post about engaging students in discussion forums, I suggested that it might be worth surveying students who have been enrolled in units where forum contributions are assessed, and units in which this is not done, to evaluate the worth of this strategy.

At 25/07/2013 4:18:44 PM Negotiator11 wrote:

I think this is a great evidence-based strategy and should be carefully planned for in the present and future. While we could have gotten this information easily from previous course graduates, unfortunately, it is extremely difficult to get their opinions and voices simply because it is virtually impossible to track them through the university services (like alumni) as they only keep their records for six months after they graduate

At 2/08/2013 2:53:05 PM Negotiator9 wrote:

If possible, still a good idea.

At 13/08/2013 4:11:05 PM Negotiator2 wrote:

Yes I do include online forum contributions as part of assessment tasks: (e.g. refer to two of your posts and explain how/if your ideas have changed") I think it is good practice as long as it is not too "threatening" and I know that some students hate it - but then some students hate speaking up in class as well. However there is a trade-off : I try to explicitly encourage the kind of "discussion" with all its attendant blurry understandings, half-formed ideas that in-class discussion is often like - which leads to learning, rather than is a description OF learning. I don't necessarily want them to feel like their contribution has to be the last word on the subject (or they won't "talk" at all). Assessing contributions I think pushes them to the latter view - that their postings have to be finely tuned reflections of a carefully considered and justified view.

Add a Reply to this Action Proposal

Return to Issue page

Figure 4.8 – Sample discussion of a proposal

Once the forum was designed and developed, I announced the start of the negotiation session. I sent an email to all participants together with new usernames and passwords. The new usernames were used to conceal the identities of the negotiators. Previous usernames indicated whether the participant was a teacher or a lecturer (ST1, UL2 and so on). In the negotiation stage, I used the username "Negotiator" and assigned a number next to it (e.g. Negotiator 1). I, too, was assigned a similar username because I was a member of the negotiations, as suggested in the A4GE.

Some of the assets of the online negotiation forum design are:

- Participants have concealed identities: participants could contribute their opinions freely without being intimidated by the status of other participants because the real identities of the participants were replaced with codes, that is, Negotiator 1, Negotiator 2 etc.
- The negotiation session was carried out asynchronously; therefore, participants did not have to provide immediate responses or comments. Rather, they were able to think about their responses and formulate their arguments at their convenience.
- Since the forum spanned six weeks, participants could access resources to enrich their statements and support them with evidence. The website also offered access to some of these resources (such as the AITSL requirements).

– My role in the negotiation forum as moderator of the discussion was less intense since I, too, had time to think about the prompts and responses I was using to scaffold the discussion.

4.6.3 Field notes and journal

Playing a dual role as the evaluator using the 4AGE and researcher of the administration of the A4GE, I was able to derive further data for this study through taking field notes from observations outside the formal interview situation, as a participant in the day-to-day life of the school. In the field notes, I recorded important university events that I thought were connected to the study because I believed that they provided information needed to develop the thick descriptions that are usually used in case study research. Therefore, any event (seminar, meetings, conference etc.) that related either to the evaluation of the program or to the practice of evaluation in general was included in my field notes journal.

Additionally, I kept a research journal, which contained my personal insights, feelings and interactions with the participants that were not captured by the interviews. The journal also included an audit trail in which I included all the decisions I made about the administration of the A4GE as well as its empirical investigation so as to capture the insights I derived about the processes of the emerging design. Any occasions where the evaluation processes deviated from the original design were noted, dated and explained for later analysis and discussion.

4.6.4 Documents

Merriam (1998) suggests that documents be used in case studies because they can potentially reveal things that happened prior to the investigation, things that cannot be investigated and things that the inquirer does not know about (p. 114).

In this study, various documents were used to complement the information derived from the interviews, the negotiation forum and the field notes. These included evaluation policies and records as well as the syllabuses of the units taught into the GDE(ST). While all these documents were content analyzed and used to enrich the situational analysis of the context of the current evaluation, the GDE(ST)-related evaluation policies and records provided a frame of reference against which I developed assertions about the usefulness and drawbacks of the A4GE.

4.7 Data analysis

In accordance with the interpretive paradigm, data analysis was carried out concurrently with data collection and the two activities were integrated (Lincoln & Guba, 1985; Skrtic, 1985). This enabled me to develop further questions throughout the data collection phase based on emerging insights from the analysis. Two analysis techniques were used in this study. For the first component of the research, the evaluation of the GDE(ST), I used constant comparison (Charmaz, 2011). For the second component, the empirical investigation of the application of the A4GE, I used constant comparison (Charmaz, 2011) and negative case analysis (Robinson, 1951). In what follows, I explain how I deployed these analytical strategies.

4.7.1 Component 1: Evaluating the GDE(ST) using the A4GE

For the purposes of analyzing the data pertaining to the evaluation of the GDE(ST), I used the content analysis protocol described in the constant comparison method (Charmaz, 2011; Glaser & Strauss, 1967). Merriam (1998) explains that while the constant comparative method had been initially conceived as a tool for building grounded theory, it has been used in other qualitative approaches to research because of its inductive, concept-building orientation. The constant comparison method suggests that newly gathered data are constantly compared to previously collected data so as to allow for conceptual categories and patterns to emerge (Charmaz, 2011). One way to achieve these comparisons is through coding. As researchers read through the data, they assign codes to important bits of data in a process known as open coding (Gibbs, 2007). Subsequently, as researchers skim through the list of codes that emerged through open coding, they can identify similar codes and themes in a process known as thematic coding (Gibbs, 2007).

While I used constant comparison to analyze the data for this first component of the research, the analysis of data from the first two rounds of data collection (the interviews) occurred separately to the analysis of the data from the negotiation session (the third round).

4.7.1.1 Protocol for analyzing data from the interviews

The analysis of the interview data pertaining to the A4GE proceeded in two stages that were repeated after the end of each interview and data collection round. In the first stage, I used open coding. I started reading through the interview scripts line by line in order to identify codes. Each relevant excerpt from the data was assigned two codes based on its content and its nature (Claim or Concern, or even suggestion as per the adaptation introduced to the 4GE). Since I was using NVivo 10 during this coding process, each code was assigned a

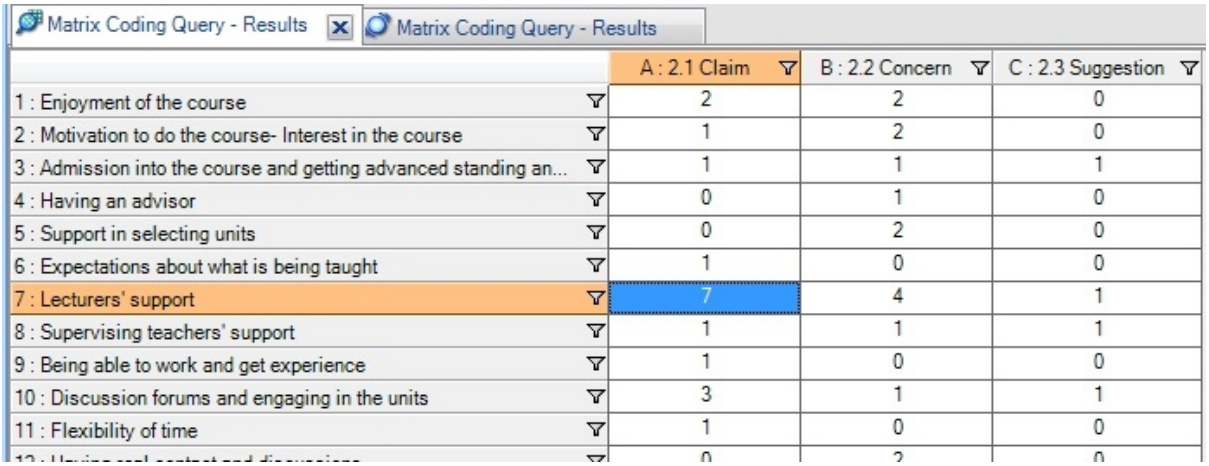
corresponding node in NVivo10. A node is a virtual container that includes all the data related to a particular code. NVivo10 enabled me to track the data excerpts to their original sources and, therefore, it was not necessary for me to assign further codes about the sources of the data excerpts. Following this process, I developed an extensive list of codes from the data. As I progressed with this process of open coding, I kept note of the emerging definitions for each code on a separate memo within NVivo 10. Importantly, since no predetermined codes were in place to guide the open coding process, I employed a variety of strategies to inductively develop and name the codes. These strategies are summarized in Table 4.4.

Table 4.4 – Techniques for developing codes

Technique	Associated questions
Asking questions	Who, what, when, where, why and how of a phenomenon. Answers to each question are potential codes.
Strauss' coding paradigm (Strauss, 1987)	What actions or interactions are occurring? What strategies are being applied? Under what conditions? With what consequences?
Lyn Richards's guiding questions (Richards, 2005)	Identify what is interesting Ask why is it interesting Ask why am I interested in that
Recording narrative structures and mechanisms that could be as revealing as verbal data (these were noted in memos)	Sudden change of theme or avoiding a particular subject Inconsistencies in talking Use of analogies Sarcasm or other emotional cues Reference to personal/other people's stories

Once I finished open coding each interview, I engaged in thematic coding and developed themes through an iterative process of comparison and contrast of the open codes. NVivo 10 was particularly useful with this task as it enabled me to instantly migrate the data relating to open codes into the emerging themes. Each theme was then assigned a corresponding node in NVivo10. To keep track of the themes and their definitions, I also used a memo in NVivo10 in which I identified and defined the open codes included under each theme.

Once all data were coded, I used a feature of NVivo10 called "Coding Query" to generate a two-dimensional matrix (see Figure 4.9) which organized the data from all the sources according to theme and nature and granted me access to the data excerpts relevant to each theme.



Matrix Coding Query - Results | Matrix Coding Query - Results

	A : 2.1 Claim	B : 2.2 Concern	C : 2.3 Suggestion
1 : Enjoyment of the course	2	2	0
2 : Motivation to do the course- Interest in the course	1	2	0
3 : Admission into the course and getting advanced standing an...	1	1	1
4 : Having an advisor	0	1	0
5 : Support in selecting units	0	2	0
6 : Expectations about what is being taught	1	0	0
7 : Lecturers' support	7	4	1
8 : Supervising teachers' support	1	1	1
9 : Being able to work and get experience	1	0	0
10 : Discussion forums and engaging in the units	3	1	1
11 : Flexibility of time	1	0	0

Figure 4.9 – Section of the "Coding Query" matrix generated with NVivo 10

The first column in the matrix illustrates the various themes that were discussed by the participants. These were the content related codes. The first row indicates the codes related to the nature of the statements (claim, concern, or suggestion). The cells in the matrix represent the number of participants who expressed claims, concerns or suggestions about each theme discussed in the interviews. In Figure 4.9, the highlighted cell clarifies that there were seven participants who expressed claims about the theme "Lecturer's support". The figure also reveals that four other participants expressed concerns about that same theme.

Coding Queries can also be performed to retrieve data relevant to different sources (e.g. Lecturer 1 or Science teachers 5 and 6) and from different rounds of data collection (e.g. round 2 of interviews). Therefore, by clicking on an entry of this matrix, I could access, for example, all the data pertinent to the science teachers in the second round of data collection. Alternatively, I could get the statements expressed by one specific university lecturer from the negotiation session and so on.

By double clicking on the highlighted cell in figure 4.6, I could access the claims from all participants in a new tab on the Nvivo interface (see figure 4.10).

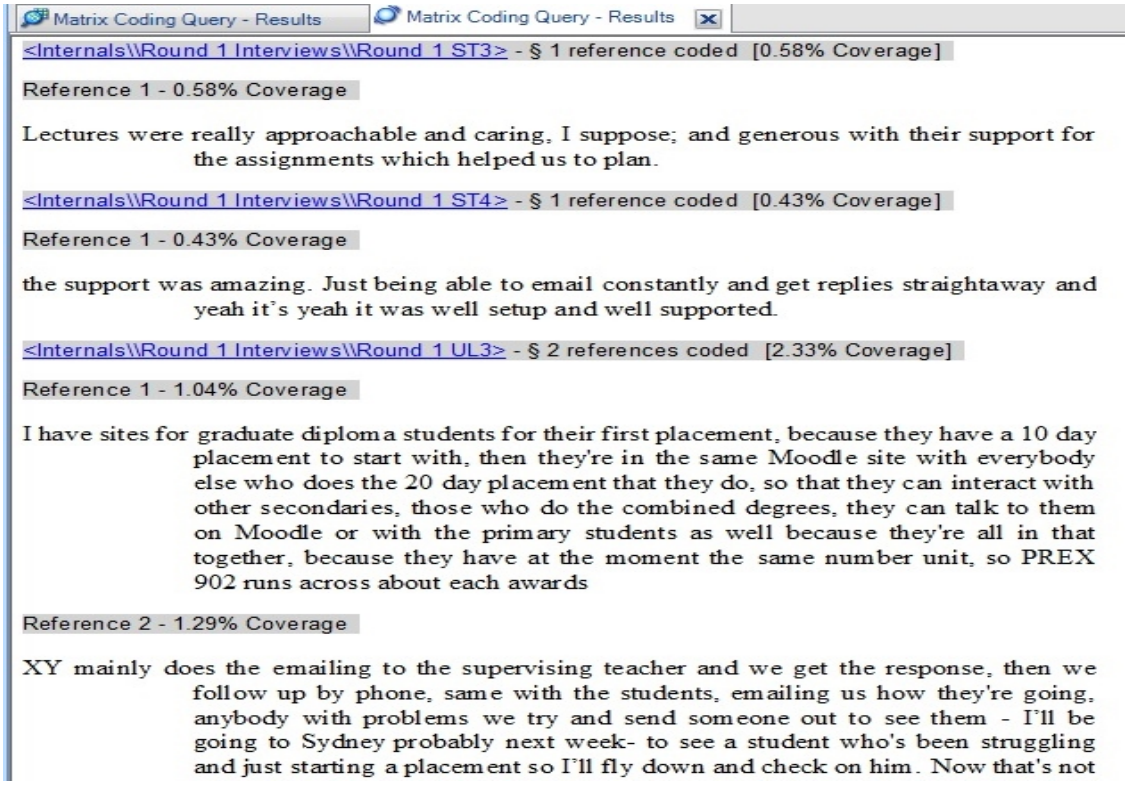


Figure 4.10 – Claims about the theme "Lecturers Support"

The data displayed in Figure 4.10 correspond to the claims made by all participants about the theme "lecturer's support".

Using the coding query matrix, I compared and contrasted the various claims and concerns to develop lists of issues that existed within stakeholder groups and between stakeholder groups. To this end, for each stakeholder group, I looked through the data for each theme and examined whether that theme was an issue for that group. That is, if informants from the group disagreed about the nature of that theme (some considered it a claim while others considered it a concern), the theme was noted as an issue for that stakeholder group. For each theme where an issue was identified, I copied the data representing the claims and concerns to a word document and content analyzed them to define the issue related to that theme and to develop a narrative that describes the areas of contention between participants. Continuing in this fashion, I was able to identify the issues for each stakeholder group. These issues were tagged as within-group issues. The next step was to identify between-group issues, which are themes perceived as claims by one stakeholder group and concerns by the other.

4.7.1.2 Protocol for analyzing data from the negotiation session

The analysis of data from the negotiation round was also carried out using the constant comparative protocol. I considered each issue that figured in the negotiation forum one at a time. Then, I content-analyzed, compared and contrasted the arguments that were advanced under that issue. I grouped similar or complementary arguments and, together with their counter arguments, I reported them in a narrative format. In these narratives, I presented the resolved issues – those which forum participants were able to resolve – as well as unresolved issues. I also discussed how the latter were redefined in the negotiations.

4.7.2 Component 2: Empirical investigation of the application of the A4GE

The analysis of data for the empirical investigation of the relationship between the theory and practice of the A4GE consisted of two steps. In the first step, I used content analysis to: sort through the data from the evaluation case study and my journal and field notes and classify instances from practice that were related to each theoretical proposition under each PEMED dimension. These data included events, activities and ideas that were associated with these propositions. Once this step was completed, I proceeded to the second step, in which I used constant comparison (Charmaz, 2011) and analytic induction, also known as negative case analysis (Robinson, 1951). That is, I considered the data set peculiar to each theoretical proposition along each PEMED dimension at a time. I compared and contrasted each bit of data from the data set to examine whether the practice of the A4GE aligned with its theoretical propositions. Where alignment was found, I noted it and I moved to the next piece of data. When misalignment occurred, I developed a description of how that proposition was enacted in practice and how it differed from the theoretical proposition. Then I moved to the second bit of data and proceeded in the same way. Where there was alignment with the theoretical proposition, I noted it, and where there was misalignment, I used that piece of data to revise and broaden my description of practice. This process continued until all the data in the data set peculiar to that theoretical proposition were exhausted. After repeating this process for all the theoretical propositions, I ended with, for some of the PEMED dimensions, two sets of information: the theoretical propositions initially formulated using the theory of the A4GE and the descriptions of practice constructed from the data. These sets assisted me in developing interpretations about how the theory and practice of the A4GE aligned and where and how they differed.

In taking the analyses a step further, I then used Miller's (2010) criteria, discussed in Chapter 2, to examine the relevant theoretical propositions that were the most informative about each

of the four criteria. Arranging the theoretical propositions under the umbrellas of the criteria and examining the extent to which they were congruent with practice, I was able to make assertions about the operational specificity of the A4GE, its range of application, its feasibility in practice and its discernible impact. These assertions enabled me to answer the research question. Additionally, and through further content analyses of my journal entries and field notes, I was able to gain insights into contextual factors that affected the practice of the A4GE in relation to its theory.

4.8 Addressing rigor criteria

According to Lincoln and Guba (2007), criteria used to ensure rigor in the traditional post-positivist paradigm include the internal and external validity of the research, its applicability, its consistency and its neutrality. Merriam (1998) argues that these criteria cannot be achieved in their conventional sense in the case of interpretive research. Lincoln and Guba (2007) explain that the assumptions of the interpretive paradigm deny the existence of a single reality that can be objectively known by the researcher and, therefore, sharply contrasts the post-positivist criteria. Alternatively, the interpretive inquirer must ensure that the parallel trustworthiness (credibility, transferability, dependability and confirmability) and authenticity criteria are all met (Lincoln & Guba, 1985). In this section, I describe the steps I have taken to enhance the rigor of this study.

4.8.1 Trustworthiness criteria

Credibility, the first of the trustworthiness criteria, is a measure of the congruency between the findings and the data (Merriam, 1998). To ensure the credibility of the findings from this study, I used a variety of data collection sources and methods (i.e. interviews, online forum, observations and field notes) and triangulated the range and depth of the findings using these varied sources (Guba, 1990). Additionally, as recommended by Miles and Huberman (1994), I used member checks. Informants were sent a copy of the transcripts of the interviews to check whether they reflected their opinions. Similarly, the second round of interviews provided another opportunity for member checking since the informants were confronted with their stated positions from the first interview round and asked to comment on it. This was a window to correct the interpretations made by the researcher about the informants' opinions and positions.

Transferability refers to the extent to which the findings from one study can be transferred to other cases or situations (Merriam ,1998). To ensure that this criterion is met, Lincoln and

Guba (2007, p. 19) suggest that the investigator should provide thick and sufficient contextual information so that "judgments about the degree of fit or similarity may be made by others who may wish to apply all or part of the findings elsewhere". In this study, the use of thick descriptions of the processes and outcomes of each of the two components of this study provided the necessary base to enable transferability judgments.

In order to address the dependability criterion, the processes within the study were noted and reported in detail and the theoretical as well as practical assumptions made were portrayed as transparently as possible. According to Shenton (2004), these activities should be clearly reported to enable another researcher to replicate the procedures in different contexts and to understand the context-bound decisions that were made during the implementation of the research design. The transparent description of the methodology allows the integrity of research results to be scrutinized, thus ensuring the dependability criterion is met. Furthermore, the specifics of the data collection, data analysis and interpretations were also reported in detail so as to clarify all the decisions made regarding those activities.

To meet the confirmability criterion, the researcher must demonstrate that the findings from the study are the result of the informants' ideas and interpretations rather than being the researchers' preferences (Shenton, 2004). To that end, Miles and Huberman (1994) suggest that researchers must reveal their own biases and dispositions. In this study, I kept a journal log to record all the decisions that I made during the data collection and analysis. This audit-trail (Lincoln & Guba, 1985) enabled me to trace back decisions I made throughout the entire investigation.

4.8.2 Authenticity

The preceding trustworthiness criteria parallel the rigor criteria of the post-positivist paradigm. However, they cannot be considered as a complete set for ensuring the rigor of an interpretive investigation because they only address those issues that are important from a positivist perspective and they, particularly, ignore the influence of the context on the investigation. For this reason, Lincoln and Guba (2007) present five additional criteria for establishing rigor that are born out of the interpretive nature of inquiry:

4.8.2.1 Fairness

To ensure that this criterion is met, the researcher must acknowledge the value-laden nature of any participant's construction. Consequently, he/she must "expose and explicate these several, possibly conflicting, constructions and value structures" (Lincoln & Guba, 2007, p.

20). To achieve fairness in the administration and analysis of the A4GE, and following Lincoln and Guba (2007), I tried to uncover the different belief systems that the participants held and used to defend their positions. These included their philosophies of teaching and learning, their views about teacher education principles, as well as their views about educational institutions and how these should be structured and run. This was particularly evident in the interviews I carried out when I was able to identify clear conflict between participants' views. Furthermore, since these conflicts were identified through the concurrent analysis of interviews, participants were given another chance to fairly refine and even change their positions during the second interview round and the negotiation forum.

Lincoln and Guba (2007) also assert that the fairness criterion:

> requires fully informed consent with respect to any evaluation procedures [which] is obtained not only prior to an evaluation effort but is continually renegotiated and reaffirmed (formally with consent forms and informally through the establishment and maintenance of trust and integrity between parties to the evaluation) as the design unfolds, new data are found, new constructions are made, and new contingencies are faced by all parties. (p. 22)

Therefore, and prior to every interview or interaction with the participants, I started by explaining the purposes and goals of the evaluation and made sure to answer all of the informants' information needs. In relation to the empirical investigation, I attempted to uncover the constructions that all participants – including myself – held concerning the evaluation processes and outcomes. Since many of the participants are also researchers, I tried to uncover their paradigmatic assumptions to clarify their views and opinions about the A4GE.

4.8.2.2 Ontological authentication

According to Lincoln and Guba (2007), since reality is based on an individual's constructions and experiences, investigators should strive to "raise consciousness … so that a person or persons (not to exclude the evaluator) can achieve a more sophisticated and enriched construction" (p. 22). In both components of this research, I used various questioning techniques during the interviews to raise the awareness of the participants to other participants' constructions, making sure to pinpoint issues as well as areas of consensus. These strategies ensured that participants were conscious of different views and enabled them to revise their own constructions based on this understanding, if they so chose.

4.8.2.3 Educative authentication

Lincoln and Guba (2007) argue that it is essential that participants appreciate other participants' constructions and understand the differing value systems underlying their different perspectives. Therefore, to ensure the educative authentication criterion is met, participants should be given opportunities to enhance their appreciation and understanding of the various constructions. In this study, and for the two components of the investigation, various venues for sharing information and interacting with the emerging findings as well as the actual data were available participants in a timely and transparent way. For example, the online interactive website provided participants with access to the various claims, concerns and suggestions presented by other informants.

4.8.2.4 Tactical and catalytic authentication

Lincoln and Guba (2007) argue that interpretive inquiries must not only facilitate but also stimulate action. Additionally, the researcher must employ techniques to ensure that participants feel and become empowered to act. In this study, I deliberately asked participants to present their suggestions about possible ways to: (1) resolve persistent issues related to the program (for the purposes of ensuring this criterion is met for the A4GE administration); and (2) improve the practices of the A4GE and its administration (for the purposes of ensuring this criterion is met for the empirical investigation of the A4GE). Furthermore, in the virtual negotiation forum, there was an entire section dedicated to proposals and suggestions from participants. Through making this clear distinction in the forum between theoretical discussions and action-oriented discussions, I stimulated participants to formulate and present their suggestions for potential actions and problem-solving strategies.

4.9 Ethical considerations

An ethics application was submitted to the Human Resources Ethics Committee at the University of New England and approval was obtained to carry out the research (HE12-072) prior to the commencement of the fieldwork. At the time of participant recruitment, a written information package and a consent form detailing the processes and objectives of data collection as well as the expected roles and duties of the participants were sent to each of the invited participants. The document informed participants that their involvement was completely voluntary and that they could withdraw their consent at any time without any kind of penalty. The document also specified that the researcher was willing to provide further explanations and clarifications if requested. Furthermore, the information package clarified

that the data would be used anonymously and any identifiers within the data would be removed prior to publishing the results. Additionally, the consent form asks for permission for recording the interviews for the purposes of transcription and later analysis. However, prior to conducting the interviews, informants were asked again for their permission to record and only after their approval was granted did I start recording. The conditions of confidentiality were also explained in both the information package and orally prior to every interview. The information package is presented with the consent form in Appendix D.

In accordance with the recommendations of the Ethics Committee (HREC), all data collected need to be stored for five years after the end of the PhD candidature. Therefore, digital data were saved on a password-protected computer and all printed data are kept in a locked cabinet. All data documents were also de-identified and coded, and the codes list is kept in a separate locked cabinet. Additionally, data shared on the website were de-identified. For example, science teachers were given codes ST1, ST2 and so forth, while university lecturers and the Head of the School of Education at the university were given the codes UL1, UL2 and so forth. These codes were kept intact for the second round of interviews and during the negotiation so that all participants could track the opinions of one another if they wanted to without knowing the real identity of the participants. It is worth mentioning that the website itself is password-protected (a guest username and password can be provided upon request) and informants were only able to access it through a user name and password for the duration of the data collection. At the end of the data collection phase, I was the only one with a valid username and password who could access the site.

One particular ethical dilemma arose from including two of my PhD supervising mentors as sources of information in the study. The inclusion of the supervisors as informants is central to this research study as they are both key players in the development and delivery of more than one essential unit in the science teacher education program under consideration. The involvement of the supervisors in the research raises the prospect of potential bias which, if not attended to properly, might jeopardize the credibility and reliability of the interpretations and results. To address this potential bias, I involved several stakeholders in a discussion about the program under consideration, including other informants who provided alternative views of the issues discussed with the supervisors. This approach enabled the triangulation of the data since different perspectives are presented. Further, the potential for conflict of interest was discussed with my supervision team and my third supervisor, who was neither involved in the GDE(ST) nor a participant in the study, monitored the research for any bias.

More importantly, since the focus of the study was on the evaluation approach and not on the evaluation of the science teacher education program in which the supervisors are involved, any potential bias concerning the program was thus marginal. Nonetheless, as a precaution and to enhance transparency, all transcripts of the interviews with the supervisors are available on request (without identifying the supervisors), to demonstrate their relationship to the analyses and interpretations derived from them.

4.10 Conclusion

In this chapter, I have explored the methodology of the current investigation, clarifying the research design for the two components and outlining the procedures of data collection and analysis. Also, I have discussed how I addressed the rigor criteria in my study. In the following chapter, I present an elaborate description and discussion of the findings from the evaluation case study.

CHAPTER 5: FINDINGS FROM THE APPLICATION OF THE A4GE

The main purpose of this study was to investigate the relationship between the theory and practice of the A4GE. In this context of research, the evaluation of the GDE(ST) was an application of the A4GE to provide a context of practice where the A4GE could be empirically investigated. Nevertheless, since the A4GE was framed as a protocol for a university faculty's review – if they were to use an interpretive framework for carrying out their reviews of academic programs, its findings represent a significant aspect of the dissertation. More importantly, since the processes for carrying out the A4GE are more integral to this investigation and more important than its outcomes, the report illustrated in this chapter describes all the stages of data analysis: from initial data reduction and summarization into themes; to closer examination of CCIs; to a final focus on the Issues and the negotiation process. This level of detail in my reporting was necessary to enable me to refer to both the processes and outcomes of the A4GE in my discussion of the application of the A4GE in Chapter 6. Therefore, this chapter is structured to show development of the findings from the data and is more methodical and detailed than would ordinarily be expected of a faculty program review.

This chapter is in three sections. In the first section, I present and discuss the various constructions, categorized under themes that emerged from the evaluation. This representation of the findings allows the CCIs of the various stakeholders to be browsed thematically. In the second section, I present lists of CCIs. This enables the reader to access the various Claims, Concerns or Issues and to understand how they are similar or different for the stakeholder groups. In the last section, I present and discuss the issues that were raised in the negotiation session and describe whether and how these were resolved.

If the A4GE were to be used in the context of a school or faculty review, the presentation of the themes (section one) would only form a minute part of the A4GE report. Section two, on the other hand would represent the bulk of that report. The third section would only include the results from the negotiation sessions.

5.1 Themes discussed in the interviews

The themes that emerged from the analysis of the interview data are summarized and defined in Table 5.1.

Table 5.1 – Themes discussed by informants during data collection

Theme	Description and aspects
1. Attitudes towards the program	These included expressions about the extent to which informants enjoyed the program, what fuelled their motivation and what they believed was interesting about the program.
2. Guidance and support	Under this theme, informants discussed their beliefs about the guidance and support they received from the university personnel during two stages of interacting with the program: prior to enrolling and after enrolling.
3. Delivery mode	This theme includes informants' ideas about the two modes of program delivery: on-campus and off-campus.
4. Content of learning	This theme includes informants' ideas about the content covered in the program as well as the extent and quality of coverage.
5. Residential school	Under this theme, I grouped the informants' arguments for or against the residential school component of the program.
6. Teaching strategies in the program	This theme includes the informants' views about the teaching strategies adopted by the various lecturers into the program.
7. Program Monitoring Processes	This theme comprises informants' views in relation to how the program is run, evaluated and modified.
8. Program structure	This is where informants presented their views about the structural aspects of the program.
9. University lecturers	Under this theme, I placed informants' discussions about the teaching backgrounds, knowledge and expertise of the lecturers.

For each theme, I present, in this section, the aspects related to that theme and how participants discussed them. I also present table summaries for each theme to show whether the aspects were Claims, Concerns, or Issues. In those tables, I outline the nature of the statements for each stakeholder group. To this end, I identify the number of participants who made positive statements about the aspect and the number of participants who made negative statements about the aspect. If, within a stakeholder group, all of the statements about an aspect were positive, that aspect was noted as a Claim for that group. If all of the statements were negative, the aspect was noted as a Concern for that stakeholder group. If some statements were positive and others were negative, the aspect was considered as an Issue for that stakeholder group. Whenever participants within a stakeholder group did not discuss an aspect, this was also noted in the table.

5.1.1 Theme 1: Attitudes towards the program

As shown in Table 5.2, science teachers expressed their attitudes towards the program and described the extent to which they enjoyed the program and what fuelled their motivation to complete it.

Table 5.2 – Aspects discussed under "Attitudes towards the program"

Aspect	Positively expressed by:	Negatively expressed by:	Nature of the statements
Enjoyment of the program	2 teachers 0 lecturers	2 teachers 0 lecturers	Issue for teachers Not discussed by lecturers
Motivation to do the program	1 teacher 0 lecturers	2 teachers 0 lecturers	Issue for teachers Not discussed by lecturers

While two teachers (ST3 & ST4) expressed their enjoyment of the program, another two (ST1 & ST2) found its content to be rather repetitive, boring and sometimes irrelevant. One of them explained that the program needed to include “more practical based learning rather than just discussions hour after hour, week after week” (ST2, Interview 1). As to the motivation to complete the program, one teacher (ST4) found the program motivating by itself while 2 others (ST1 & ST2) were only externally motivated and considered the program as a degree they were doing so that they can maintain or get a job. University lecturers did not comment on this theme. One of them (UL4) explained that teachers’ attitudes are a relative matter and that lecturers always strive to make sure that most of their student teachers end up developing favorable attitudes towards the program.

5.1.2 Theme 2: Guidance and support

Table 5.3 summarizes the aspects that were discussed under the theme ‘guidance and support’. The table shows that informants discussed their beliefs about the guidance and support they received from the university personnel during two stages of interacting with the program: prior to enrolling and after enrolling. Aspects discussed under this theme include admission into the program, recognition for prior learning, lecturers' support throughout the units and the supervising teachers' support during the professional experiences.

With regard to the first aspect, admission into the program and recognition of prior skills, two teachers (ST1 & ST4) expressed contradictory feelings about how the university handled their previous experiences, and whether and how it was accounted for. While one teacher (ST4) was satisfied with the way the school of education’s admission protocols aligned with

the department of education requirements, another teacher (ST1) felt that, because of her seven-years of teaching experience, she already knew most of what was covered in the program and her experience should have been taken into account during enrolment so that she does not have to take courses whose content she already knows.

Table 5.3 – Aspects discussed under "Guidance and support"

Aspect	Positively expressed by:	Negatively expressed by:	Nature of the statement
Admission into the program and recognition of prior skills	1 teacher 0 lecturers	1 teacher 0 lecturers	Issue for teachers Not discussed by lecturers
Support in selecting units	0 teachers 0 lecturers	2 teachers 0 lecturers	Concern for teachers Not discussed by lecturers
Lecturers' support throughout the units	2 teachers 3 lecturers	1 teacher 2 lecturers	Issue for teachers Issue for lecturers
Supervising teachers' support during the professional experiences	1 teacher 0 lecturers	1 teacher 0 lecturers	Issue for teachers Not discussed by lecturers

In relation to the second aspect, support in selecting units, two teachers (ST1 & ST5) did not feel well-supported during their selection of units, and felt that they missed some opportunities and even job prospects because they were not aware of the consequences of their selections. One of them stressed the important role that academic advisors could – but do not – play in that respect, particularly that student teachers often “don’t think about the implications of their selections at the time because they just go for survival and choose those units which they know they’re going to achieve well on instead of those they might need further down the track” (ST5, Interview 1).

Concerning the third aspect, lecturers' support throughout the units, two teachers (ST3 & ST4) felt that lecturers provided enough support for them to learn throughout the units and coursework. They also thought lecturers were approachable, caring and generous with their support for the assignments. On the other hand, while not contradicting her peers in that the support during the coursework was generous, one teacher (ST6) believed that the lecturers’ support was minimal and even absent during the practicum and that she didn’t have anyone from the lecturers to turn to during the practicum to receive feedback on her performance. That teacher also explained that it was more important to her to get feedback from the university lecturers about her performance than it was from the supervising teacher.

Three lecturers (UL3, UL4, & UL6) thought they supported student teachers well across the various units and practical experiences. One way through which they provide support for student teachers is by allowing them to choose the topics for the assignments. UL4 (Interview 2) explained that "a lot of the units have the students do lesson plans and/or programming, so they will email me and say look I am teaching this, can I please do my program about this instead of what's in the assignment because I am teaching it in this case and I'd say yes, absolutely!". UL3 (Interview 1) explained the role of the Professional Experience (PREX) office in offering support for the students during their practical experiences: "We send emails to the supervising teacher and we get the response, then we follow up by phone. The same happens with the students, emailing us how they're going. If anybody has problems, we try and send someone out to see them. Now that's not everybody, but certainly if there's a problem we also have eight liaison officers who take trips to most areas of NSW … but we don't get all of them, I'd love to but 1800 students in year, is a bit many".

Two lecturers (UL4 & UL6) explained that, in order to provide scaffold for student teachers during their professional experiences, they had established websites and virtual discussion forums as part of their units. UL4 (Interview 1) explained:

> There is a section on the discussion board which has to deal with discussion and feedback about their PREX, offering mutual support. So, tell us your story, if you've got problems, people will, the student and I will give ideas, give advice … so that will be kind of a virtual little support network within this unit.

Nevertheless, not all lecturers agree that the support they are offering for teachers during their PREX is sufficient. Two lecturers (UL7 & UL8) explained that it was up to the supervising teachers to solely assume that function. Arguing that this was problematic, UL7 (Interview 1) explained that lecturers "don't really know other than [through their] assessment task how [students] are going to manage during their PREX. And the difficulty is that if [students] are not managing well, there is no other opportunity to develop that". UL8 explained that this is further complicated for student teachers who are taking the program off-campus because, in the absence of any connection between the lecturers and the supervising teachers, it is hard to monitor what and how the student teachers are learning during their PREX. Although UL8 acknowledges that the supervising teachers are experts in their domain and can provide meaningful support for students, she explains that better communication between lecturers and supervising teachers will radically enhance the quality of feedback that student teachers will get.

With respect to the fourth aspect, supervising teachers' support, one teacher (ST5) said that the supervising teacher offered meaningful support by allowing her to experiment in the classroom with what she had learned during the program. In contrast, another teacher found the support that she got from her supervising teacher was very minimal and non-constructive and was reduced to "filling in the evaluation forms required by the university" (ST1, interview 1). While university lecturers did not comment thoroughly on this aspect, three (UL3, UL7, & UL8) of them acknowledged that students' experiences in the practicum depended to a large extent on the professionalism, character and style of the supervising teachers.

5.1.3 Theme 3: Delivery mode

This theme includes informants' ideas about the two modes of program delivery: on-campus and off-campus. As shown in Table 5.4, the aspects discussed include the off-campus learning experience and the flexibility of the off-campus program.

Table 5.4 – Aspects discussed under "Delivery mode"

Aspect	Positively expressed by:	Negatively expressed by:	Nature of the statement
The off-campus learning experience	1 teacher 1 lecturer	2 teachers 1 lecturer	Issue for teachers Issue for lecturers
Flexibility of the off-campus program	2 teachers 1 lecturer	0 teachers 0 lecturers	Claim for teachers Claim for lecturers

Concerning the first aspect, the off-campus learning experience, ST4 explained that the online forums were useful venues for sharing opinions and experiences and for providing mutual support. UL6 agreed with this perspective, arguing that the discussion boards and forums, which are considered as the main classroom space in the online environment, present venues for interaction between student teachers and provide experiences that are similar to what happens in real settings. ST1 disagreed and thought that the forums were mostly populated with irrelevant posts and, therefore, were not as good as the face-to-face learning set-ups. Nevertheless, in replying to this concern during the second interview round, UL6 argued:

> I can certainly see there's a lot of irrelevance [on those forums] … but it's interesting. There's a lot of social stuff that's happening in the discussion boards which some students think that they're just wasting time because they're just chattering. But evidence suggests that that's sort of the real, the social stuff is sort of

> the glue that holds the community together. So often in that social stuff you've got a lot of good embedded knowledge plus it also helps students build up rapport and relationships … this is interesting because I also noticed that one of the other issues they are talking about is that, in off-campus education, students often felt quite isolated. (UL6, Interview 2)

ST4 echoed UL6's statement and argued that:

> [Participating in the online forum was] no different than sitting in a tutorial; you know when they've got eight students sitting in a tutorial, there are going to be people making irrelevant comments. There are going to be people making inappropriate comments. As adults, I think we just need to use our own discretion as to what we take on-board; but in terms of the content, I found it extremely relevant and I posted a few questions, not that many. I did feel like I had the ability to respond to some other people's questions because of my experience in teaching and I felt that I was actually quite helpful … and [one lecturer] said at one stage that my responses were helping others and that made me feel quite empowered. (ST4, Interview 2)

Arguing from a different vantage point, UL1 explained that the online learning environment is still not as beneficial as the real life face-to-face environment, and that this is concerning, particularly that the technology is not there yet that would make the social experiences and interactions of off-campus student teachers as rich as those of the on-campus student teachers:

> I don't think the technology is actually there at that point where you can say that the experience is as valuable and as rich as it would be when they're on-campus because they're still learning alone. So there's a lot to be said for Vygotsky's social constructivism. I mean that [theory] underlies what we teach here about the way students learn at school but we don't really incorporate that very much into our online components because many students are learning in isolation. I think it's a better experience, more meaningful, if you're learning with others at the same time, you know, the social experience. (UL1 Interview 1)

UL1 further explained:

> Our external students have a sense of isolation from other students and also from the course coordinator, then they're not really sure whether they're doing things the right way. They lack a little bit of guidance simply because they can't turn to one person on the other side and say "am I going OK here? Am I keeping up? Am I approaching the assignment correctly?" they can write on the discussion board or write or email me etc. or ring me and get information but it's not really the same. (UL1 Interview 1)

ST3, who experienced both off-campus and on-campus education, agreed with UL1 and explained that:

> Having a course at the UNE with tutorials with lecturer twice a week or whatever happens to be was a huge boost rather than being an external student having no real contact with anyone else in the cohort. With my fellow colleagues who were in the

> course with me, we build on tons of knowledge and helped each other out … You get the online forms and things, but nothing is as good as sitting in a classroom and asking ideas and having a decent discussion about things and sitting in a lobby and talking about assignments or talking about all of the other stuff. (ST3 Interview 1)

For the second aspect, flexibility of the off-campus program, two teachers (ST1 & ST4) thought that off-campus education gave them greater flexibility to study at their own pace when compared with on-campus education and that this presented a great advantage to off-campus education. ST1 explained that taking off-campus studies allowed her to work and gain more experience at the same time. UL6 added that the use of an asynchronous medium of communication not only offers greater flexibility, but also that is also gives student teachers plenty of time to think about things in more depth. This, UL6 (Interview 2) argued, is a great advantage when compared to one-to-one interaction where "students are rushed to think about things and answer questions on the spot because the they are working within a one- or two-hour timeframe".

5.1.4 Theme 4: Content of learning

This theme includes informants' ideas about the content covered in the program as well as the extent and quality of coverage. Table 5.5 lists the various aspects that fall under this theme and which include the extent of coherence of the content across the different units in the program, appropriateness of the content to secondary science teaching, amount and quality of resources made available for teachers, relatedness of content to HSC requirements and so forth.

Concerning the first aspect, developing teachers' Pedagogical Content Knowledge (PCK), teachers' arguments were formulated in relation to how much the program focused on developing their content specific teaching resources and repertoire. Three teachers (ST2, ST5 & ST6) were concerned that the program did not provide them with practical experiments and resources in science that could be readily implemented in class:

> The resources that I found were lacking were those specific to science teaching in terms of practical activities specific for science lessons; so, practical experiments that will be used for demonstrating/developing the skills of students … carrying out practical investigations, giving ideas of what sort of practical investigations specific to the content, to give out or to use with students. I didn't feel these were given in the program. Another example may be even a simple demonstration, say I might have wanted to show a quick exciting demonstration where I might have wanted to, maybe, you know, make something explode or make a bit of noise or something that would go colourful just for pure engagement for pure fun, a very simple and exciting

chemical reaction specific to the content I was never shown or given an idea of what I was only able to get. (ST2 Interview 1)

Table 5.5 – Aspects discussed under "Content of learning"

Aspect	Positively expressed by:	Negatively expressed by:	Nature of the statement
Developing teachers' PCK	4 lecturers 1 teacher	1 lecturer 3 teachers	Issue for teachers Issue for lecturers
Relevance and usefulness of content learned	2 lecturers 2 teachers	1 lecturer 4 teachers	Issue for teachers Issue for lecturers
Learning the content of the HSC syllabuses	2 teachers 0 lecturers	2 teachers 0 lecturers	Issue for teachers Not discussed by lecturers
Writing programs and lesson plans	2 teachers 1 lecturer	0 teachers 0 lecturers	Claim for teachers Claim for lecturers
Duplication of content	0 teachers 0 lecturers	1 teacher 5 lecturers	Concern for teachers Concern for lecturers
Appropriateness of content for secondary science teaching	0 teachers 0 lecturers	3 teachers 0 lecturers	Concern for teachers Not discussed by lecturers
Learning to deal with stakeholders (e.g. parents and colleagues)	0 teachers 0 lecturers	2 teachers 0 lecturers	Concern for teachers Not discussed by lecturers
Learning how to do an accreditation portfolio	0 teachers 0 lecturers	2 teachers 0 lecturers	Concern for teachers Not discussed by lecturers

ST6 added that the experiments and practicals, among other resources, should be trialed and tested in order to be meaningful for teachers, particularly those in their first year of teaching. She explained that having these resources would boost teachers' performance, as they tend to be reluctant to experiment with resources in the first couple of years of teaching.

ST5 also argued that student teachers need to have opportunities during the program to perform some of the most basic experiments and get feedback on them so as to learn how to teach it and also to avoid accidents. ST5 argued that while these resources could be obtained from the Internet, teachers need to have an opportunity to trial them during the program and get professional feedback from lecturers about how to administer them in a classroom. She explained that lecturers could also point out potential risks associated with practical activities:

> the added value of doing it in the university context is so that you can experience what is likely to go wrong, the things that can go wrong when you do those experiments so things breaking. You know the first time I did crystallization, I think, thankfully, it was a senior class and we had safety equipment on but I blew up and that's pretty basic! Because I just didn't know that they would expand with the salt that was lifting them and explode and so it can be really dangerous and here with a junior class they're really silly so it's just actual experience with the things that can be right and wrong. (ST5 Interview 2)

Furthermore, ST5 argued that lecturers can help teachers develop a repertoire of resources that match the main points in the syllabuses so that teachers can have a repertoire to draw on when teaching the content:

> The list of experiments on the internet is endless and occasionally I will think of an experiment that I know and I will seek it out on the internet just to refresh, you know, what equipment did I need and so forth, but I know what experiment I am looking for. If you don't know what you're looking for it's going to be really hard to find. (ST5 Interview 2)

In contrast to ST2 and ST5, ST6 and ST4, who has more experience in teaching, found that some units, particularly those concerned with science teaching, were very beneficial as they provided ample opportunities for developing content-specific teaching resources.

While teachers' arguments about PCK centered on the development of content specific teaching resources, the lecturers' arguments revolved around the extent to which the program and its content are practice oriented. Three lecturers (UL1, UL4 & UL8) thought that the course was practically oriented, bearing a strong emphasis on developing the teachers' PCK. UL1 explained:

> The course has a practical basis that is aimed at equipping graduates with the practical skills and knowledge that they will need in their first year, first couple of years of teaching … and by practical, I mean safety and risk assessment skills that they absolutely need, knowledge of the departmental policies relating to these sorts of things. (UL1, Interview 1)

Another lecturer argued that the workshops they offer during the program present ample opportunities for student teachers to develop their PCK: "In every workshop, we give them practical examples of different teaching and learning strategies [and] we then ask them to see how they could relate that to their subject areas" (UL8, Interview 2). UL4 also stressed that lecturers emphasize lesson planning, and that every lesson that teachers plan can be taken away and used. She explained that: "sometimes, teachers share their lesson plans and post it up on the discussion board and then they get different lesson plans that they can use" (UL4, Interview 2). Nevertheless, UL4 agrees with the teachers that providing more resources to

teachers will expand their science-teaching repertoire and will eventually improve the program.

Concerning the second aspect, the relevance and usefulness of the content learned, two teachers (ST3 & ST4) noted that the content they learned in the program was mostly relevant to and aligned with what was going on in school. Their experiences in schools reinforced what they learned at university. ST4 (Interview 2) explains:

> I think [the program] was fantastic in terms of its relevance and its proximity to reality. And I think that comes from the staff more than anything and the staff programming … the way that they have done it or the way the program has been setup and it's been done with teaching in mind with the result of getting at the other end a good teacher.

In contrast, four teachers (ST1, ST2, ST5 & ST6) said that they found several elements of the program to be irrelevant to their teaching. They argued that these elements were either theoretical in nature or not applicable in the context of school teaching (like the theories of teaching and learning (ST1 & ST2) or classroom behavior management (ST6 & ST2)) or they were too elaborate to the extent that they became irrelevant (like Aboriginal education (ST5 & ST6) and literacy education (ST6)):

> In science teaching, where we did a lot of learning how to write programs, they do all the stuff in theory. We need to be able to put that in perspective. So, like, they could give us or should give us, say, examples of practicals that … have been tried and tested since that was the starting point … whereas we don't get that … like if they tell us how to do things in theory and … in theory and in practice are two different things. (ST6, Interview 1)

In trying to explain the reason for this perceived disconnect between the content learned and its application in schools, ST5 (Interview 2) explained:

> Perhaps if the different philosophies of teaching and learning that they look at, if they're not put into a practical analogy, like they don't actually show what they look like in the classroom then they won't really have any relevance or make any sense. So it would have been much better if lecturers were able to model what these theories look like by giving a lesson that is teacher directed or giving a lesson that is student directed or, you know, giving a lesson that is coming from the perspective of whoever they're trying to model.

In relation to the lecturers' positions on this aspect, UL4 and UL7 said that the program provides useful content that equips student teachers with tools and resources that are not only relevant but also necessary to their work. UL7 (Interview 1) explained that the program is entirely "based on literature and experience, statistics and not sort of 'this is what I think'

because no one cares about what we think, it's got to be what's actually going on in the real world". UL4 (Interview 2) also explained that: "there isn't a big emphasis on writing theoretical essays that are sort of disconnected from the application part of it". She added that: "If you're looking and thinking about it from Bloom's taxonomy there's a lot of emphasis on the top of the pyramid on the evaluation, synthesis, and application of the ideas that they're learning in the program" (UL4, Interview 1). Nevertheless, UL4 acknowledged that the perceived disconnect expressed by teachers could have resulted from teachers not being able to implement their lesson plans and programs in a classroom during the program and that they had to wait until they were actually teaching in order to do that. She elaborated her point by saying:

> I know a unit that I used to be involved in – I don't know if it's running anymore – that was specifically designed to address that issue. The unit was taken by on-campus students before they went to their first prac. It involved lesson plans where it was compulsory for them to take it into the classroom and teach. So that they then got feedback on how that process worked. (UL4, Interview 2)

Regarding the third aspect, learning the content of the Higher School Certificate (HSC) syllabuses, two teachers (ST3 & ST4) thought that the program covered the content of the HSC in a thorough way through "lesson planning and preparation and the programming and knowing the programs back to front" (ST4, Interview 1). Two other teachers (ST1 & ST5) disagreed with this and stated that the content of the HSC syllabuses was not covered properly throughout the program. ST1 explained that the assignments only covered a relatively small portion of the syllabuses and ST5 explained that the program covered nothing about open-ended investigations despite their great importance in secondary science teaching.

Concerning the fourth aspect, writing programs and lesson plans, two teachers (ST1 & ST4) found that the program taught them well how to write programs and develop lesson plans. ST4 (Interview 2) explained: "I can walk into any school now knowing that I have got this 10 week program, particularly units in Biology I have got two of them that I did there and I have got the foundation to do it again". One lecturer (UL4) agreed with them and argued that the program provides plenty of opportunities for teachers to develop programs and lesson plans that they can be used in classrooms.

Concerning the fifth aspect, duplication of content, one teacher (ST6) thought there were redundancies in the program across the various units: "I'll say also a lot of the subjects are really … they sort of cover the same thing in multiple times" (Interview 1).

Similarly, five lecturers (UL1, UL2, UL3, UL5 & UL8) expressed concern that they did not know how their units integrate with other units in the program and that, as a result, there was unnecessary duplication of material. UL1 (Interview 1) explained that some student teachers experience redundant materials because of historical changes to the structure of the program:

> students sometimes take a long time to get through a course and they may have enrolled quite a while ago, when the course was different, and then they still got to work their way through the course including your unit, so they're not all following the same pipeline through, and so sometimes it's hard to know whether a student has done X before they've done Y as they should've been or the wrong way around.

UL3 (Interview 1) explained that redundancies exist because lecturers do not have a detailed map of the program, and while they might all know what units exist in the program, they do not necessarily know what is being covered in those units:

> I think we duplicate materials sometimes, and sometimes we miss things out and nobody knows because we don't have an overall course map, content map. So we have a program structure, like I know that GDE has EDLT 400 for learning and teaching, those who do science they've got four science curriculum units. But I don't know what they do in them, but I know they've got them. It would be really nice to know, do they teach them how to write them a science lesson plan or do I have to do it? And if they teach them, do they include quality teaching? Because all the NSW government schools use quality teaching framework, are they putting it on their lesson plans? Because eventually, they're going to have to, and who is teaching them how to program? Do the coordinators do that or don't they?

Concerning the sixth aspect, the appropriateness of content for secondary science teaching, three teachers (ST2, ST3, & ST6) thought that some units were more tailor-made towards the primary education student teachers and that several things that were discussed in these units do not apply to secondary science teachers. ST6 (Interview 1) explained:

> I found that a lot of it is based at primary school level. So, actually they're a lot different from high school teaching. Primary teachers can do a lot of things that we cannot do because they've got the kids for the whole day. You start the morning with these things … if this student plays out you do this with them and that sort of stuff. We can't really put a lot of those skills into practice because I might have seen them today and then I'm not going to see them again until, say, Friday, or in some cases I'm not going to see them again until next week.

ST2 and ST3 expressed another related concern and said that the program did not prepare them to teach junior science, as they had no idea as to what was covered in those curricula. Those teachers emphasized the importance of this missing component particularly that secondary science teachers are typically expected to teach junior science. None of the university lecturers discussed this particular aspect.

With regard to the seventh aspect, learning to deal with stakeholders (e.g. parents and colleagues), two teachers (ST1 & ST2) argued that despite the importance associated with those skills, they were not taught how to negotiate challenges with peers and discuss students' performance with parents. ST1 explained that for most schools, dealing with parents is a routine and central activity and teachers should be taught how to use professional communication methods to inform parents about their children's progress or problems. ST2 also explained that the program should include a component to help teachers develop the necessary knowledge and skills to negotiate with peers. He argued, "all the political garbage that goes on in the school is what would eventually drive [him] out of teaching" (ST2, Interview 2). The university lecturers did not discuss this aspect.

Concerning the eighth and final aspect, learning how to construct an accreditation portfolio, two teachers (ST2 & ST3) believed that the accreditation process was not properly addressed in the program although it is considered as a big step in their professional work. Again, none of the university lecturers discussed this aspect.

5.1.5 Theme 5: Residential school

Under this theme, I grouped the informants' arguments for and against having a mandatory residential school in the program (see Table 5.6). The residential school is a "mode of teaching and learning that brings students together at a particular location for a period of intensive interactive learning experiences" (UNE, 2013). Residential schools are particularly designed for off-campus student teachers to come to the university and learn some of the practical aspects of science teaching.

Table 5.6 – Aspects discussed under "Residential school"

Aspect	Positively expressed by:	Negatively expressed by:	Nature of the statement
Arguments for having a mandatory residential school	0 teachers 2 lecturers	0 teachers 5 lecturers	Not discussed by teachers Issue for lecturers

Two lecturers (UL1 & UL8) argued that there should be a mandatory residential school for science teachers as it enhances external students' practical experiences, particularly because the course, they said, was a bit lacking in terms of teaching student teachers about safety, coordinating experiments and things of the sort.

UL1 referred to the positive feedback received from student teachers to argue that the residential schools should be mandatory:

> The students, again from their evaluations, always said that it was a very worthwhile experience and they were glad they came and that it was the most positive aspect of their course, etc ... And we also felt better about it because we knew that ... I mean they're better equipped to go into the classroom because they had these hands-on stuff to do with safety or coordinating an experiment or something like that. (UL1, Interview 1)

He further added that the benefit of having a residential school is for off-campus student teachers to experience the same sort of practical hands-on opportunities that the on-campus student teachers get and to offer them opportunities for collecting meaningful science resources and developing their PCK:

> The residential school did definitely contribute to the building up of resources and ideas for the students because one of the features of it was something called the science circus and this was where each student is asked to find a small experiment or demonstration that's suitable to junior secondary school that they find on the web or in the textbook or something like that and then to prepare that and then do a presentation when they come up here] Now I provide a lot of online links in my units but that's not really the same because learning is social constructivism, and these intensive schools were social constructivism in action. When a student is learning in isolation that's a very different thing. When they're in a group then they really bring or develop an enthusiasm for the work and for new ideas and get energy and there's a lot more input and collaboration and so after the session students would still be talking and sharing ideas and things like that. (UL1, Interview 1)

UL1 further added that the residential schools are especially beneficial for science teachers since they will be teaching science for years 7-10:

> Our students, I know they got science degrees but they usually got science degrees in one major or maybe 2 major areas, but when you're teaching years 7 to 10, you have to know a whole range of integrated disciplines in science. They don't have that and it is hard to help them develop that. (UL1, Interview 1)

UL8 agreed with UL1 about the importance of having a mandatory residential school for external science teachers for the same reasons advocated by UL1. However, she did acknowledge the costs and burdens incumbent on the students:

> I don't think [the online environment as opposed to the residential school] offers the same thing but I think there are a lot of students who haven't got a lot of money, it's difficult to come and afford it on top of all the other classes. (UL8, Interview 2)

While UL1 and UL8 argued from a pedagogical perspective for keeping the residential school mandatory, five other lecturers (UL2, UL3, UL4, UL7 & UL9) took a rather market-

driven standpoint and argued that residential schools should no longer remain mandatory. UL2 (Interview 1) explained that point of view:

> Students were unhappy with coming here, because it would cost them money, they had to pay for the airfare, pay for accommodation, and so on … the students would always give great feedback, but it was also seen as reducing enrolments! You would get a lot of phone calls from students saying "I see there's a mandatory residential school, do I have to come?" and you would say "yes" and they would say "Oh right" and they wouldn't enrol … You'll find that pressure to say "if we want students coming here, there's a certain proportion who wouldn't enrol if we had intensive schools but we want them so don't have intensive schools.

UL5 (Interview 2) echoed this position, explaining that student teachers are using residential schools as a filter when they're looking at universities in the first instance.

Three lecturers (UL4, UL7 & UL9) acknowledged the importance of the residential school but suggested having a virtual alternative that would eliminate the financial burdens often associated with residential schools. In that respect, UL9 (Interview 2) argued that:

> With the residential schools and face-to-face teaching in general I think it's a terrific idea. In fact that's a nice way to develop a relationship with students and teachers is actually face-to-face but having said that that's quite a traditional view as a world of teaching in that there are other ways of doing it and we can develop that off-campus … so because a lot of our decisions are market driven now, we have to sort of respond to students who cannot afford to or don't have the time to travel here to teach. I think that encourages us to think outside the box to develop that relationship online or through a whole range of ways of doing that and I think that's possible.

5.1.6 Theme 6: Teaching strategies in the program

This theme includes the informants' views about the teaching strategies adopted by the various lecturers in the program. As can be seen in Table 5.7, the aspects discussed here include the nature and quality of instruction during the units and during the practicum as well as the nature and usefulness of the assignments and assessment strategies used in the program.

Concerning the first aspect, the quality of instruction in the units, only one teacher (ST4) thought that some of the online modules were really well laid out and had useful tools for instruction (like podcasts and short videos). Unlike ST4, five teachers (ST1, ST2, ST3, ST5 & ST6) complained about the teaching strategies employed in the program. For example, two of the teachers (ST2 & ST5) who took the course on-campus, thought that lecturers were not addressing the multiple learning styles of student teachers and were adhering to a single mode of delivery, lecturing:

> I also didn't like that there was a lot of discussions at the science education seminars or tutorials, I'm more a person who likes to be shown things or shown videos or I learn by practical sessions rather than just listening to someone talk and try to take notes at the same time … (ST2, Interview 1)

Table 5.7 – Aspects discussed under "Teaching strategies in the program"

Aspect	Positively expressed by:	Negatively expressed by:	Nature of the statement
Quality of instruction in the units	1 teacher 0 lecturers	5 teachers 0 lecturers	Issue for teachers Not discussed by lecturers
Quality of instruction in the professional experiences	3 teachers 0 lecturers	2 teachers 2 lecturers	Issue for teachers Concern for lecturers
Nature and usefulness of assignments	1 teacher 5 lecturers	3 teachers 0 lecturers	Issue for teachers Claim for lecturers

Similarly, the three other teachers (ST1, ST3, & ST6) who took the course off-campus expressed concerns about the teaching style which they thought was mostly based on passing on information through the website and discussing it without a major focus on practice.

> You go to Moodle and it's all up there! You just do the readings, write an essay, hand it in, do the readings, write the assignment, hand it in ... That's about the extent of it. (ST1, Interview 1)

ST3 (Interview 1) went on to argue that: "a lot of the discussion boards are based on how to do the assignments and what [students] need to do in order to pass the assignments". He argued that these discussions need to include more discussions about the importance of the theories they are learning, why these are important from a pedagogical point of view, and how these can be actually put into practice" (ST3, Interview 1). Furthermore, two teachers (ST1, ST6) were concerned that lecturers were not modeling the theories they were teaching about:

> I found that a few [lecturers] tell you how to do something like behavioural management, they would say, this is how you do it, and they do something completely different. So, it was basically, do as I say, not what I do. If you are trying to do something, if you're teaching teachers how to do it … you should really be doing it yourself … if you're teaching behaviour management or whatever you are teaching, you still need to be able to put that stuff into practice while you're trying to teach it to university students. I found a lot of them didn't do that. (ST6, Interview 1)

Concerning the second aspect, the quality of instruction in the professional experiences, three teachers (ST4, ST5, & ST6) explained that they learned a great deal from their supervising

teachers and that their interactions with those teachers were beneficial and constructive. ST4 (Interview 1) explained: "In terms of practical ideas, demonstrations, engaging activities, and so on, I feel more confident learning what I learned of a teacher who has been teaching for 35 or more years". ST5 (Interview 2) clarified that the learning he got from the supervising teachers transcended the skills and knowledge acquired at university and covered "how the school runs, how to build rapport and trust with the students, and what's professional interaction compared to non-professional interaction and that sort of things".

Nonetheless, one teacher (ST2) complained that the practicum was lacking one particular component, reflection about teaching. He explained that this reflective activity should be an integral part of the program when the teachers are having their PREX and that it should be coordinated by the university lecturers who are qualified to give a pedagogical perspective on the challenges that teachers face in schools:

> [S]itting in a classroom with a lecturer that no doubt has knowledge of hostile classroom is just not as effective for your own development until you are in a classroom and having your own experience. You need to do it and reflect on that, go into a practical situation and then reflect on it. It's just … how it works, four weeks or eight weeks; these are several teaching periods but these are not enough! It's all about reflecting and negotiating with your lecturer as well as with your supervising teacher" (ST2, Interview 2)

ST3 also explained that it would have been preferable to be visited during the practicum by one of the lecturers he was familiar with in order to get better instruction and constructive feedback:

> During my first PREX, I was not visited at all and during the other two, I was visited for maybe an hour or so and then that was it, an hour-and-a-half maybe. That was very, very quick and very brief. It wasn't from the person that I have had any contact with. It was from the university, but I had no idea who that person was. They just had the university badge on but it didn't have any connection with me. It would have been great to have one of my science teacher or someone in my field come to see me, someone I was familiar with in order to receive adequate instruction and meaningful feedback. (ST3, Interview 1)

Two lecturers (UL1 & UL2) discussed this aspect and agreed with the teachers' position that the role of the lecturers during PREX was minimal and could be improved.

UL1 explained that greater communication between the lecturers and the PREX office and better resourcing could improve the experiences of teachers during their PREX:

> There is that this jump between the supervisors at the school and the lecturers here. We don't connect with them. All of that PREX is organized through the PREX office, so

> some lecturers don't have that contact with the schools … that's not to say it's a bad idea but this is just the way that it is. So it's almost like the students are taken out of our hands for those four weeks and I think a lot of lecturers don't really know what goes on in the schools and with the PREX. So a good suggestion for improving that articulation might be greater communication between lecturers and the PREX office. You know PREX is a really important thing and yet it seems to be one of those areas of the school that does not get the resourcing that it needs for supervisors. (UL1, Interview 2)

In terms of the resources available to provide adequate instruction for student teachers during their PREX, UL4 (Interview 2) further explained: "I don't quite know what the PREX office resources are, but I know they got a huge number of student teachers and three or four people there. So it can't be, no matter how hard they work, it can't be enough".

Concerning the third aspect, the nature and usefulness of the assignments, one teacher (ST4) found that the assessment tasks were very practical and could be easily used in classrooms. In contrast, another teacher (ST1) found that the dominant type of assignments in non-science teaching units is the essay type, which is not practice oriented and not useful:

> You take a course on behavioural management it's really 6 different theories on behaviour management, write an essay comparing and contrasting them ... like so what? Big deal! Like am I actually going to implement that behaviour management plan and theory into my classroom? The reality is that this theory looks great on paper but it does not ... it's not helpful! It's completely unhelpful ... so most of these assignments are unhelpful!! Having the few exceptions being some of the science stuff that I've done, but beyond that, the rest of the stuff is worthless. (ST1, Interview 1)

Additionally, two teachers (ST2 & ST3) considered many assignments to be not very useful either because they were too theoretical in nature,

> [T]here were quite some theoretically-based assignments which now I can't remember too many of them, which is a sign at how irrelevant they were, whereas I can remember some other ones which I think might have been a bit better which were meaningful and which helped me gain some understanding of some of the skills involved. (ST3, Interview 2)

or unreliable as they were not trialed and tested:

> [T]he lessons plans and curriculum plans and similar assignments that I had at university, I wasn't sure whether it would work or not, yes you get feedback on the assignment, but I knew that the programs that were being used in schools, the assessment tasks that were being used in schools, they were working as they were. (ST2, Interview 1)

Five lecturers (UL1, UL4, UL6, UL7 & UL8) argued that the assignments were meaningful, practically oriented and aimed at connecting theory to practice. UL1 (Interview 1) explained

that: “the assignments are fairly practical and combine students' readings and theory with classroom practice with building out resources for their own teaching and skills development”.

Two lecturers (UL4 & UL6) argued that the assignments are “the main tools through which lecturers are engaging the students with the materials” (UL4, Interview 2). UL6 explained how the use of the “learning through assessment” model motivates student teachers to learn meaningfully:

> We’re using a model called “learning through assessment”. I think the traditional thing is you learn some content, you do the activities, and then you go off and then you do the assessment and that's relation between what you're doing here and then what you're assessing. But what we do here is we actually get them working on their assessment tasks and in completing the assessment task they're actually learning the course content. So instead of you learn this body of knowledge and then you might have to be able to assess, ideally you need to assess every outcome, but you might not necessarily do that. In our case, everything they do goes towards an assessment and the students got a motivation because they've got an assessment task to complete and they've got a motivation to actually learn their content. (UL6, Interview 2)

One advantage that lecturers mentioned was that the assignments were made flexible enough to allow student teachers to choose the most relevant context (in terms of science discipline) and to develop the assignments from within these contexts:

> So what we do is we make the assessment very open. So we get the science people come along and they will deliver in their science context … so we're very aware that students that come in with, they’ve already got a certain context in content knowledge so we keep the assignments as broad as possible. (UL4, Interview 1)

Clarifying their positions in response to the concern expressed by student teachers that essays are dominant in the assignments, UL7 (Interview 1) explained:

> We try to get away from the theory side of stuff because we’d really be defeating the purpose. However, some of the AITSL (Australian Institute for Teaching and School Leadership) requirements, they have not only a certain body of content for example you know in terms of how to use the technology safely or ethically but basically also how to use the technology to promote good learning outcomes. So in order to do that we have to teach them how to use the technology, I mean it’s not necessarily easy! In some respects, it’s easier to sort of send students off and do the 2000 word essay.

UL8 (Interview 2) further argues that in some places, the use of the essays in the assignments is justified by the need for student teachers to learn the pedagogical explanations underlying their teaching practices:

> In our units what we try and do is that we try and develop authentic assessment, things that they would be doing out in the workplace. So for example, they might

have to develop an explanation about something within the key learning area; this is very practical but also is based on theory because it talks about all the different points or aspects or characteristics you should have in an explanation.

5.1.7 Theme 7: Program monitoring processes

This theme comprises three aspects that relate to how the program is run, evaluated and modified. As outlined in Table 5.8, the first aspect tackled the amount of communication and collaboration that goes into these monitoring processes. The second aspect included informants' views about some of the processes inherent to the program. The last aspect relates to the evaluation of the program.

Table 5.8 – Aspects discussed under "Program monitoring processes"

Aspect	Positively expressed by:	Negatively expressed by:	Nature of the statement
Collaboration between lecturers	0 teachers 0 lecturers	0 teachers 5 lecturers	Not discussed by teachers Concern for lecturers
Processes related to the administration of the program			
– Admitting student teachers into the program	0 teachers 1 lecturer	0 teachers 3 lecturers	Not discussed by teachers Issue for lecturers
– Assigning workloads to lecturers	0 teachers 0 lecturers	0 teachers 4 lecturers	Not discussed by teachers Concern for lecturers
– Trimesterization	0 teachers 1 lecturer	0 teachers 4 lecturers	Not discussed by teachers Issue for lecturers
– Meeting the academic board requirements	0 teachers 1 lecturer	0 teachers 3 lecturers	Not discussed by teachers Issue for lecturers
– Accreditation	0 teachers 0 lecturers	0 teachers 2 lecturers	Not discussed by teachers Concern for lecturers
Program evaluation	0 teachers 2 lecturers	0 teachers 4 lecturers	Not discussed by teachers Issue for lecturers

With respect to this theme, none of the teachers made any comment about any of three aspects. This is probably the case because these processes are not tangible to the teachers and are only visible to the lecturers who have the insiders' perspective.

With respect to the first theme, collaboration between lecturers, five lecturers (UL1, UL3, UL4, UL6 & UL8) were concerned about the small amount of collaboration that goes into the

program design, delivery and monitoring. UL1 (Interview 1) expressed: “It's a bit of a silo mentality in that lecturers, even within the same program like the GradDipEd, don't really have much to do with each other”.

While two lecturers (UL1 & UL4) argued that the communication is great at the level of discipline specific teams (like the science education team or the social science education team), UL4 suggested that collaboration should be enhanced at the program level.

Three lecturers (UL3, UL6 & UL8) discussed the challenges resulting from the lack of communication between lecturers. UL8 (Interview 1) argued that lecturers feel as if they lost ownership of the program because they have to implement decisions that are handed for them in a top-down fashion:

> The trouble is everything has been set in place for new courses and all of those things, so I have to try and make sure I can work within it, and make it work which is always interesting that you make somebody else's ideas work that you've never actually agreed with in the first place.

UL6 (Interview 2) argued how the lack of collaboration at the program level could cause redundancy in the program and decrease coherence across the units. UL6 gives an example of how more collaboration could reduce unnecessary workloads of student teachers:

> In the GradDipEd, because you’ve got so many different teams, I think there possibly is a lot of overlap. There is duplication. Our teams could sit down and actually improve that; I mean for example the PDHP team, the PE for want of a better word, they had a lesson plan pro forma. You know that they were getting their students to design lessons using this lesson plan framework. So I used it, I said you know that students already know how to use it, they’re happy with it. I’m not going to get them to do something different so I was able to borrow stuff from them. It was consistent and suddenly the students weren’t focusing about getting the things in the right boxes because they’d seen the framework or they’d seen their little template before.

UL3 (Interview 1) argued that the lack of communication between the lecturers and the PREX officers has impacted on the preparedness of teachers before taking their PREX:

> Since science isn't my area, so it's very much dependent, I think on the subject unit coordinators or lecturers and myself being able to work more closely together than we do because I don't know what they do get taught, and they probably don't know what I’m doing … The science lecturers wouldn't know what preparation I’m doing very often, and I don't know if they are sending [teachers] prepared.

In light of these challenges, UL8 (Interview 2) argued that there need to be in place some structure to enhance communication and discussions about program-related issues:

> I feel that people are so busy and there are no avenues for [collaboration] anymore. The school meetings cover many things, role awards, research reports, teaching reports, so I don't think discussions about award-specific issues happens there. I talked about the forum for the GradDipEd in the past interview. I don't think this happens anymore. I think communication and collaboration has got far worse over the last few years. When we had those forums at least everyone could go and listen and give their two cents worth and know what was happening. I think if people are collaborating then I don't know about it.

With regard to the second aspect, the processes related to the administration of the program, the lecturers discussed five processes.

1. The first was related to the way student teachers are admitted into the program. In that respect, three lecturers (UL5, UL6 & UL7) found that the admission requirements into the course were not appropriate. For example, UL5 and UL6 argued that applicants who have been away from university for many years should not be admitted into the course only based on their academic achievement:

> Someone could have undertaken their undergraduate degree 20-25 years ago and not have worked in the field and have forgotten, most of them may well have got distinctions and high distinctions in their undergraduate degree, but if they haven't used something for 25 years they would have forgotten a lot of it. So I guess that would be my major concern with the admission requirements and that affects people in there. (UL5, Interview 1)

UL7 (Interview 1) argued that it was inappropriate that there were no interviews to filter the quality of teachers admitted into the course. She explained:

> When we fill out an application form, we all look great, but teaching is more than that. Maybe there are some professions, dare I say, like Law or Medicine where you just need to be good at that. You don't actually need to, dare I say, have a personality and to be able to relate and understand and read and respond to people, that is what teaching is all about.

Nevertheless, UL4 counter argued from a market-driven perspective and explained that, if interviews were put in place, the enrolment rates of student teachers could drop dramatically and that, with the current levels of funding allocated to the school of education, the school would be facing a potential economic catastrophe.

2. The second process, related to the administration of the program concerned the allocation of workloads to lecturers. Four lecturers (UL1, UL3, UL4 & UL8) argued that the workload assigned to them has imposed constraints on their performance as they have to balance teaching, research and professional development in addition to other duties (like unit coordination and attending meetings). UL1 (Interview 1) explained:

> I don't think that teaching into these courses is really ideal because all the lecturers have other responsibilities. If teaching was our only job, I reckon the quality of teaching of units and courses would be much higher than it is ... but you also have a lot of pressure to publish and to do research and to do service to the university.

3. The third process related to the administration of the program was trimesterization.

One lecturer (UL5) believed that this process was beneficial in terms of reducing the time student teachers spend in their degree. She argued that: “a part-time student who is reducing say six years of study to five years, it's a very significant change; it's an extra year of income” (UL5, Interview 2). Alternatively, four lecturers (UL3, UL4, UL6 & UL8) argued that the move towards trimesterization was concerning. They believed there was no sound pedagogical argument for the move and that the process was impacting badly on teachers and lecturers alike. For example, UL3 (Interview 1) explained that for teachers going on PREX, it was harder for them to prepare for their PREX:

> Trimesters and school holidays are now very difficult; semesters, it used to fit nicely: two school terms and two university terms, it doesn't anymore. We have students now from the program, out on PREX in schools, and while the assessment tasks about safety in the lab used to be usually done before they go, now they're only here for two weeks and then they are out in schools, so while they can experience it first-hand, they haven't had the preparation as much as they might have done in the past because of the change”

For UL4, the time assigned for marking was considerably reduced because of trimesterization. This was particularly challenging for those lecturers who coordinate units that attract a large number of students. For UL6 and UL8, trimesterization made it more difficult for them to deliver the same content in a shorter period of time without having to make compromises in the content and focus of the units. UL6 (Interview 1) clarified:

> With trimesterization the time has essentially gone from what would have been a 15-week unit to 12 weeks, but they're all still meant to be delivering 150 hours on instruction. So we still have to give 150 hours of instruction, but now we've only got 12 weeks to do it, so what sounds a lot of students will struggle with the workload finding time to do the work.

4. The fourth process related to the administration of the program was meeting the academic board requirements. While UL4 argued that the requirements set by the academic board at the university are flexible enough and give the lecturers a large degree of freedom in modifying the units while they are being taught, three other lecturers (UL1, UL2 & UL6) believed that the academic board sets very rigid constraints on the lecturers in terms of

what they can modify. They said that it is concerning that changes to the unit had to wait for twelve months before they take place which can mean up to two iterations of the unit before changes are in place. UL6 explained:

> [T]he difficulty I find here is that you can do your evaluation, but if you want to make substantive changes, there's often up to a year to a year-and-a-half before you can put any substantive changes like that so in terms of turnaround … you know they call it often the quality control processes, what they do is I mean you know I've got a unit that needs to make changes then I've to wait a year for this year that can be another two interactions of the unit before I can make changes.

5. The fifth process related to the administration of the program was accreditation. Two lecturers (UL1 & UL3) argued that the constant accreditation requirements set by external agencies like AITSL and NSWIT were sometimes disturbing for both student teachers and lecturers. These requirements sometimes imposed changes to the duration of the program and imposed unexpected extensions in the duration of candidature of student teachers who were struggling to navigate their way through these changes.

Concerning the third aspect, Program Evaluation, two lecturers (UL4 & UL5) argued that the existing evaluation mechanisms carried out by the academic board and those performed for the accreditation of the program were sufficient and effective:

> I suppose that during the mapping process [referring to the accreditation] things got looked at, so that was a formal evaluation. There's informal evaluation happening all the time in individual units, like I evaluate my units every time I teach it and I keep a little check, print off of what I have to change next time, what I need to change in the assignment, what I could add, what I could take away, lots of people do that. That kind of stuff is happening all the time, the units are not static things that are set and forgotten. (UL4, Interview 2)

Moreover, UL5 argued that the unit evaluations were supplying lecturers with the necessary information for improving their units and that these were sufficient for improving the program overall.

In contrast, four lecturers (UL1, UL2, UL3 & UL8) expressed concerns about how program evaluation was happening. UL1 and UL2 argued that proper evaluation was not happening in the first place, and that the only time reflection was done on the program and not on individual units was when a new program was added or an old one was removed:

> Evaluation of the program?! I haven't seen anything ... Wow ... I haven't seen any sort of exit surveys of the GradDipEd. And I'm aware ... I mean they publish the student evaluations of all the units that go into the GradDipEd, because all the units taught in the school, you know they publish those evaluations ... So if your unit is doing really

> well, everyone in the school will know that, if yours is doing poorly, they all know that as well, so there's transparency there. But no one ... no one takes all of the units that are just the GradDipEd units and puts them into a thing as a GradDipEd. Maybe the course coordinator does that, but nothing comes to the unit coordinators. (UL1, Interview 2)

> Here, the only reflection done on programs rather than on units, the only reflection done on that is when the head of the school decides we need a new course. (UL2, Interview 1)

UL2 and UL3 criticized how the decisions about the longevity of the programs were solely made in relation to attrition rate and not based on pedagogically sound arguments:

> I think it was in a school meeting, the Head of School mentioned that he may be getting rid of some courses because of low students enrolments in all the units in those courses, now these are just little courses that we run. Well, that's probably how they are evaluated. The Head of School looks and says "are we getting enough students in that course?" if not, it's obviously not a very attractive course, so we get rid of it. (UL2, Interview 1)

> So evaluation happens at that level, and unfortunately, it may only seem about getting money in the door how many students will enrol in course. Even our units are judged the same way. In fact at the last school meeting, everybody also said we got to look again at our units; any small units with small enrolments they might have to get rid of them. (UL3, Interview 1)

UL3 and UL8 also expressed concerns about the course accreditation being not sufficient to make evaluative judgments about the program; UL3 (Interview 1) argued that “there are some things related to the program such as the duplication of materials, the scope and breadth of coverage of materials across the various units” that transcend the scope of the unit evaluations. UL8 agreed with UL3 and suggested that a combination of evaluation methods should be sought after:

> I think people now, with all the new accreditation processes, believe that if it goes through NSWIT [New South Wales Institute of Teachers] and AITSL and then TESQA [Tertiary Education Quality Standards Agency], then it’s all right, but it’s not always the case. Students and lecturers can be very good judges … but students can also be poor judges of what a course is. Because a course makes you work, they might say it’s bad! Or because they don’t like the way a particular lecturer talks or does things, they might say it’s bad. Whereas it might be very good for them, and that lecturer is just enlightening them to the ways they’re going to have to cope without things. However, I think it’s probably – like research, I think if you can triangulate from all different angles, it’s probably better to do that. (UL8, Interview 2)

5.1.8 Theme 8: Program structure

This is where informants' views about the structural aspects of the program were grouped. As shown in Table 5.9, aspects discussed included embedding the PREX units within other units in the program and the requirement of having a science degree for being admitted into the program.

Table 5.9 – Aspects discussed under "Program structure"

Aspect	Positively expressed by:	Negatively expressed by:	Nature of the statement
Embedding the PREX units within other units	0 teachers 2 lecturers	3 teachers 3 lecturers	Concern for teachers Issue for lecturers
Having a science degree as a prerequisite for being admitted into the program	2 teachers 2 lecturers	0 teachers 0 lecturers	Claim for teachers Claim for lecturers

Concerning the first aspect, embedding the PREX units within other units in the program, three teachers (ST2, ST3 & ST6) expressed concerns about the fact that the units were still running while they were taking their PREX units. ST2 (Interview 1) explained:

> Whilst I had to go on prac there were numerous occasions where I had to miss out on lectures, I had to catch up on assessment tasks, practical investigations, because I was away on PREX so I had to spend a lot of extra time catching up on those which ended up compromising the quality of my assessment tasks!

In contrast to the teachers' point of view, two lecturers (UL3 & UL9) argued that embedding the PREX units within other units was beneficial from pedagogical and economical perspectives. UL3 argued that embedding the PREX in units could help lecturers draw connections between the content they teach and the professional experiences of the student teachers and to adjust their teaching towards the PREX. Furthermore, this way of structuring the PREX units enables student teachers to experience what the profession looks like before they finish their qualification so that they can withdraw from the program early on during their studies in case they did not like teaching:

> The GDE is the only one where they actually get 10 days to observe real teachers in real classrooms. This means that they actually get to see what schools look alike these days, not what they remember, and it's not a huge number but we get a percentage who say "forget it, teaching isn't for me" because they see it early on, first start and say "No! I can't cope" and that means that they withdraw very quickly and don't get the huge debt of money waiting til near at the end. (UL3, Interview 1)

UL9 (Interview 1) takes a more economically driven perspective and argues that the embedding the PREX units within other units means that student teachers will get funding for those units:

> Previously, the PREX units were sitting outside other units, unfunded, and there was no HECS [Higher Education Contribution Scheme] related money so student, when they pay their fees, there was no unit there, this was just an extra additional requirement that students had to pay for. Embedding the PREX in those units … that still might need to be played with a bit because the units would be running while students are on PREX, but at least it's funded now!

Three other lecturers (UL1, UL3 & UL5) expressed concerns as to the way the PREX units were integrated within other units. UL1 (Interview 2) argued that the benefits of embedding the PREX units within other units came at the expense of flexibility for student teachers and scope and depth of content coverage for lecturers:

> [B]efore when they were individual units [students] could choose whenever in the school year [they] wanted them, so it would be 20-day Prac at some time during the year; now, it has to be a unit that [students] enrol in for that trimester which has some sort of assessment associated with it as well.

> I used to have weekly programs of 12 weeks of teaching, and students would know what is expected each week. When they put the PREX units in, that really interfered with it. So, if students are on PREX from weeks 3 to 7 then they can't be giving their full attention to what's on the program for those weeks. So what's happened is that there was a directive from higher up that we should rather than have our students try to fit some of the teaching around their PREX so they've got 12 weeks of teaching, we only designate eight weeks for teaching. So there will only be eight weeks to cover the course material and the other four weeks are for PREX. My students lose four weeks and even if most of them aren't going to do it then I'm still required to have eight weeks of teaching so that it fits with the other units across the way. So everyone is down to eight weeks of teaching from what used to be 12 from what used to be 14; so it's getting squeezed all the time and when I look at my program from last year to this year and then I'm thinking, well what can I afford to cut out of this now, there's no longer discretionary things that I think are not so important or maybe picked up in another unit. Now you're taking out some really important things.

UL3 & UL5 argued that because the PREX units are currently done on top of a unit's full load, embedding them within other units is not fair for the students:

> Professional experience has no credit points, it's done on top of the full load; there's no recognition of the workload of professional experience for the student so they've got four weeks in the school which includes planning, reporting, assessment but they have to contact the school, they have to try and find out what they will be teaching, do a bit of preliminary work, and so forth. Now, all that is happening while they're doing four other units on a full time load. Then they're four weeks in the school

> while they're still studying and then they have to catch up when they come back. (UL3, Interview 1)

Concerning the second aspect, having a science degree as a prerequisite for being admitted into the program, two teachers (ST2 & ST5) and two lecturers (UL3 & UL5) argued that having a science degree as a prerequisite for being admitted into the program was beneficial for the student teachers as it (1) provided them with extant content knowledge to teach at the secondary level with confidence,

> It's an asset of the program, that they come with a full degree behind them particularly that they're being trained to teach into years 11 and 12 as well, they need that thorough knowledge of whatever their subject is. (UL3, Interview 1)

(2) offered them a degree to fall back on in case they decide to leave the teaching profession,

> I possibly will leave teaching one day and go into science field and so I'm thankful that I have done extended sciences in my degree. If it was basic science in my degree then I wouldn't be able to have such an opportunity to be able to work in the industry. So I'm lucky that I have a full science degree. (ST2, Interview 2)

and (3) enabled them to upgrade their qualifications to a master in teaching in just 6 additional months.

> [B]ecause of the articulation that the program has to the current Master of Teaching Secondary, people can come in, they can get the shorter award and then they can teach because they're fully qualified and then should they want to, they can then do the extra six months and upgrade to the Master of Teaching Secondary. (UL5, Interview 1)

5.1.9 Theme 9: University lecturers

As shown in Table 5.10, under this theme, I placed informants' discussions about the teaching backgrounds, knowledge and expertise of the lecturers.

Table 5.10 – Aspects discussed under "University lecturers

Aspect	Positively expressed by:	Negatively expressed by:	Nature of the statement
Lecturers' professional knowledge	1 teacher 2 lecturers	0 teachers 0 lecturers	Claim for teachers Claim for lecturers
Lecturers' experiences in school teaching	1 teacher 1 lecturer	2 teachers 0 lecturers	Issue for teachers Claim for lecturers

Concerning the first aspect, lecturers' professional knowledge, one teacher (ST2) and two lecturers (UL4 & UL8) explained that lecturers teaching into the program were knowledgeable and could relate well to students.

Concerning the second aspect, lecturers' experiences in school teaching, one teacher (ST4) felt that lecturers had relevant school experiences, which they could easily draw upon in their teaching. This made their teaching more relevant, ST4 added:

> I found that the lecturers were pretty much school teachers who have stepped up. So they have that foundation of teaching and so they have that experience to draw upon, and so they are not theorizing about what could happen in a classroom. They are talking about what has happened in a classroom to them, so it's much more relevant. (Interview 2)

However, two other teachers (ST1 & ST2) disagreed and said that university lecturers seemed to lack relevant school experiences to draw upon in their teaching, which made their advice seem a little detached from the school contexts. ST1 (Interview 1) said:

> I'm curious as to when was the last time my lecturers have actually taught in high school? that I think should be mandatory, I think they should be teaching in a high school so they can offer constant practical experience, because I just don't think that aspect comes through in their teaching.

ST2 (Interview 1) added:

> I've been there and I've done my PhD I've been stuck in that world! It's a very different world from being in a classroom and being on the ground. I don't feel like any of my teachers have been in the classroom, you know as much as I loved my science education units, it's still not feeling that any of them has been in a classroom recently!

While UL8 argued that many academics who teach into the course have a decent career in school teaching, she acknowledged the teachers' point of view and explained that there could be, sometimes, a trust issue between lecturers and student teachers and that some protocol should be put in place to restore this trust. UL8 explained further that lecturers need to prove their teaching skills before the student teachers in one way or another.

5.2 Lists of claims, concerns and issues

In the previous section, I illustrated in detail the various themes and CCIs that were discussed by the participants. In this section, I presented lists of claims, concerns and issues, and use them to briefly discuss and analyze the commonalities and difference in the CCIs of the two stakeholder groups.

5.2.1 The claims constructed from the data

Table 5.11 summarizes the Claims that were expressed during the interviews. The table also shows the relative strengths of these claims as evident by the number of stakeholders who discussed them.

Table 5.11 – List of claims constructed from the data

	Theme	Aspect	Nature of the statement
Common Claims across stakeholder groups	Delivery mode	Flexibility of the off-campus program	Claim for 1 teacher Claim for 1 lecturer
	Content of learning	Writing programs and lesson plans	Claim for 2 teachers Claim for 1 lecturer
	Program structure	Having a science degree as a prerequisite for being admitted into the program	Claim for 2 teachers Claim for 2 lecturers
	University lecturers	Lecturers' professional knowledge	Claim for 1 teacher Claim for 2 lecturers
Claims expressed by only one group	Teaching strategies in the program	Nature and usefulness of assignments	Issue for teachers (1 Claim; 3 Concerns) Claim for 5 lecturers
	University lecturers	Lecturers' experiences in school teaching	Issue for teachers (1 Claim; 2 Concerns) Claim for 1 lecturer

As can be seen in Table 5.11, there were four aspects related to four themes which were Claims of the two stakeholder groups. These aspects are: the flexibility of the off-campus program; writing programs and lesson plans, having a science degree as a prerequisite for being admitted into the program; and the lecturers' professional knowledge. The Table also features the aspects that were Claims of one group of stakeholders but not by the other. For example, while the "nature and usefulness of the assignments" was the aspect most frequently cited by university lecturers as an asset of the GDE(ST), it was an issue for teachers with one teacher considering it as a Claim and three others considering it a Concern.

5.2.2 The concerns constructed from the data

Table 5.12 summarizes the Concerns that were expressed in the interviews. The table also shows the relative strengths of these Concerns as evident by the number of participants who discussed them.

Table 5.12 – List of concerns constructed from the data

	Theme	Aspect	Nature of the Statement
Common Concern across stakeholder groups	Content of learning	Duplication of content	Concern for 1 teacher Concern for 5 lecturers
Concerns expressed by only one group	Guidance and support	Support in selecting units	Concern for 2 teachers Not discussed by lecturers
	Content of learning	Appropriateness of content for secondary science teaching	Concern for 3 teachers Not discussed by lecturers
		Learning to deal with stakeholders (e.g. parents and colleagues)	Concern for 2 teachers Not discussed by lecturers
		Learning how to do an accreditation portfolio	Concern for 2 teachers Not discussed by lecturers
	Teaching strategies in the program	Quality of instruction in the professional experiences	Issue for teachers (1 Claim; 5 Concerns) Concern for 2 lecturers
	Program monitoring processes	Collaboration between lecturers	Not discussed by teachers Concern for 5 lecturers
		Assigning workloads to lecturers	Not discussed by teachers Concern for 4 lecturers
		Accreditation	Not discussed by teachers Concern for 2 lecturers
	Program structure	Embedding the PREX units within other units	Concern for 3 teachers Issue for lecturers (2 Claims; 3 Concerns)

As can be seen in Table 5.12, only one Concern (duplication of content across the program's units) was expressed by participants from both stakeholder groups.

On the other hand, nine concerns were expressed by one group and not by the other. Interestingly, with the exception of the “quality of instruction in the professional experiences” and “embedding the PREX units within other units”, all other aspects were a concern for one group and not discussed by the other. This observation suggests that the stakeholder groups might have different considerations for evaluating the program and, while one group expresses concerns about something, the other group seems indifferent to it.

5.2.3 The issues derived from the analysis of claims and concerns

From the analysis of Claims and Concerns, I was able to derive three categories of Issues (see Table 5.13). The first category included the aspects that were an Issue for both stakeholder groups. The second category contained the aspects that were an Issue for one stakeholder group but were not discussed by the other group. In Table 5.13, these Issues are denoted as ‘Issues peculiar to only one group’. The third category of Issues contained the aspects that formed an Issue for one group and either a Claim or Concern for the other group. In this last category, denoted as “Issues that exist between stakeholder groups” in Table 5.13, I would have also included those aspects that were Claims for one group and Concerns for the other, but there was no such case from my data.

As can be seen in the table, four Issues were common to the two stakeholder groups: “lecturers' support throughout the units”, “the off-campus learning experience”, “developing teachers’ PCK”, and “relevance and usefulness of content learned”. Participants for discussion in the negotiation session also nominated all but the first. This indicates the importance of these aspects to both stakeholder groups.

Another pattern that presents itself in Table 5.13 is that 11 out of the 19 Issues (more than half) featured in the table were discussed by only one stakeholder group and not by the other. Again, this observation reinforces the inference made earlier about stakeholders’ concerns stating that the criteria used for program evaluation by the two stakeholder groups might be different.

Table 5.13 – List of issues derived from the analysis of claims and concerns

	Theme	Aspect	Nature of the statement*
Issues for both stakeholder groups	Guidance and support	Lecturers' support throughout the units	Issue for teachers (2 Cl; 1 Co) Issue for lecturers (3 Cl; 2 Co)
	Delivery mode	The off-campus learning experience	Issue for teachers (1 Cl; 2 Co) Issue for lecturers (1 Cl; 1 Co)
	Content of learning	Developing teachers' PCK	Issue for teachers (1 Cl; 3 Co) Issue for lecturers (4 Cl; 1 Co)
		Relevance and usefulness of content learned	Issue for teachers (2 Cl; 4 Co) Issue for lecturers (2 Cl; 1 Co)
Issues peculiar to only one group	Attitudes towards the program	Enjoyment of the program	Issue for teachers (2 Cl; 2 Co) Not discussed by lecturers
		Motivation to do the program	Issue for teachers (1 Cl; 2 Co) Not discussed by lecturers
	Guidance and support	Admission into the program and recognition of prior skills	Issue for teachers (1 Cl; 1 Co) Not discussed by lecturers
		Supervising teachers' support during the professional experiences	Issue for teachers (1 Cl; 1 Co) Not discussed by lecturers
	Content of learning	Learning the content of the HSC syllabuses	Issue for teachers (2 Cl; 2 Co) Not discussed by lecturers
	Residential school	Arguments for having a mandatory residential school	Not discussed by teachers Issue for lecturers (2 Cl; 5 Co)
	Teaching strategies in the program	Quality of instruction in the units	Issue for teachers (1 Cl; 5 Co) Not discussed by lecturers
	Program Monitoring Processes	Admitting student teachers into the program	Not discussed by teachers Issue for lecturers (1 Cl; 3 Co)
		Trimesterization	Not discussed by teachers Issue for lecturers (1 Cl; 4 Co)
		Meeting the academic board requirements	Not discussed by teachers Issue for lecturers (1 Cl; 3 Co)
		Program evaluation	Not discussed by teachers Issue for lecturers (2 Cl; 4 Co)
Issues that exist between stakeholder groups	Teaching strategies in the program	Quality of instruction in the professional experiences	Issue for teachers (3 Cl; 2 Co) Concern for 2 lecturers
		Nature and usefulness of assignments	Issue for teachers (1 Cl; 3 Co) Claim for 5 lecturers
	Program structure	Embedding the PREX units within other units	Concern for 3 teachers Issue for lecturers (2 Cl; 3 Co)
	University lecturers	Lecturers' experiences in school teaching	Issue for teachers (1 Cl; 2 Co) Claim for 1 lecturer

* Cl stands for claims; Co stands for concerns

5.3 Issues addressed in the negotiation round

Table 5.14 outlines the nine issues that were presented on the negotiation forum after having been nominated as most important by the participants. In what follows, each issue is examined at length. The participants' constructions concerning the issues are explored in a narrative format. The narratives feature the arguments put forth by the different negotiators as well as their suggestions for improvement in relation to each issue.

Table 5.14 – Issues raised in the negotiation forum

Issue title	Definition of the issue
The off-campus learning experience	The issue here is about how to use discussion boards and forums (and the online environment in general) to improve student teachers' experiences.
Developing teachers' PCK	There is disagreement about whether the program provides student teachers with enough resources to develop their base-level pedagogical content knowledge (PCK).
Relevance and usefulness of content learned	There is discrepancy in the perceived relevance of the content learned in the program.
Arguments for having a mandatory residential school	The issue is less about actually having a residential school or not than it is about whether the goals of the residential schools are being achieved in any other way.
Quality of instruction in the professional experiences	Although not an issue, participants decided to discuss this Concern as they believed that the practicum was not being monitored properly while it should have been.
Nature and usefulness of assignments	The issue here is about agreeing on a balance between theoretical and practical aspects of the assessment tasks
Collaboration between lecturers	There is a common concern among lecturers about the minimal amount of collaboration in the school when it comes to course-related matters.
Program evaluation	Despite there being in place mechanisms for course content "evaluation", the issue here is whether there needs to be similar or other mechanisms for program implementation evaluation and monitoring.
Embedding the PREX units within other units	The issue here concerns the identity and form of the PREX units.

5.3.1 Issue 1: The off-campus learning experience

The issue here was about how to use the online environment to improve the off-campus students' learning experiences. In the two rounds of interviews, teachers and lecturers expressed mixed views about the relevance of the discussions that take place on the

discussion boards and whether online tools could be used to promote a social constructivist learning environment. Therefore, the question derived from the informants' input, which guided the negotiations was: What sort of learning/teaching settings can and should be created through an online environment to promote more vivid experiences for external students?

Four lecturers (UL1, UL4, UL5 & UL6) responded to this question. All argued that the online environment should be developed so as to simulate a face-to-face constructivist learning environment. UL6 explained that the content on the discussion boards, for example, should be designed to promote student reflection and encourage discussions between students. The role of the lecturer in that respect is that of a facilitator who creates opportunities where these reflections and discussions can take place in a meaningful fashion.

UL5 added that the discussion boards could be used by student teachers and lecturers alike to monitor students' learning throughout the units by exploring whether and how the student's ideas and understandings are evolving. She explained:

> I think the key consideration is how the online discussion boards and forums are used pedagogically. There are some units that require students to post statements about their particular view/understanding of an issue in week 1, before they have engaged with any of the online learning materials, readings etc. This posting is a mandatory formative assessment task and students have to revisit this posting in subsequent summative assessment tasks to gauge how their views/understandings have changed/developed. They may end up critiquing their original views or have had original views affirmed and strengthened, or they may position their views in relation to a range of theoretical positions presented in the associated literature or in relation to other initial postings etc. This seems to be a constructive use of online discussions that prompts purposeful reflection. (UL5, Negotiation Forum)

UL4 agreed that online tools should be used to monitor students' learning and argued that various tools could be developed further to catalyze students' engagement:

> [L]inking forum questions to online surveys they complete relating to some contentious issue, or engaging images or current media stories or AV resource to enhance student engagement with them. (UL4, Negotiation Forum)

> [P]rogress tracking which apparently gives the coordinator an immediate visual indication of the engagement of the cohort with various key aspects of the unit – useful to spot at-risk students early on and respond in a timely way. (UL4, Negotiation Forum)

UL1 suggested that a key consideration for promoting effective use of the discussion boards among other online tools is "whether or not to require student participation as a component of students' assessment". UL1 (Negotiation Forum) explained:

> In some units this is required, and accounts for 5-10 per cent of the unit grade. The advantage of this is that students are far more inclined to engage with online discussion topics and to interact with the lecturer and other students. When done well, this approach can lead to better learning outcomes. One disadvantage is that it is difficult to specify anything about the quality of the contributions. Students may simply add generic comments. On the other hand, it can be argued that this happens in face-to-face tutorials anyway. Another disadvantage is the additional time required of lecturers to check and assess these contributions. To see them in context, a lecturer would need to read the thread, so having students simply cut and paste their forum contributions to their assignments would be inadequate. I know of a number of lecturers who tried this type of approach but have now removed it from their assessment regime because of the additional time and trouble involved"

Several suggestions have been proposed in the interview for resolving this issue. I have summarized them into one proposal and placed it on the negotiation forum so that participants could discuss it further. The proposal is: to manage the online forums, lecturers can dedicate a special forum for social stuff (e.g. virtual coffee shop), scan the forums for important questions which then can be migrated into a FAQ board to make it quicker for student teachers to find important stuff, and archive some forums every now and then and keep them read-only.

In discussing the suggestions related to developing FAQs and archiving forums, UL6 (Negotiation Forum) clarified that "the Moodle forum tool [used at the university] performs very poorly as it lacks both a compile post option – in which all of the posts in a thread can be merged into a single text document – and a bulk move option where old posts can be moved out of a current forum and into an archive.

UL1 (Negotiation Forum) presented another proposal related to the issue of off-campus learning and suggested that lecturers need to evaluate whether it is worth including forum contributions as assessment. He argued "it might be worth surveying students who have been enrolled in units where forum contributions are assessed and units in which this is not done, to evaluate the worth of this strategy".

In discussing UL1's proposal, UL4 cautions that the use of forum contributions as part of students' assessment might impose restrictions on students' participation and should thus be carefully planned for before any evaluation is undertaken:

> I do include online forum contributions as part of assessment tasks (e.g. refer to two of your posts and explain how/if your ideas have changed). I think it is good practice as long as it is not too "threatening" and I know that some students hate it – but then some students hate speaking up in class as well. However there is a trade-off: I try to explicitly encourage the kind of "discussion" with all its attendant blurry understandings, half-formed ideas that in-class discussion is often like – which leads to learning, rather than is a description OF learning. I don't necessarily want them to feel like their contribution has to be the last word on the subject (or they won't "talk" at all). Assessing contributions I think pushes them to the latter view – that their postings have to be finely tuned reflections of a carefully considered and justified view. (UL4, Negotiation Forum)

5.3.2 Issue 2: Developing teachers' PCK

The issue here was about the extent to which the program provides student teachers with enough resources to develop their base-level pedagogical content knowledge (PCK). From the two rounds of interviews, I did not identify reflective discussion questions related to this issue and, therefore, on the forum, there were no questions displayed. Nevertheless, participants were encouraged to pose their own questions. UL5 initiated the discussion by posing a question. UL5 first explained that the program is accredited by AITSL and therefore successfully meets the requirements set by AISTL, which specify what graduate teachers should know and be able to do. Then UL5 asks whether the issue concerns the accreditation standards themselves or how well the program meets those standards:

> In order to be accredited, teacher education institutions present submissions to the relevant regulatory authority outlining how programs such as the Grad Dip Ed meet the graduate teacher standards. Is this issue addressing whether the graduate standards, upon which accreditation is based, are sufficient? The standards are generic. Given that this study concerns the preparation of science teachers only, is this question indirectly asking whether standards should be subject-specific, as in the USA? Is this question a critique of how well the Grad Dip Ed meets certain standards/descriptors as specified in the Australian Professional Standards for Teachers, e.g. Standard 2: Know the content and how to teach it. (UL5, Negotiation Forum)

UL5's question was not discussed by any other participant and, therefore, the issue remained unresolved. On the other hand, one suggestion has been proposed in the interviews for resolving this issue and was discussed by two lecturers (UL4 & UL6). The suggested action to develop students' PCK is to offer, within the science education units, a series of dedicated workshops that provide student teachers with ready-to-use resources and that allows them to try out scientific experiments.

UL4 (Negotiation Forum) agreed with this proposal and explained that the Australian Academy of Science offers freely available resources that are mapped to the Australian

curriculum and directing student teachers to such resources might help enhance their PCK. Additionally, UL6 recommended expanding the notion of PCK to include the Technological Knowledge dimension. In an attempt to encourage input about the implications of this suggestion, I asked whether this recommendation had implications to the way units are run. I also argued that introducing this technological knowledge dimension suggests some coordination between lecturers so as to infiltrate the TPCK into the contents of other units; however, the discussion was not followed up by any other participant.

5.3.3 Issue 3: Relevance and usefulness of content learned

The issue was about the discrepant perceptions of teachers and lectures about the extent to which the content is both relevant and useful. Throughout the two rounds of interviews, teachers and lecturers expressed mixed views about the relevance of the content and its usefulness in the school context. The question derived from the informants' input and which guided the negotiations was: Is there a gap between the actual relevance and perceived relevance of the program's content or is this program deficient in connecting theory to practice?

UL4 argued that teachers would be the best informants about this question as they are able to identify the irrelevant aspects of the course or aspects that are in disconnect with what happens in schools:

> The answer I'm guessing might be a bit of both. It is a problem if newly practising teachers who are looking back at their program and evaluating it in the light of their school experience can't see the relevance of what they have learned. Would be good to know if there are specific aspects they could point out that were or weren't relevant so we can have a good look at it – any ex-students out there want to make any suggestions? It needs to come from you so please let us know. (UL4, Negotiation Forum)

However, none of the student teachers took part in the negotiation forum and the discussion was terminated at this point. On the other hand, three suggestions have been proposed in the interviews for resolving this issue but only the first was negotiated.

The first suggestion made to enhance the relevance of the content was to make the units focus less on different aspects of teaching and learning – that is, pedagogical knowledge – and more on how to teach the various dot points of the junior and senior science syllabuses – that is, pedagogical content knowledge. In that respect, UL1 explained that the perceived relevance of the program was a relative matter and that student teachers might not appreciate

the relevance of some of the important aspects in the program because the context of the school where they are first employed makes those aspects seem irrelevant:

> "Our students complete 12 units of study, only 4 of which are in science education (or 2 if they are majoring in two teaching subjects). The other 8 units are in generic teaching and learning units or in educational context units. The relevance of these is not always apparent to the first year out student struggling with science lesson plans and resources. Take Aboriginal Education as an example, depending on where the student is first employed, the relevance of this unit might seem very high or very low. The relevance of units which seek to equip our graduate students with a general understanding of aspects of student diversity always varies with school context, whereas the science syllabuses are a constant" (UL1, Negotiation Forum)

UL1 then suggested that perhaps more science education specific units should be made available to students. UL4 agreed with that suggestion and added that student teachers would both benefit from this and enjoy it. Furthermore, UL4 pointed out that these units should be made for both on-campus and off-campus students.

The other two suggestions that were derived from the interviews and which were not negotiated were:

- Having the lecturers model the various teaching and learning theories that they teach about in order to operationalize and concretize these concepts for student teachers thus making them more meaningful.
- Developing and implementing transparent processes that aim at connecting the learning that happens at the university to its application in schools during the PREX so that student teachers can experience minds-on and hands-on the relevance and usefulness of what they learn.

5.3.4 Issue 4: Arguments for having a mandatory residential school

The issue here was about whether or not the program should include a mandatory residential school and whether the goals of the residential schools could be achieved in any other way. The questions derived from the informants' input, which guided the negotiations were: What are the specific goals to be achieved by having a residential school? How can each of these goals be achieved otherwise?

The discussions held in the interviews about the importance of the residential schools were mainly derived from arguments coming from incompatible vantage points. Proponents of the residential school argued that there are pedagogical reasons for making them mandatory while opponents of making residential schools mandatory argued from a market-driven perspective. In the negotiation session, the discussion was focused on the pedagogical reasons for keeping the residential schools mandatory or replacing them with a virtual alternative. In

that respect, UL4 presented her view on the goals of the residential schools and argued that all of them could be pursued using a virtual alternative:

> I guess goals would include: 1) Interacting with fellow students, 2) Having dedicated time to focus on the program (many OFF students are busy juggling multiple lives and residential schools take them out of these other lives), and 3) Doing hands-on scientific activities to enhance students' confidence and expertise in the PCK and also the practicalities of divvying up equipment, planning for safety, monitoring safety issues etc. (UL4, Negotiation Forum)

Then UL4 (Negotiation Forum) argued that using virtual environments can also achieve these goals. In relation to enhancing student interactions, student teachers can "interact online asynchronously via discussion forums or even synchronously via Adobe Connect or Virtual worlds or some other alternative". With regard to dedicating time for the program, UL4 argued that if student teachers really wanted to, they could do this with whatever tools are already available for them. She added, "the fact that many students do not sign up for the res schools is testimony in part to the demands of juggling roles" (Negotiation Forum). Regarding the last goal, UL4 explained that the PREX should provide ample opportunities for student teachers to develop their PCK:

> The majority of our students have done their science degrees and have had (at least) three years of practical hands-on science so there should be some understanding of how to conduct themselves in a lab and use science equipment. Sure this doesn't directly translate to TEACHING those skills/content areas though. This should be done in the students' practicum placements. There is also scope for role playing these activities in virtual worlds, using AV material to raise some of the issues and so forth. (UL4, Negotiation Forum)

UL1, on the other hand, extended the list of goals achieved by residential schools and argued that, while the goals listed by UL4 could be achieved through other means, there are others that cannot be replaced by a virtual component. He listed the following goals:

> 1) the opportunity to meet other students face-to-face; and 2) to have frank and non-public discussions about assignments and other aspects of the course; 3) the opportunity for students to build confidence in their capacities to run practical lessons in a school lab; 4) opportunities to have immediate peer and lecturer feedback on ideas and teaching skills. (UL1, Negotiation Forum)

Then UL1 argued that residential schools should be kept mandatory:

> Face-to-face learning at university offers situated cognition and better facilitates the "social construction" of knowledge and skills, as opposed to the individual construction available via off-campus learning. It also offers a mode of learning – through speaking and listening – which is more familiar to most students and is the

> general mode in schools. By comparison, off-campus learning is mostly through reading and typing. (UL1, Negotiation Forum)

The issue was not discussed any further and therefore remained unresolved. Furthermore, no suggestions were presented or discussed by the negotiators concerning this issue.

5.3.5 Issue 5: Quality of instruction in the professional experiences

Although not discussed by lecturers in the interview rounds, teachers decided to push this issue to the negotiation forum as they disagreed on the extent to which PREX was being adequately monitored to ensure that adequate learning occurred in the program. While three teachers argued that the supervising teachers were helpful and that they learned a lot from them, two others argued that the lecturers were performing poorly in providing feedback for student teachers and encouraging reflective practice.

Two questions guided the negotiation about this issue:

1. How do we reinforce the learning and acquisition of educational theories and skills during PREX?
2. What kind of support do lecturers need to offer for student teachers during PREX?

With regard to the first question, UL4 explained that most of the units have an online forum where student teachers can discuss their experiences and provide mutual support. She further argued that the role of the lecturers in connecting the learning from the program to practice could be enhanced if lecturers share the information about the teaching and learning theories student teachers have been studying with the supervising teachers. No other lecturers further discussed this question.

For the second question, UL1 argued that he encourages reflective practice with student teachers upon their return from the PREX and that on-campus student teachers enjoyed this activity. He further explained why this activity was more challenging for the off-campus student teachers and argued that further discussions about this matter are needed:

> One aspect of the practicums which concerns me is the lack of debriefing and reflection among off-campus students. When my on-campus students come back from prac, we have an informal coffee session where we share the best and worse experiences on prac. I facilitate a guided discussion in which students talk about what they would have done differently, what they learned, and what they wish their units had better prepared them for. These sessions are very interesting, fun and valuable. For the off-campus students I have Moodle discussion forums dedicated to prac experiences in my units to afford them the same opportunity, but these are not well used. I suspect students are less willing to share their experiences with others they

> don't know, and I imagine they are certainly less willing to write about their weaknesses or share their teaching disasters. Also, off-campus students are undertaking prac at different times, so there is no designated "PREX time" as there is for on-campus students. It could be argued that a written reflection on their PREX could be used as part of their assessment, but there is very little capacity to add to assessment workloads. I would welcome suggestions from other participants about how this might be addressed. (UL1, Negotiation Forum)

UL4 argued that she faced similar challenges with her off-campus student teachers and that there should be in place some protocols that lecturers can refer to in helping student teachers in distress while on PREX.

In relation to the suggestions about resolving the challenges associated with the PREX, two propositions were made by the teachers during the interviews:

- Establishing schools-university partnerships thus promoting a culture of teacher training in those schools characterized by reflective practice.
- Invest in virtual tools that help lecturers play a more efficient and transparent role in the students' PREX

While these suggestions have not been negotiated, both UL1 and UL4 embraced those ideas. Furthermore, UL1 (Negotiation Forum) explained that "a number of initiatives have been undertaken in the School around virtual supervision" but that there are still challenges related to its implementation due to "a lack of resources, a lack of agreement on how to proceed or a lack of will".

5.3.6 Issue 6: Nature and usefulness of assignments

The issue here is about the extent to which the assignments were practically oriented. While five lecturers argued that most of the assignments are practically oriented, three teachers found the assignments to be theoretically oriented and unreliable for use in a real classroom.

While no questions have been derived from the interview rounds, there was one suggestion made to enhance the practicability of the assignments and which were made open to negotiation in the forum. The suggestion was to have more variety in the types of assessment used and to include some hands-on elements like performing a mini-lesson plan.

Both UL4 and UL6 agreed with this suggestion and explained that hands-on experiments are the essence of science teaching and learning. However, UL4 argued that incorporating hands-on elements in the assignments could be difficult to implement, particularly for the off-campus students. UL8 extended the suggestion further and proposed that, if lecturers want to

link the theory and practice of learning and teaching, they ought to develop assessment tasks that simulate real life contexts. UL8 gave an example to explain her proposal: “we can ask them 'If you are a practicing teacher and you have to present your assessment plan to your students' parents, how do you intend to do this?’” UL1 agreed with UL8 and explained:

> Quite a few of my students complain about the de-contextualised nature of some of their assignments. Even though the lecturer setting the assignment may see the longer-term connection to teaching practice, sometimes this is unclear to students. It is a common comment made to coordinators of curriculum units (maths, science, English etc.) and teaching and learning units that these are the most relevant and enjoyable within the course. Even in more theoretical units, placing assignments in relatable school-based contexts might increase student motivation to undertake them. (UL1, Negotiation Forum)

5.3.7 Issue 7: Collaboration between lecturers

Although not an issue, lecturers decided to discuss this concern further in the negotiation round as they were worried about the minimal amount of collaboration in the school when it comes to course-related matters. Three questions were derived from the lecturers’ discussions in the interviews and were used to guide the negotiation of this issue:

1. What specific course-related matters could be enhanced through more collaboration?
2. What sort of collaboration channels and structures need to be in place to enhance those ties between lecturers?
3. How can we create collaboration channels between lecturers teaching into one course?

However, only the first question was discussed. In that respect, UL4 argued that enhanced collaboration at the program level could inform lecturers further about the detailed program structure so as to know what is covered, in which units and to what extent. UL6 added that, through collaboration, lecturers can achieve some uniformity across the various units in the program and that this would not only reinforce the learning that student teachers carry from one unit to the next but that it could also remove some of the burdens from students. Finally, UL8 argued that increased collaboration between lecturers reduces “overlap” of issues and raises awareness of important topics, which are neglected when lecturers do not collaborate.

One proposal has been suggested by the lecturers during the interviews: allocating time for lecturers to get actively involved in collaborative activities that focus on the program enhancement. UL4 argued that even if it happens once a term, this would be quite beneficial for lecturers. Alternatively, she suggested that there could be in place some induction or

information website for lecturers starting to teach in a particular program that specify who the unit lecturers are and other relevant information about the program and the units.

5.3.8 Issue 8: Program evaluation

Despite having mechanisms for the evaluation of the program's content (accreditation through the academic board), the issue here is whether there needs to be similar or other mechanisms for evaluating the program's implementation. Three questions were derived from the lecturers' discussions in the interviews and were used to guide the negotiation of this issue:

1. What aspects of a program need to be evaluated? How and by whom?
2. What resources (personnel, structure, funding) need to be in place to establish a monitoring feedback loop?
3. Does this process of program implementation evaluation need to be institutionalized?

Concerning the first question, UL4 argued that all aspects of the program need to be evaluated. Neither UL4 nor any other lecturer elaborated their ideas about what these aspects were. UL4 suggested that the program evaluation should involve the program coordinators, the school Teaching and Learning Committee, the unit coordinators, and the student teachers through some participatory model. UL1 explained that graduate teachers' feedback would be the most important source of data for that kind of evaluation.

In relation to the other two questions, UL4 suggested that there needs to be an evaluation culture in the school that encourages lecturers to take part in those activities. UL1 (Negotiation Forum) discussed this point at length:

> I'm picking up that there seems to be an underlying reluctance to advocate systematic and effective course evaluation because of the additional workload this would require of staff. Certainly I have been hesitant in my responses to what are quite sound and reasonable questions and suggestions because of the additional effort which might be required to implement a more integrated evaluation. Reflection and Evaluation are two of those processes which should be part and parcel of our professional duties, but which are often squeezed out by more pressing priorities. Academics have high workloads and tend to work long hours, so even the most reasonable suggestions about ongoing course evaluation can be met with a little reluctance, as there is so little room left to devote to it. Hence I think any system of course evaluation would need to be a greater part of the university culture – e.g. university wide, integrated into school processes and more explicit in workload expectations. Otherwise it would just be spasmodic and ineffective"

In discussing the elements of an effective evaluation framework, UL1 and UL4 argued that evaluations should be collaborative, oriented towards improvement, participatory in nature and evidence based. They further added that they have to be relatively easy and quick for people to contribute to. UL1 explained that this is necessary particularly that the school of Education offers around 20 courses and that evaluation "would need to be organized so as not to be too much of an imposition on staff to provide feedback on all the courses they teach into" (UL1, Negotiation Forum).

5.3.9 Issue 9: Embedding the PREX units within other units

The issue here concerns the identity and form of the PREX units. In the interviews, three teachers and two lecturers expressed concerns about the fact that the units were still running while student teachers were taking their PREX units and that embedding the PREX units within other units has placed constraints about when student teachers can take those PREX. On the other hand, two lecturers presented opposing views, arguing that embedding the PREX in other units can help lecturers draw connections between the content they teach and the professional experiences of the student teachers and to adjust their teaching towards the PREX. The question derived from the informants' input and that guided the negotiations was: What configurations can be thought of for PREX (particularly in terms of allocating credit points and embedding it, or not, within other units) that acknowledge the loads of professional experiences on student teachers and lecturers and that maximize the support provided during PREX? None of the participants discussed this question and therefore the issue remained unresolved.

Some lecturers made one suggestion during the interviews. It called for allocating credit points to the PREX units and redefining its identity in ways that reflect its "independent" nature and reinforces its connections to other units in the program. While UL1 explained that PREX units have recently been allocated credit points, none of the other lecturers discussed this suggestion further.

5.4 Resolved and unresolved issues

Following Guba and Lincoln (1989), issues that are discussed in the negotiation forum can be completely resolved if participants agree on a course of action for addressing them. Otherwise, these issues would remain unresolved, although they could be redefined and put on an agenda for future negotiation. In this evaluation, I have used this criterion to differentiate between resolved and unresolved issues. Accordingly, I presented the resolved

issues as those about which lecturers who participated in the forum agreed. I also presented the unresolved issues and discussed how these were redefined based on the negotiations. Nevertheless, since only university lecturers took part in the negotiation forum, I am cautious about advocating the significance of the resolved and unresolved issues classification. While three teachers indicated their willingness to take part in the forum, none contributed to the negotiations of the various issues and, therefore, the negotiations were biased, being only carried out by university lecturers and myself.

Three of the issues raised in the negotiation forum were classified as resolved. Concerning the first issue, the off-campus learning experience, university lecturers expressed an agreement that these experiences should be improved based on the development of online tools that promote a social constructivist environment. With regard to the second resolved issue, collaboration between lecturers, negotiators agreed that further collaboration regarding many program-related aspects needs to take place between lecturers, and that the school of education should endeavor to allocate time for lecturers in which these collaborative activities can take place. The third resolved issue concerns the evaluation of the program. In that respect, negotiators agreed that there need to be better evaluative mechanisms to monitor aspects of the program, which are collaborative, oriented towards improvement, participatory in nature and evidence based. Furthermore, negotiators argued that they should be given time to engage in these activities, which should be recognized as part of their job and as part of the school's culture.

The remaining six issues raised in the forum were classified as unresolved. Three of these issues were not discussed by any of the participants and were therefore classified as unresolved. The first was related to embedding PREX in other units, the second concerned the relevance and usefulness of the content learned, and the third concerned the nature and usefulness of the assignments.

With regard to the fourth issue, developing teachers' PCK, I did not generate any questions to guide the negotiations. Nevertheless, UL5 initiated the negotiations by asking a question. However, other negotiators refrained from answering that question even though it was posted midway through the negotiation forum timeline. Nevertheless, UL5's question pointed out another dimension to the issue that had not been previously discussed in the interviews. The added insight related to the extent to which the accreditation standards were a good measure of a program's worth to those studying it. Notably, two lecturers did discuss the proposals made for improving the teaching and learning activities to enhance teachers' PCK. UL4

suggested that lecturers need to redirect student teachers to resources that are trustworthy and freely available, such as the Australian Academy of Science. UL6 also suggested including a Technology dimension to the PCK framework adopted.

The fifth unresolved issue concerned the arguments for making the residential school mandatory. While the negotiators raised the discussion to a new level and based their arguments on pure pedagogical grounds (whereas these discussions were previously partly pedagogical and partly economically driven), there was no consensus as to what the goals of the residential school were and whether or not these could be achieved through virtual alternatives.

The sixth and final unresolved issue concerned the quality of instruction in professional experiences. This issue, which was initially raised by teachers, has been discussed as a concern by lectures. The lecturers who discussed this issue agreed that the quality of instruction in professional experiences was problematic and they embraced both suggestions made by the teachers to establish schools-university partnerships thus promoting a culture of teacher training in those schools, and investing in virtual tools so as to gather more data about these experiences and thus make evidence-based decisions.

5.5 Conclusion

In this chapter, I outlined and discussed the findings derived from the evaluation of the GDE(ST) using the A4GE. In particular, I have discussed the themes that were discussed by the various stakeholders and have described the claims, concerns, and resolved and unresolved Issues constructed throughout the evaluation. In the following chapter, I discuss the findings derived from the empirical investigation of the implementation of the A4GE and which draw heavily on the insights that emerged and were reported in this chapter.

CHAPTER 6: FINDINGS FROM THE EMPIRICAL INVESTIGATION OF THE A4GE

The purpose of this chapter is to answer the research question posed at the beginning of this study:

> **How congruent are the underlying theory and practice of an adapted version of the 4GE model in the context of evaluating a secondary science teacher preparation program in an Australian university?**

To answer this question, I examined the extent to which each of Miller's (2010) four criteria was met. Therefore, I investigated whether the theory underlying the A4GE provides guidance for its application, specifies the conditions under which the model works best, specifies how the model is enacted in practice and delimits the evaluation's impact.

To organize my data analysis, I used the conceptual framework described in Chapter 2 and established a set of connections between the theoretical propositions of the A4GE derived using the PEMED and each of Miller's four criteria. To develop those connections, I considered each theoretical proposition of the A4GE as defined by the PEMED (see Section 2.2.1 in Chapter 2) and examined whether that proposition was most useful in providing me with: (a) guidance about how to act and behave throughout the various activities, (b) information as to when the model should or should not be applied, (c) information as to whether the model was applicable in practice, or (d) information as to whether the model has had an impact that can be attributed to it. From this, I established tentative explanatory connections between the theoretical propositions and the criteria of operational specificity, range of application, feasibility in practice, and discernible impact, respectively. The connections that I derived through this process are shown in Table 6.1. For example, the first theoretical proposition in the table logically and tentatively connects to the "range of application" of the A4GE and this is indicated by an "x" in that cell of the table.

Table 6.1 – Connections between the theoretical propositions of the A4GE and Miller's criteria

PEMED Dimensions	Theoretical Propositions of the A4GE	Operational specificity	Range of application	Feasibility in practice	Discernible impacts
Views about social programs	Program development is an evolving process that is predicated on stakeholders' experiences with the program		×		
	Programs are never ideal; they are at best the most sophisticated interpretation of how to enact a particular policy within the constraints of the cultural, physical and psychological contexts	×			
Paradigm of evaluation	The A4GE is interpretive and responsive	×			
	The A4GE stakeholders are partners during the whole evaluative process	×			
	The A4GE evaluator is an orchestrator of the evaluation processes	×			
Views about utilization	Process use: stakeholders involved in the evaluation learn about the evaluand and the evaluation processes				×
	Instrumental use: The A4GE must have an action orientation				×
Purpose of the evaluation	The A4GE aims at developing the constructions of the stakeholders about the program		×		
	The A4GE is best carried out during or after a program has been implemented		×		
Scope of the evaluation	The A4GE does not predetermine the focus and scope of an evaluation		×		
Operational procedures	Data collection in the A4GE is hermeneutic and dialectical			×	
	Data analysis in the A4GE is based on constant-comparison and yields the claims, concerns, issues and program improvement suggestions of stakeholders			×	
	Issues are negotiated in a virtual asynchronous negotiation forum and are either resolved or redefined			×	
	The A4GE evaluator is responsible for using interactive technologies to disseminate evaluation information			×	
	In the A4GE, discussions about theory and praxis are carried out separately			×	

6.1 Answering the research question using Miller's criteria

Using the connections in Table 6.1, I developed a response to the research question by analyzing how the theoretical propositions related to the criteria "operational specificity", "range of application", "feasibility in practice" and "discernible impact", respectively, were enacted in practice.

6.1.1 Criterion 1: Operational specificity of the A4GE

Miller (2010) describes operational specificity as the extent to which the theory underlying a program evaluation model provides guidance to the evaluator. Accordingly, to assess a model's operational specificity, evaluation researchers examine the extent to which the theoretical propositions of the model make them aware of alternative propositions and how their own proposition differs from these alternatives.

In what follows, I examine the four theoretical propositions of the A4GE that helped me to understand the model's operational specificity. These propositions are presented in Table 6.2. I discuss how these propositions were enacted in practice and, where relevant, whether and how they were modified in the context of practice. More specifically, for each of the four theoretical propositions, I answered three questions:

1. What does this proposition provide guidance about and what is the theoretical proposition advocated in the A4GE?
2. How and where did this proposition influence the decisions that I made as the A4GE evaluator?
3. Was there alignment or misalignment between the theoretical proposition and its practice? What were the observed differences, if any?

Table 6.2 – Propositions related to the operational specificity of the A4GE

PEMED Dimensions	Theoretical propositions of the A4GE	Congruence between theory and practice
Views about social programs	Programs are never ideal; they are at best the most sophisticated interpretation of how to enact a particular policy within the constraints of the context of implementation	Yes
Paradigm of evaluation	The A4GE is interpretive and responsive	Yes
	The A4GE evaluator is an orchestrator of the evaluation processes	Yes
	The A4GE stakeholders are partners during the whole evaluative process	No

6.1.1.1 Proposition 1 – Programs are never ideal; they are at best the most sophisticated interpretation of how to enact a particular policy within the constraints of the context of implementation

According to the A4GE, programs are never ideal as they are shaped by the contexts where they are implemented. Since these contexts are constantly changed by political and social influences, programs are in a state of endless flux. This proposition provides guidance to the A4GE evaluator about how to evaluate the program in relation to the context of implementation. Accordingly, the A4GE evaluator needs to understand the cultural, organizational and psychological contexts within which the program is embedded and, more importantly, how these contexts affect the program and judgments made about it.

Based on the data I collected throughout the evaluation, I found that this theoretical proposition was congruent with my implementation of the A4GE. Indeed, this proposition influenced many of the decisions that I made as the evaluator. For instance, prior to starting the evaluation, I carried out a detailed situational analysis of the context and assessed the appropriateness of the A4GE to the evaluation context. I also familiarized myself with the context to better understand the stakeholders' arguments, particularly when they were referring to the organizational context of the GDE(ST).

Additionally, this proposition made me aware that comments made about the GDE(ST) needed to be interpreted in relation to the contextual setting of the program. Therefore, in carrying out the evaluation, I did not place any particular emphasis on measuring adherence of the program to its written policies, nor did I encourage informants to make such statements. Instead, I urged stakeholders to present their CCIs about the GDE(ST) in relation to what can and should be done given the cultural, economic, organizational and psychological contexts. To that end, during the data collection and analysis, I was discussing, clarifying and reporting stakeholders' CCIs and suggestions as well as the contextual considerations related to them such as their practical limitations.

For instance, when lecturers expressed their views about the necessity of having a mandatory residential school for off campus student teachers, seven out of the eight participating lecturers expressed positive beliefs about the importance and multiple benefits of residential schools. Therefore, if I were to note the lecturers' positions about this matter without attending to their considerations of the contextual hindrances, I would have reported that the majority of them agreed with the necessity of having mandatory residential schools. However, because I was attentive to their explanation of the contextual factors, I noted that

five lecturers would not have encouraged having a residential school at the time, particularly because requiring attendance at residential school could inflict a noticeable economic burden, and the school would incur significant costs in delivering a residential school. UL2 (Interview 1) explained that "prospective candidates were reluctant to enroll in a program where attendance to a residential school was mandatory". UL1 (Interview 2) added that, given the minimal number of external students who expressed their willingness to attend the residential schools, there was "no compelling argument to justify the expenditures of the School of Education to run these residential schools". Therefore, while lecturers expressed claims about the usefulness of this component of the GDE(ST), they expressed concerns about its implementation in relation to contextual factors that discouraged the school from holding residential schools.

It is worthwhile mentioning that the view that programs are never ideal was also explicitly expressed by the majority of stakeholders (UL1, UL3, UL4, UL5, UL6, UL7, UL8, ST2, ST3, ST4 and ST5) who indicated that the GDE(ST) can only improve within the constraints of financial, logistical and human configurations. UL1 (Interview 1) explained that "while lecturers have to carry out research, be involved in duties within the School of Education and engage in personal professional development, their teaching efforts are essentially compromised, and this, in turn, affects the quality of the program".

Based on all of the above, I conclude that, with regards to this proposition, the practice of the A4GE was congruent with its theory.

6.1.1.2 Proposition 2 – The A4GE is interpretive and responsive

The A4GE is developed around the principles of the interpretive paradigm. Therefore, the model emphasizes social construction of knowledge as a primary means for collecting and analyzing data about programs. Furthermore, the A4GE emphasizes that knowledge construction is shaped by cultural and ideological backgrounds as well as the sophistication and openness of the participants. The A4GE is also theorized as being responsive, thus seeking multiple stakeholder views – including the evaluator's – and recognizing their contributions when developing the aims and questions of the A4GE. Moreover, in contrast to pre-ordinate evaluations, in which the criteria of the evaluation are predetermined by the evaluator and/or those who commission the evaluation, the A4GE evaluation criteria are derived through discussions with stakeholders.

According to this proposition, the A4GE evaluator must encourage stakeholders to disclose their previous experiences with the evaluand and to reflect on these experiences. Through analyzing personal stories, the A4GE evaluator, together with the participants, can then identify the participants' constructions about the program. The evaluator uses these constructions to formulate the questions for the interviews and to focus the scope of the evaluation.

I believe that this theoretical proposition was congruent with my implementation of the A4GE and has provided me with the required sensitivity while interacting with participants throughout the data collection, data analysis and reporting processes.

During the two rounds of interviews, I was asking participants about their stories and experiences with the program. To that end, I initiated all the interviews of the first round of data collection with this question: "Based on your experiences with the GDE(ST), what do you think are the strengths and weaknesses of the program?". I also asked for participants critiques of the arguments put forth by others. For me, it was not enough to only know their opinions about certain things, I consistently wanted to learn the reason for their opinion. Additionally, during each interview, I was careful to record even the smallest CCIs and suggestions that were presented by informants. I considered each piece of information as equally important and used it to develop a comprehensive list of CCIs. With more and more interviews, I kept expanding the list to represent all of the CCIs.

Accepting the interpretive nature of the A4GE, and realizing that informants were disclosing very personal information, I tried to build a trusting relationship with the participants and shared the evaluation data on an interactive website to maximize the transparency of the evaluation. When asked about the value of the website as a tool for information sharing, all respondents expressed positive attitudes and some emphasized its usefulness in providing them with necessary information to develop informed opinions about the issues raised. In that respect, UL5 (Interview 2) explained that it was "particularly useful to have a clearer insight into what others had expressed and the reasons for their arguments". UL8 (Interview 2) also noted that she felt she could "better relate to what others expressed" when she could access their stories and narratives.

Acknowledging the interpretive nature of the A4GE also made me aware of some limitations associated with this mode of inquiry. For instance, it was not possible to tell whether the informants were biased towards expressing more positive opinion about the program than

they truly believed. As explained by UL5 (Interview 2), some informants would become "defensive" if they felt that their job might be compromised by the evaluation. Therefore, during the data collection, I tried to clarify any suspected biases so that I could better portray the CCIs of the participants.

In line with the responsive nature of the A4GE, I developed the questions for the interviews based on informants' responses. For the first interview, I used open-ended questions to provoke the first interviewee's insights about the GDE(ST). In subsequent interviews, the questions were developed based on my analysis of the responses from previous interviewees. For instance, before I carried out the second interview in round one, I prepared a list of claims and concerns that were derived from the first interview. The list included (but was not limited to) the following claims: appropriateness of instructional strategies, usefulness of the assignments and quality of support for students. It also included the following concerns: lack of common vision of how the units articulate with the course, having optional residential school and lack of proper evaluations of the program. Using these items, I was able to develop probing questions to gauge the second interviewee's perspective on these matters. This approach enabled me to integrate, within each interview, the aspects that informants wanted to discuss. In that sense, the content of each interview was responsive to the evaluation needs of the stakeholders.

Furthermore, during the data collection, I asked informants who they thought should be included in the evaluation and who would be particularly able to answer their questions. Whenever possible, I integrated their suggestions within the evaluation processes. For example, I only included the Head of the School of Education after some participants suggested that his input was valuable to them. Nevertheless, it was not always possible for me to enact all suggestions. While some participants suggested including members from AITSL, and school principals and supervisors in the evaluation, due to time constraints and other logistic considerations, I refrained from involving such stakeholders in the evaluation.

In sum, I conclude that, with regards to Proposition 2, the theory and practice of the A4GE were congruent.

6.1.1.3 Proposition 3 – The A4GE evaluator is an orchestrator of the evaluation processes

Proposition 3 aims to increase the A4GE evaluator's awareness of his/her own role throughout the evaluation processes. The complexity of the roles of the A4GE evaluator is

linked to the notion of managing multiple stakeholders who bring competing interests and diverging views to the evaluation. In the A4GE, the evaluator is not only a participant in the evaluation, but also a mediator of a dialogue between stakeholders. During this mediation process, the A4GE evaluator is responsible for unveiling different stakeholder perspectives and presenting those to other stakeholders to assist the group in developing a shared view. This view about the evaluand translates into either a consensus about agreement or a consensus about disagreement. The A4GE evaluator is also a learner who seeks to comprehend the various perspectives and competing arguments presented during the evaluation. But he/she is also a teacher who educates stakeholders about each other's perspectives and about their roles in the evaluation.

In the reported case study, I assumed all these roles. First, I was a participant in the evaluation. To this end, I presented my arguments and opinions about the program in the same way as the participants. I also invited other stakeholders to comment and reflect on my perspectives.

Second, I mediated a dialogic process. During the interviews, I assimilated participants' conflicting views into a shared view about the program. For example, after each interview, I identified issues and noted them together with an explanation of the arguments presented on each side. In subsequent interviews, these issues were presented to the interviewees to stimulate their thoughts about them. I used these issues as venues for promoting professional dialogue with participants. Furthermore, through pointing out differences and similarities between stakeholders' points of view, I canvassed a shared view about the program. In order to make this process as smooth as possible, and to build a productive environment, I refrained from using a confrontational style. Instead, I acknowledged participant's perspectives and treated them as equally important. Additionally, acting as a mediator, I tried to promote a sense of direction to the dialogues. Thus, I asked participants to suggest solutions to the issues presented before them, specifically focusing on program improvement.

While it was relatively easy to mediate the dialogic process during the interviews, it was more challenging to do that in the online negotiation forum. Indeed, there were several instances where the negotiators would only comment once by responding to the questions that were presented by me, the evaluator. Although participants seemed to have incorporated the answers posted by others into their own, the forums did not generate extensive dialogues about issues. In these instances, I intervened and made some comments to encourage further dialogue, but my efforts were in vain. For example, I regularly included comments such as

"what do you all think?" or "do you agree with this?" where I invited participants to disclose whether they agreed or disagreed with some points raised in the forums.

A third role that I assumed in the reported case study was that of a learner. During the data collection phases, I developed my understanding of stakeholders' perspectives. To achieve that, I exercised active and reflective listening techniques. Hence, I attentively listened to what was being said, I developed appropriate follow-up questions to seek further clarifications and I used paraphrasing to ensure a shared understanding about the interviewee's views.

The fourth role I assumed in the evaluation case study was that of a teacher. At the time of recruitment, I explained the methods and goals of the evaluation to the participants in an invitation letter. Then, at the beginning of each interview, and with each participant, I again explained these processes to ensure that the evaluation was completely unambiguous. I also communicated the expectations I had concerning the roles that participants needed to assume. Therefore, I clarified that they needed to disclose their experiences and opinions about the program, and that they also needed to comment on other participants' input. I also explained their roles in relation to giving feedback about the evaluation model and making suggestions about how to improve the model. Additionally, during the data collection, I educated stakeholders about the points of view of others by exposing them to the arguments and statements developed by others, and sharing with them the experiences of others.

As part of being a teacher in the dialogic process, the A4GE evaluator is supposed to provide training for participants, if needed, so as to build their level of sophistication and to assist them in actively engaging in the evaluation. In this vein, I developed an interactive tool on the evaluation website where participants could learn how to extract data from the evaluation to obtain more information about other participants' opinions. Additionally, prior to the negotiation session, I developed a video tutorial to showcase the goals and processes of the online forum. Half way through the forum, I also sent an email to all participants to remind and encourage them to respond to each other's comments and arguments.

Based on all of the above, I conclude that with respect to this proposition, my implementation of the A4GE was congruent with its theory.

6.1.1.4 Proposition 4 – The A4GE stakeholders are partners during the whole evaluative process

According to the A4GE, stakeholders are not mere sources of information but partners in the entire evaluation process. This proposition aims to increase the evaluator's awareness about the roles of stakeholders in the evaluation processes and how to encourage stakeholders to realize and achieve those roles. The range of activities in which stakeholders need to be involved extends from designing the evaluation to selecting other informants, deciding on issues to discuss, contributing their own arguments and experiences about these issues, making suggestions regarding how the program can be improved, and, finally, enacting those suggestions in practice. Accordingly, the A4GE evaluator must actively optimize stakeholders' involvement in these activities.

While the informants in the reported case study were not seen as mere sources of information, their involvement was, nevertheless, limited to certain processes that were decided upon by me as the evaluator. Therefore, the proposition of the A4GE portraying stakeholders as partners in the evaluation process was not congruent with practice.

Regarding the degree of involvement of stakeholders in designing the evaluation, in the reported case study, it was not possible for me to negotiate the design of the evaluation with stakeholders. This was predominantly because I was a PhD student and did not have the resources or the time to do it. Nevertheless, throughout the data collection stage, I repeatedly asked participants for their feedback about the evaluation model so that I could capture their views about how evaluation needs to be carried out and what their expectations were in relation to the evaluation processes and outcomes. Based on their answers, I tried to accommodate as many elements in my design as possible. For instance, UL1 suggested that it would be better if the evaluator shares both the data (anonymous and de-identified) and summaries of data with participants to accommodate the diverse needs of participants. He explained that "some lecturers might feel compelled to read the summaries and only attend to the detailed data for the themes and issues that concern them the most" (Interview 1). Another instance where participants made suggestions about the design was when they suggested including the results from the formal student evaluations of the component units as part of the data. However, this was not possible to implement given the confidentiality of those evaluations.

Concerning the degree of involvement of stakeholders in contributing information to the evaluation, I tried to maximize informant input whenever possible. For instance, I explicitly

addressed and emphasized the role of the informants as partners in the program improvement process prior to their involvement through an information sheet. In that sheet, I clarified the centrality of their contributions. Furthermore, during each interview round, I made an explicit statement about the importance of their opinions and arguments in the progression of the evaluation. However, in accordance with the requirements of the university ethics committee, the information package clearly signaled that participants were free to withdraw their participation at any time. In the reported case study, five out of fourteen participants withdrew after the first round of interviews. Only one (ST1) explained that she had contributed everything she had and that she had nothing further to add. The others clarified that, while they realized the importance of their contributions, they were too busy to commit to the evaluation. In that respect, UL3 explained in an e-mail that, while she believed that she could be a great contributor in subsequent data collection, she had to withdraw as she could not "accommodate additional tasks" in her loaded schedule. Three more participants withdrew from the evaluation before the third round but did not indicate the reasons for their withdrawal. Therefore, in spite of my repeated explanations about the importance of their full participation during each contact session, some participants were not committed to the evaluation in a way that reflected authentic partnership.

Some of the processes in the evaluation were easier for me to engage participants as partners. These processes included nominating other stakeholders to take part in the A4GE, prioritizing which issues to discuss in the negotiation forum, and making suggestions about how to improve the GDE(ST). On the contrary, it was not possible for me, given that the study was a part of my PhD, to empower participants to enact any of the suggestions they made to improve the GDE(ST). I believe that my implementation of the A4GE as a PhD study lacked the authority to secure the commitment of the participants, particularly that it was conducted at a time when the GDE(ST) was being replaced by another program with a form and shape yet to be determined.

Based on the preceding discussion, I conclude that, concerning this theoretical proposition, there was no congruence between the theory and practice of the A4GE.

6.1.1.5 Summary for the criterion "operational specificity"

In relation to Miller's first criterion, evidence from my case study suggests that the theory of the A4GE as articulated by the PEMED dimensions and associated propositions was operationally specific and provided me with adequate guidance for approaching the

evaluation through understanding the context of the program, using an interpretive and responsive approach, and assuming multiple roles as the evaluator. Nevertheless, while the theory of the A4GE made me aware of the roles of other participants in the evaluation and alerted me about the need to achieve partnership with them, in the reported case study, I was not able to assist other participants in enacting their roles as partners in the evaluation to the degree specified in the theory of the model.

6.1.2 Criterion 2: Range of application of the A4GE

Miller (2010) describes the range of application of a program evaluation model as the extent to which theory describes the contextual circumstances under which the model is applicable and feasible. Accordingly, to describe a model's range of application, evaluation researchers examine the prescribed limits of the model's application as well as the conditions that enhance its successful application.

In what follows, I examine the four theoretical propositions of the A4GE, as defined by the PEMED, which helped me to understand the model's range of application and thus explore this criterion. These propositions are presented in Table 6.3. I discuss how these propositions were enacted in practice and, where relevant, whether and how they were modified in the context of practice. For each of the four theoretical propositions, I explored three questions:

1. What implications does this proposition have on the conditions in which the A4GE can and cannot be applied?
2. How were these conditions evident (or absent) in my case study?
3. Was there alignment or misalignment between the theoretical proposition and its practice? What were the observed differences, if any?

Table 6.3 – Propositions clarifying the range of application of the A4GE

PEMED Dimensions	Theoretical propositions of the A4GE	Congruence between theory and practice
Views about social programs	Program development is an evolving process that is predicated on stakeholders' experiences with the program	Yes
Purpose of the evaluation	The A4GE aims at developing the constructions of the stakeholders about the program	Yes
	The A4GE is best carried out during or after a program has been implemented	Partly
Scope of the evaluation	The A4GE does not predetermine the focus and scope of an evaluation	Yes

6.1.2.1 Proposition 1 – Program development is an evolving process that is predicated on stakeholders' experiences with the program

According to the A4GE, the development of a program is an ongoing process that is primarily based on the experiences of stakeholders who interact with the program. This development is incremental by nature and is better achieved through stakeholders' participation in the process. Thus, if the A4GE is to be implemented successfully, stakeholders involved in the evaluation should share this view about program development and recognize the importance of their role in promoting that process. Furthermore, the success of the A4GE depends on the extent to which stakeholders are willing to disclose and debate their private experiences.

In the reported case study, both of these conditions were met to a satisfactory degree. For instance, the view that a program is best developed through the participation of stakeholders was, reportedly, shared by participants. When asked their opinions about how the GDE(ST) should be monitored, both lecturers and teachers suggested that the monitoring should engage all those who have interacted with the program, ranging from student teachers to the Head of School. In that respect, ST2, ST4, ST5, UL1, UL3, UL4 and UL6 pointed out the usefulness of the A4GE as a method for promoting effective engagement of stakeholders with program development. For example, UL3 (Interview 1) explained that the most important characteristic of the A4GE was its participatory nature as it "promotes multiple views from people who have actually had some experience with the program". Similarly, other participants explained that the evaluation model is a "venue for professional dialogue" (ST5, Interview 1) and a "place for sharing professional expertise" (UL6, Interview 1).

In the reported case study, most participants expressed their enthusiasm to share their opinions and experiences related to the GDE(ST). For example, ST4 (Interview 1) explained that he "strongly believe[s] in the potential of this study in bringing about meaningful changes to some of the program's components such as the PREX" and that he was pleased to contribute back to the university he so loved. ST2 (Interview 1) made a comparison between this type of evaluation and the exit survey questionnaire he received from the university after he completed his degree, and noted that he felt more confident voicing his opinions in front of a real person, knowing that the person will actually listen to him and discuss those opinions with him. Similarly, lecturers expressed enthusiasm about sharing their experiences with others. Two lecturers said that this type of evaluation encourages open dialogue, and better discourses can emerge when people talk about things that most matter to them. Three other lecturers expressed that they would be willing to change their practices to accommodate

the suggestions made by teachers and colleagues. One lecturer (UL2) proposed that more data would be needed for practices to change, and the data should be obtained by a variety of means, including evaluations that are similar to the A4GE.

Based on these findings, I argue that the theory and practice of the A4GE were congruent in relation to this proposition.

6.1.2.2 Proposition 2 – The A4GE aims at developing the constructions of the stakeholders about the program

The purpose of the A4GE is to generate the constructions of the stakeholders in the form of lists of CCIs and suggestions about the program. The evaluation, thus, does not answer predetermined questions, nor does it focus on predetermined issues or goals. Instead, the only goal of the A4GE is to elicit the CCIs and suggestions of participants and to develop their formulations through dialogue and negotiation. The implication from this proposition is that if the A4GE is to be implemented successfully the participants in the evaluation need to accept that the outcomes of the A4GE are lists of CCIs and suggestions, and that stakeholders' consensus about issues and their resolutions is an unlikely outcome of the A4GE. Therefore, participants need to be willing to accept agreements as well as differences and to tolerate the persistence of unresolved issues.

In the reported case study, this condition was satisfied. Teachers and lecturers alike expressed positive attitudes towards the goals of the A4GE. For instance, ST3, ST4, UL5 and UL6 explained that the most important outcome from the evaluation was the mutual learning it generates by uncovering insights from those who have had experience with the program. UL3 (Interview 1) added that "through revealing claims, concerns and issues, some of which are not usually discussed in formal settings or when using other models of evaluation, the A4GE can improve the way lecturers teach and reflect about their teaching". Furthermore, participants in the case study did not reportedly care about consensus and issue resolution as much as they cared about learning about the numerous CCIs in ways that enabled them to comprehend them and act upon such understanding. While most participants anticipated that consensus about issues and their resolution was an unlikely outcome, many argued that it was not as important as the discussions that generated issues. In that respect, ST4 (Interview 2) explained: "I don't think an agreement is going to emerge from our discussions. I don't think it is rational to expect that we will all agree on what needs to be changed and how … I think what matters the most at this stage is that we are all discussing the [GDE(S)] and contributing ideas that can form a basis for future improvement".

With regards to this proposition, I conclude that the findings from my case study suggest that the theory and practice of the A4GE were congruent.

6.1.2.3 Proposition 3 – The A4GE is best carried out during or after a program has been implemented

Proposition 3 suggests that the A4GE is carried out best during or after the implementation of the program, that is when the stakeholders involved in the evaluation have had the time to interact with the program and to develop their experience repertoires. As such, the A4GE is used as a monitoring process to improve the program processes. The implication of this proposition is that the success of the A4GE depends on the extent to which participants perceive the A4GE as a process for bringing about insights about the program and not as a tool to measure the program's impact.

In the reported case study, the A4GE was carried out several years after the program was implemented. Interestingly, though, during the evaluation, a decision was made by the School of Education to terminate the program and replace it with another one. All lecturers were aware that the program was going to be replaced by a two-year equivalent program in compliance with new AITSL mandates. I assume that teachers were not aware of this decision since none of them commented on the subject.

When asked about their opinions regarding the timing of the evaluation, participants expressed two points of view. One group comprised lecturers who emphasized the importance of reaping the lessons learned from the implementation of the GDE(ST) which could then be used to avoid potential challenges while implementing the new program. This view was particularly shared by science teachers, who believed it would be beneficial to use the lessons learned from implementation of the GDE(ST) to improve it (not knowing that the program was to be terminated). ST5 (Interview 1) stated that "this evaluation is really important because it captures our experiences with the [GDE(ST)] … I think it would make a great tool for reflective practice, particularly for the lecturers". The second group comprised two lecturers (UL2 & UL7) who questioned the utility of the evaluation, since the program was going to be terminated. For example, UL2 (Interview 1) said: "I think this evaluation would have been more useful three or five years ago … but now, I mean the program is going, it is being replaced, so it is not as useful". These lecturers were not involved in the second round of data collection. While this belief might have contributed to their decision to withdraw from the evaluation, I cannot establish a causal link between the two events,

particularly that two teachers and one lecturer from the first group also withdrew at that point without commenting on the reasons for their withdrawal.

Concerning this theoretical proposition, I conclude that there was only partial congruence between the theory and practice of the A4GE.

6.1.2.4 Proposition 4 – The A4GE does not predetermine the focus and scope of an evaluation

Proposition 4 suggests that the focus of the A4GE and its scope are not determined *a priori*, but rather through an emerging design as the evaluation proceeds. Therefore, the A4GE is best applied in a context where participants accept the flexibility of the evaluation design. Also, participants should accept that they are as responsible as the evaluator for shaping the focus of the evaluation through intensive engagement in the evaluation.

Both conditions were met in the reported case study, though the latter was satisfied to a lesser degree. As previously outlined in my discussion of Proposition 2, participants in my study were positive regarding being able to raise their own CCIs about the program. UL1 (Interview 2) noted that the A4GE made him think "outside the box" about the things that mattered the most. UL5 (Interview 2) added that the flexible boundaries of the A4GE "generated innovations in terms of the content of the discussions and the way these were debated". While drawing on comparisons with alternative evaluations that have predetermined focus, one student (ST5, Interview 2) made a comment about how she "hated" to do the evaluation of units at the university because there was limited room for describing the things that mattered to her the most, and no room at all to discuss them. She explained that, through her engagement in the A4GE, she felt empowered because she had the "authority to discuss" her own challenges and concerns and be heard.

Only half the participants in the reported case study accepted the responsibility of shaping the focus of the evaluation. While I tried to clarify, at the beginning of each round of data collection, the importance of participants' contributions to the development and refinement of the focus of the evaluation, some participants withdrew as the evaluation progressed. One participant (ST1, Interview 1) explained that she had made her contributions and did not see the point in refining the evaluation any further. Other participants withdrew but acknowledged the importance of their contributions in the second phase. The remaining participants expressed a keen interest in committing to the evaluation and contributing further

to shaping the discussions. For instance, UL6 (Interview 2) stated: “I am looking forward to the negotiation stage to see how the whole process will unfold”.

I conclude that, with regards to this proposition, the theory and practice of the A4GE were congruent.

6.1.2.5 Summary for the criterion “Range of application”

Based on my case study, I argue that there was considerable congruence between the theoretical propositions concerning the A4GE’s range of application and its practice. The A4GE offered an elaborate description of the conditions required for the successful implementation of the model. These conditions were significantly met in my study Participants shared the view that their personal experiences and those of other stakeholders were essential to the evaluation and development of the program. Some stakeholders disclosed their willingness to change their practices to accommodate suggestions made in the evaluation. Furthermore, participants expressed positive attitudes towards the goals of the A4GE. Concerning the timing of the evaluation, participants expressed divergent views regarding the benefits of carrying out the A4GE. While some emphasized the importance of developing lessons from the implementation of the GDE(ST), others argued that the evaluation was redundant. Lastly, only half the participants demonstrated responsibility in focusing the evaluation, with the other half, while acknowledging the importance of their contributions, did not continue their participation in the evaluation.

6.1.3 Criterion 3: Feasibility in practice of the A4GE

Miller (2010) describes the feasibility in practice of a program evaluation model as the extent to which theory can be applied in practice. Accordingly, to describe a model’s feasibility in practice, evaluation researchers need to examine aspects, such as the design quality, effectiveness, efficiency and robustness of a model, in addition to the model’s side effects.

In what follows, I examine the five theoretical propositions of the A4GE as defined by the PEMED that helped me to understand the model’s feasibility in practice. These propositions are presented in Table 6.4. I discuss how these propositions were enacted and whether and how they were modified in practice. For each of the five theoretical propositions, I answered three questions:

1. What does this proposition imply about the procedures that need to be made by the evaluator?
2. How did I implement those steps in my case study?

3. Was there alignment or misalignment between the theoretical proposition and its practice? What were the observed differences, if any?

Table 6.4 – Propositions clarifying the feasibility in practice of the A4GE

PEMED dimensions	Theoretical propositions of the A4GE	Congruence between theory and practice
Operational procedures	In the A4GE, data collection during the interviews is hermeneutic and dialectical	Partly
	Data analysis in the A4GE is based on constant-comparison and yields the claims, concerns, issues and program improvement suggestions of stakeholders	Yes
	Issues are negotiated in a virtual asynchronous negotiation forum and are either resolved or redefined	No
	The A4GE evaluator is responsible for using interactive technologies to disseminate evaluation information	Yes
	In the A4GE, discussions about theory and praxis are carried out separately	Partly

6.1.3.1 Proposition 1 – in the A4GE, data collection during the interviews is hermeneutic and dialectical

In the A4GE, data are gathered from the interviews through a hermeneutic dialectical process involving multiple stakeholders. The methodology of the A4GE is hermeneutic because the evaluator essentially seeks to explain and interpret the ideas of stakeholders involved in the evaluation. It is dialectic because the evaluator compares and contrasts divergent views to construct, with other stakeholders, a more sophisticated shared view. In the A4GE, the evaluator uses the steps described by Guba and Lincoln (1989) to carry out the hermeneutic dialectic cycle twice, once for each interview round.

In the reported case study, I was able to apply the dialectic hermeneutic cycle described by Guba and Lincoln (1989) twice, once in each interview round. However, due to time constraints, I introduced a focusing process prior to the second cycle where I asked participants to nominate fewer issues to discuss in the second interview round.

In the first round of interviews, the first interview was highly unstructured. My approach was to encourage the interviewee to share his experience with me about the program and to highlight the aspects of the program that he thought were particularly good and others that needed improvement. During the interview, I asked the interviewee to share personal stories

and give specific examples to illustrate his point of view. My purpose was to elicit his claims and concerns. The second interview was more structured, informed by the preliminary analyses from the first interview. I asked the second interviewee about his claims and concerns concerning the GDE(ST). Additionally, to meet the dialectic intent of the A4GE, I invited him to comment about the claims and concerns of the first interviewee. In particular, I used probing questions to compare and contrast his views with those of the first interviewee. Therefore I asked: "what do you think about this aspect of the program?" and followed up that question with "Participants X and Y said this and that, what do you think about that? Do you agree?"

I continued in this fashion with the remaining interviewees until the first round of data collection was complete. Using the dialectic approach for comparing and contrasting claims and concerns, I was able to identify issues between members of each stakeholder group as well as issues that existed between stakeholder groups.

In the second round of interviews, I wanted to achieve two objectives: I wanted to clarify further the CCIs of the participants and, more importantly, I wanted participants to nominate issues which they wanted to discuss in the negotiation forum. In the first interview of the second round, I trialed the hermeneutic dialectic cycle advocated by Guba and Lincoln (1989). Therefore, I asked the first interviewee to comment on all of the 41 topics that emerged from the first round and which I classified under claims, concerns or issues. The interview took two hours and forty-five minutes. Therefore, I modified my approach because I realized that most participants were not willing to commit that much time. I decided to ask participants to nominate up to ten CCIs that were particularly meaningful to them prior to the interview. Then, during the interview, I used the hermeneutic dialectic cycle to discuss the CCIs that were nominated.

Based on these findings, I conclude that the theory and practice of the A4GE were only partially congruent in relation to this proposition.

6.1.3.2 Proposition 2 – Data analysis in the A4GE is based on constant-comparison and yields the claims, concerns, issues and program improvement suggestions of stakeholders.

In the A4GE, data are analyzed using constant comparison. This technique of data analysis requires the evaluator to code the statements made by the interviewees. The evaluator examines each statement and compares it to other statements that are either similar or different. Each statement is then assigned two codes: the first indicates the content of the

statement and the other indicates its nature (i.e., a claim, concern or suggestion). The codes are inductively developed and, therefore, undergo definition changes as more statements are examined and coded over the course of the analytical process.

Based on my application of the A4GE, I argue that the model is specific in relation to how the evaluator derives the CCIs and suggestions for improvement from the data, and that its methodology was feasible in practice. Furthermore, using NVivo to perform the coding and data analysis was both feasible and effective.

While the A4GE does not offer guidance about how to use computer-assisted data analysis software to analyze data, I argue that, based on my experience with the A4GE and the analytical procedure described in Chapter 4, the analyses that NVivo permits (or other computer assisted data analysis software for that purpose) are difficult if not impossible to complete by hand. Given the difficulty of completing the required analyses to the depth required by the model's formulation, I argue that the use of computer-assisted software presents the following advantages:

- Developing the codes inductively and changing their names easily without having to change those names on paper
- Merging codes while carrying out thematic coding: the data assigned to these codes will merge automatically
- Ease of access to the data
- The luxury of working on one file and one computer
- Possibility to compare across group characteristics, such as gender, years of experience, level of educational attainment, and so forth.

Based on my experience in the reported case study, I argue that the theory and practice of the A4GE were congruent in relation to this proposition.

6.1.3.3 Proposition 3 – Issues are negotiated in a virtual asynchronous negotiation forum and are either resolved or redefined

In the A4GE, the agenda for negotiation comprises the issues that were nominated by stakeholders during the second round of interviews. These issues are then negotiated in a virtual forum. The medium for negotiations is virtual, because the evaluator seeks to conceal the identities of the negotiators in an effort to eliminate power differentials that sometimes dominate face-to-face negotiations. The mode of communication is asynchronous to give

participants time to think through their comments and contributions. In the negotiations, participants present their arguments and counterarguments until a consensus is reached, or until the time and resources available for the negotiations are exhausted. Importantly, in the A4GE, consensus could be either a consensus on agreement or a consensus on disagreement.

The 4GE was specific in relation to how I should develop the negotiation agenda and carry out the negotiations. However, given that this proposition contains modifications to the original formulation of the 4GE, it did not specify the steps required to develop a virtual medium for the negotiations. In Chapter 4, I described how I enacted this proposition in practice.

In my case study, none of the teachers were involved in the negotiations even though three initially indicated their willingness to. The lecturers, however, were not told that this was the case. Given this situation, it is not possible for me to make a comment as to whether concealing the identities of stakeholders in my study had eliminated or reduced power differentials between stakeholder groups in the negotiations.

Concerning my role as a mediator of the negotiation process, my main tasks included clarifying the points of view by making summaries and drawing connections between the comments made by the various negotiators, and encouraging dialogue by developing probing questions and inviting reflections about what was being said. My interventions in the negotiations were to enable consensus among all negotiators about agreement or disagreement. That is, my purpose was to achieve a shared understanding among participants about the issues either by having participants agree on the resolution of issues or agree on the existence issues and their non-resolution. To this end, I made regular interventions that summarized the points made in each forum and, where appropriate, I use direct questions and asked participants whether they agreed or disagreed with the ideas.

Based on the data from my case study, I found that consensus of any form (that is, on agreement or disagreement) was not achievable. First, only lecturers were involved in this part of the evaluation. This automatically meant that no consensus could be achieved since the evaluations excluded a major stakeholder group. In the report of the evaluation case study, I mentioned that any reported consensus was biased, being among lecturers and myself as the evaluator. Therefore, I cautioned about the significance of any reported consensus.

Second, most lecturers presented only one comment on the questions and proposals. This observation could indicate that the remaining participants felt accountable for responding to

the questions and proposals but not to engage in debates. However, it is noteworthy that when lecturers made comments, they integrated the comments made by others on the forum into their own responses. I conclude that, while there has been some interaction between the stakeholders on the forum, it was minimal and consequently not congruent with the requirement of intensive participation formulated in the theory of the A4GE.

A related worthwhile observation about the forums is that more than half the questions and proposals (that is 17 out of 31) were not discussed. This ratio can be explained in relation to the large number of questions and proposals made. I believe that, given the time constraints of most participants, better negotiations could have occurred if the number of issues and their related questions and proposals, had been smaller.

Based on the data presented above, I argue that negotiations of issues did not occur in my case study. What happened instead was a continuation of the hermeneutic dialectic process that started in the interview rounds. Negotiators were responding to the prompts presented before them in the forum and were integrating other participants' responses. However, there was no sense of direction in the forum that indicated whether agreement was being formed or disagreement was being consolidated.

Based on all of the above, I argue that the theory and practice of the A4GE were not congruent in relation to this proposition.

6.1.3.4 Proposition 4 – The A4GE evaluator is responsible for using interactive technologies to disseminate evaluation information

In the A4GE, the evaluator should design, develop and use appropriate interactive tools to disseminate the findings of the evaluation in a timely manner. This step is essential to enhance the engagement of the participants and to promote the transparency of the evaluation. Since this proposition is based on the adaptations introduced to the 4GE, the A4GE does not specify the steps needed to achieve this. However, in the reported case study, I was able to enact this proposition by developing an interactive website for sharing the data and emerging findings from the evaluation. In Chapter 4 and Appendix A, I described how I enacted this proposition in practice.

Participants in the study expressed positive attitudes about the use of the interactive website as a means for sharing the data. Some found it particularly useful in presenting a large amount of information in a compact yet meaningful way. For instance, UL1 and UL5 explained that one asset of the website is that the user can choose what to look at and extract

the information he/she needs without having to go through pages of interview data. ST2 (Interview 2) added that while it took him a little bit of time to get used to the website, he found it "particularly useful to be able to look at the data selectively yet examine what has been exactly said, and not through the lens of analyses carried out by someone else". Another participant explained that the structure of the website enabled him to make comparisons between the stakeholder groups and draw useful conclusions: "I liked the idea of being able to separate between science teachers and lecturers and see how the comments were different" (ST3, Interview 2).

Based on these findings, I argue that the theory and practice of the A4GE were congruent in relation to this proposition.

6.1.3.5 Proposition 5 – In the A4GE, discussions about theory and praxis are carried out separately

In evaluations, discussions about theory and praxis can be easily confused. In the A4GE, these two types of discourses should be explicitly distinguished. In the interviews as well as in the negotiation forum, discussions addressing theoretical aspects of an issue should be distinguished from those focused on what could be done in practice to resolve that issue.

This proposition was an amendment to the original formulation of the 4GE where the assumption was that action and theory are both discussed at the same time. Therefore, I cannot make a statement as to whether the A4GE was specific in how practitioners can enact this proposition. Instead, I describe how I enacted this proposition in practice.

In the reported evaluation, I made a conscious effort to distinguish between theory and practice and I encouraged participant to make that distinction as well. During the interviews, I used probing questions to raise the participants' awareness about the distinction and to elicit their responses about theory and praxis separately. For instance, when discussing a particular issue, I would ask them for their point of view about why they think there is an issue and how it affects them. Then I would ask them what could be done to address the issue in practice. While I made an effort to draw these distinctions, I was also keen to minimize my interruptions and allow participants to express themselves freely and openly. As a result, I identified several instances where participants would be integrating their discussions of issues with suggestions about their resolution. This observation was more common among teachers than it was among lecturers. For example, when asked to discuss the issue of scarcity of resources, ST2 (Interview 2) suggested about what should be provided for him as a student

teacher rather than explaining why these resources would make him better prepared. Similarly, when asked about why the quality of instruction was a concern for ST2 and ST6, both teachers replied by providing immediate suggestions about how they would change existing practices.

On the online forum, the task was simplified because the distinction was more pronounced. The structure of the negotiation forums clearly indicated a separation of discussions about the theoretical nature of the issue and the practical aspect related to the issue and its resolution. I also emphasized this dichotomy further in the video tutorial sent to participants prior to the negotiations. As a consequence, it was easier for participants to make that distinction between theory and praxis and this was evident in their responses. Indeed, I identified only one instance (Forum 1, Negotiator 2) where a participant had embedded a comment about practice while contributing to the theoretically oriented discussions.

Based on these findings, I argue that the theory and practice of the A4GE were partially congruent in relation to this proposition.

6.1.3.6 Summary for the criterion "Feasibility in practice"

Based on the data from my case study, I argue that many of the processes described in the A4GE were feasible in practice. For instance, I was able to apply the dialectic hermeneutic cycle described by Guba and Lincoln (1989). I was also able to use constant comparison as described in the A4GE to derive the CCIs and suggestions for improvement from the data. In addition, I designed and used an interactive website to share the findings of the evaluation in a timely manner, and developed a virtual asynchronous negotiation forum where I carried out the negotiations.

However, the findings also revealed that other processes were more difficult to implement and required modifications, such as the focusing mechanism prior to the second interview round. Furthermore, I was only partially able to differentiate the discussions about theory from discussions about praxis. Finally, in this case study, the process of negotiation was not feasible in practice and no consensus of any form was achievable.

6.1.4 Criterion 4: Discernible impact of the A4GE

Miller (2010) describes the discernible impact of a program evaluation model as the extent to which theoretically conceptualized impacts are, in fact, happening in practice. Accordingly, to describe a model's discernible impact, evaluation researchers need to examine whether and to what extent the model actually generates impact.

In what follows, I examine the two theoretical propositions of the A4GE that helped me in understanding the model's discernible impact and thus explore the fourth criterion. These propositions are presented in Table 6.5. I discuss how these propositions were enacted in practice and, where relevant, whether and how they were modified in the context of practice. Specifically, for each of the two theoretical propositions, I answered three questions:

1. What does this proposition imply about the expected outcomes of the A4GE? How are these outcomes recognizable?
2. Was I able to capture those outcomes in my case study?
3. Was there alignment or misalignment between the theoretical proposition and its practice? What were the observed differences, if any?

Table 6.5 – Propositions clarifying the discernible impact of the A4GE

PEMED dimensions	Theoretical propositions of the A4GE	Congruence between theory and practice
Views about utilization	Process use: stakeholders involved in the evaluation learn about the evaluand and the evaluation processes	Partly
	Instrumental use: The A4GE must have an action orientation	Partly

6.1.4.1 Proposition 1 – Process use: Stakeholders involved in the evaluation learn about the evaluand and the evaluation processes

Although the term "process use" only appeared in 1997 (see, Patton, 2008), that is eight years after the 4GE was developed, it is implicitly embedded within the formulation of the model. The 4GE and, hence, A4GE specify that stakeholders involved in the evaluation will not only learn about the evaluand through the CCIs and improvement suggestions but will also gain insights into the evaluation experience through engaging in the processes of inquiry and discovery.

Process use of the A4GE can be recognized through observed or reported changes in participants' intentions and/or actions. These changes, however, do not necessarily appear immediately and may only become manifest in the long run, when the challenge of determining whether process use occurred or not. Therefore, and particularly for long-term changes, it is not always possible to determine whether the observed or reported changes are discernible impacts that can be attributed to the evaluation. Nevertheless, it is still likely that

a researcher/evaluator can recognize more immediate changes that surface during the evaluation.

In the reported case study, I was able to identify some instances where immediate process use occurred. For example, participants reportedly learned about the GDE(ST) and about the various perspectives of their fellows in the evaluation. UL5 (Interview 2) explained that the evaluation helped her "appreciate the perspectives of other colleagues regarding some issues which [she] didn't even realize were problematic". UL6 (Interview 2) also pointed out that the evaluation was particularly helpful in pointing out challenges that lecturers and teachers encounter, and which are usually not discussed to the extent that they deserve. Science teachers also shared their views about the learning they derived from the evaluation. ST4 (Interview 2) argued that without knowing what the concerns of lecturers are, it is likely that a student candidate in the GDE(ST) makes uninformed statements undermining the program. He added: "through engaging in the evaluation, I learned about what it truly means to put yourself in someone else's shoes and really understand where [lecturers] come from and what they can actually do".

Furthermore, participants in the evaluation described other learning they experienced through engaging in the processes of collaborative inquiry and discovery of the A4GE. UL1 (Interview 2) for example, explained that the A4GE was particularly useful for helping people "think outside the box" and come up with innovative solutions that are accepted by others. Another lecturer expressed how engaging in "professional dialogues" (UL7, Interview 2) expanded her horizons and made her appreciate further the importance of reflection in improving the quality of the discourse about the GDE(ST). UL8 stated that the A4GE generated an interest among participants to carry out discussions at the level of the program, which she thought was "beneficial and enlightening" (Interview 2).

The previous examples of process use are based on the reported narratives from the participants about the learning they derived from being involved in the A4GE. However, there are other observations that I made during the evaluation and which indicate occurrences of process use. Throughout the second and third rounds of data collection, I noticed that participants were becoming more inquisitive about the CCIs and suggestions. They were asking more questions and requiring further clarifications. For example, lecturers were asking for additional information, such as "what kind of support the student teachers needed in their PREX?" (UL8, Interview 2), or "what kind resources did student teachers find to be missing from the units?" (UL4, Interview 2). These activities are indicative of a learning process

whereby participants seek more information to clarify and define their understandings. Additionally, during these two rounds, participants were referring to the data from previous interviews, which they could access through the website, and were comparing and contrasting their opinions with those of others. For instance, ST2 stated: "I think that the website is all very well structured and I found that it was easy to see my opinions and also to compare with others ... I found, though, which is interesting, that the largest differentiation between views was between students and lecturers. The lecturers had one opinion, which I think is really defensive, and the students had ones what were similar to mine" (Interview 2). These analytical actions also indicate that participants are interacting with the evaluation data to generate learning and are, thus, considered as indicators of process use.

Based on these findings, I argue that the theory and practice of the A4GE were at least partially congruent in relation to this proposition.

6.1.4.2 Proposition 2 – Instrumental use: The A4GE must have an action orientation

The A4GE posits that the evaluation should yield a course of action to be followed, and the evaluator should stimulate stakeholders to follow it and generate and preserve their commitment to do so.

While the outcomes of the A4GE included suggestions for improving the program, in the reported case study, I did not stimulate stakeholders to enact these suggestions. Since the evaluation was part of my PhD study, I was bound by time and resource constraints. Additionally, I did not feel sufficiently empowered to initiate change, especially since I was already struggling to simply keep participants motivated to continue their participation in the evaluation. In sum, I did not feel that the evaluation, in that context, had the critical weight to drive change.

Nevertheless, I was able to report some occasions where the findings from the evaluation generated participants' reactions concerning their current practices. For instance, building on the suggestions regarding establishing partnerships between the supervising teachers in schools and the lecturers, UL1 (Interview 2) suggested that he was going to prepare a small handbook to explain to the supervising teacher what his expectations were concerning student teachers' learning during PREX. Similarly, two other lecturers (UL4 and UL8) reported that they were going to adjust components of their assignments to respond to the concerns expressed by teachers. UL8 (Interview 2) explained that she wanted to make the topic of the

assignment “more flexible in accommodating the diverse backgrounds of student teachers, and more responsive” to their declared needs.

While these instances are potential examples of occurrence of instrumental use, I cannot confirm whether these actions were actually implemented because I did not monitor them further. Therefore, I argue that the theory and practice of the A4GE were only partially congruent in relation to this proposition.

6.1.4.3 Summary for the criterion “discernible impact”

Examining the discernible impact of the A4GE requires extensive follow-up. Therefore, in the reported case study, I only focused on the self-reported impacts and the few observations that I was able to make during the evaluation. Based on the data derived from my case study, I suggest that process use did occur, as evidenced by the variety of learning that the evaluation engendered. Participants reported learning more than the outcomes of the evaluation, that is, the CCIs and suggestions. They also learned about the importance of engaging in professional dialogues and reflective discourses where they can share experiences. Nevertheless, it was not possible for me to provide evidence that instrumental use occurred beyond the few descriptions reported by participants.

6.2 Answering the general research question

The general research question in this study was:

> **How congruent are the underlying theory and practice of an adapted version of the 4GE model in the context of evaluating a secondary science teacher preparation program in an Australian university?**

The findings from my study suggest that there is considerable congruence between the theoretical propositions of the A4GE and its practice.

Concerning the operational specificity of the model, I argue that the A4GE offered me sufficient guidance concerning how to conduct the evaluation through the use of an interpretive, responsive and context-sensitive approach. The A4GE also raised my awareness about the roles that I and other participants needed to assume. Nevertheless, in practice, I noticed that it was challenging for me to assist participants in enacting their role as partners in the evaluation as it was difficult to maintain their commitment and pro-activeness.

Concerning the range of application of the A4GE, I found that the theory underlying the model’s formulation clarified necessary conditions where the A4GE can be implemented. In

my case study, these conditions were met to a considerable extent and there was significant congruence between the theoretical propositions concerning the A4GE's range of application and its practice. Indeed, participants shared the view that their personal experiences and those of other stakeholders were essential to the evaluation and development of the program. Additionally, they expressed positive attitudes towards the goals of the A4GE and explained that consensus was not as important as the CCIs and suggestions. Furthermore, while more than half the participants withdrew from the evaluation, most acknowledged the centrality of their contributions to shaping the focus of the evaluation. The remaining participants demonstrated their responsibility in this process by continuing their engagement until the end of the evaluation even though their engagement was not as intensive as theorized in the A4GE.

With respect to the timing of the evaluation, participants in my study expressed two opposing views regarding the benefits of carrying out the A4GE. While some emphasized the importance of learning from the evaluation to inform the development of the new program, others found the evaluation pointless given that the program under evaluation was about to end.

Concerning the feasibility in practice of the A4GE, I found that many of the processes described in the A4GE were feasible in my case study. Hence, I was able to apply the dialectic hermeneutic cycle, use constant comparison to derive the CCIs and suggestions for improvement from the data, design and use an interactive website to share the findings of the evaluation in a timely manner, and develop a virtual asynchronous negotiation forum to carry out the negotiations. However, I found that some other processes needed modification before they were applied in practice. For instance, prior to the second application of the hermeneutic dialectical cycle, I introduced a focusing mechanism whereby participants nominated up to ten issues to discuss during the second interview round. My findings also suggested that differentiating the discussions about theory from discussions about praxis was only feasible in the negotiation session and not in the interviews.

Importantly, the findings from my case study revealed that the negotiation process, as conceptualized in the A4GE, was not feasible in practice. Instead, my observations of interactions during that stage suggest that negotiations were replaced by hermeneutic dialectic discourses in which participants were responding to the prompts (questions and proposals) on the forum by sharing their own stories and referring to those of others.

Concerning the discernible impact of the A4GE, I found evidence of numerous occurrences of process use as demonstrated by the variety of learning that the evaluation engendered. Participants in my study reported learning about both the program and the evaluation processes. Nevertheless, I was not able to provide substantial evidence that instrumental use occurred beyond the descriptions described by participants.

In order to give a more comprehensive understanding for the relationship between the theory and practice of the A4GE, I identified two contextual factors that affected my practice of the A4GE and hampered my adherence to its theoretical formulation. In the following, I explore these factors and discuss where and how they supported or challenged me in enacting the A4GE.

6.2.1 Factor 1 – Evaluation resources

In the reported case study of the A4GE, all of the processes of the evaluation were carried out within limitations of time and availability of human resources. The application of the A4GE was part of my PhD study and took about fourteen months to complete. During that period, I recruited participants, carried out data collection and analysis, designed a website and an online forum, and developed a case study report. Being a single evaluator, I had to carry out all these processes by myself. Whenever possible, I sought professional assistance to complete the tasks efficiently and on time. For example, I used the services of a transcription company to assist me with the transcriptions of interviews. Each interview was de-identified and then sent to an overseas institution where it was transcribed within three to four days. This enabled me to carry out an analysis of the interview before conducting the following interview. This process was crucial since the A4GE is essentially responsive. I also employed a web developer to assist me with the development of the website and the online forum. However, the lack of assistance in carrying out the data collection and analyses imposed limitations on the sample size, which was restricted to what would be manageable for the evaluation; using additional, trained evaluators helping me in the task would have enabled me to increase the number of participants, thus enriching the findings.

Given the time constraints, I had to make compromises in some of the processes of the A4GE. For instance, I was not able to engage stakeholders in the design of the evaluation or in the data analysis. Whereas the theory of the A4GE suggests that participants should be partners in all processes, I was bound by time and could not enact this proposition. Markiewicz (2005) argues that involving stakeholders in the planning of evaluations

enhances the likelihood that evaluation findings are used. In my study, I identified instances of process use but could not ascertain that instrumental use occurred. While my findings align with the argument put forth by Markiewicz (2005), I do not claim a relationship between evaluation use and stakeholders' engagement in evaluation design, because the design of my study does not allow for such conclusions.

Another compromise to the processes of the A4GE was my decision to limit the number of issues discussed in the second round of interviews to the relatively few that were nominated by the interviewees. In that respect, I was bound by the amount of time that other participants were willing to invest in the evaluation. During the recruitment process, most participants indicated that they would be happy to contribute to the evaluation as long as it would not take more time than indicated in the information package sent during the recruitment process, that is, around 2.5 hours for the entire evaluation.

Time constraints also affected the quality of interaction that occurred in the negotiation forum. While the discourse was hermeneutic and dialectical in the negotiation forum, it was not a negotiation process. Given the large number of prompts in the forums, the absence of negotiations can be partly explained by the incongruence between the time and effort necessitated by negotiations and the time that participants were willing to invest. Taking into consideration the voluntary nature of participation of the parties involved, it was not feasible to achieve a higher level of commitment and involvement during the negotiations.

6.2.2 Factor 2 – Organizational context of the program

The application of the A4GE was influenced by the organizational context of the program. Evaluation practices that typically occur in the School of Education are characterized by being indicator-based and non-interpretive. They are also commissioned by authorities within the School, participative in nature in that they use multiple stakeholders, but non-collaborative in that participants are mere sources of information and do not engage in other aspects of the evaluation. Against these norms, the A4GE was introduced as an unusual type of evaluation, which made it challenging to fully engage the participants. Furthermore, the use of the A4GE was not commissioned nor was it supported by any authority figures that are usually associated with evaluations (such as the Head of School or the Academic Board). As such, it was challenging to "sell" the evaluation as having potential for instrumental use and actual improvement. To add to these complications, the use of the A4GE was hindered by the weak collaboration channels in the School of Education at the level of programs. While there

exist organizational structures that enable lecturers' collaboration at the level of discipline-based teams, professional discourses about programs are not commonplace.

Based on all of the above, the challenge before me, as the evaluator, was to convince participants of the need to collaborate at the level of the program and engage in a time-demanding interpretive evaluation. This challenge partly explains the lack of evidence related to instrumental use, particularly that the evaluation lacked the critical weight to retain their commitment.

6.3 Conclusion

In this chapter, I outlined the findings from my empirical investigation of the theory of the A4GE. I was able to examine the extent to which the A4GE's highly theoretical propositions aligned with its practice by providing insights into the model's operational specificity, range of application, feasibility in practice, and discernible impact.

While the findings revealed a considerable degree of alignment between the theory and practice of the A4GE, they also revealed some practical limitations of the model's theory. To provide further explanations to the observed differences, I discussed two contextual factors that were significant in moderating the congruence between theory and practice. In the following chapter, I discuss the key findings derived from this study and critically examine the limitations of enacting this model in general. I also discuss the contributions and implications of the study, and develop recommendations for future research on and practice of the A4GE.

CHAPTER 7: CONCLUSIONS AND IMPLICATIONS

The purpose of this study was to examine the congruence between the theory and practice of the A4GE in evaluating a secondary science teacher preparation program.

The general research question was:

> **How congruent are the underlying theory and practice of an adapted version of the 4GE model in the context of evaluating a secondary science teacher preparation program in an Australian university?**

To answer this question, I synthesized pertinent literature to develop a theoretical tool, the PEMED, which I used to clarify the A4GE's theoretical propositions. I then collected data about each of these propositions by implementing the A4GE in a case study of a pre-service science teacher preparation program. Making use of Miller's (2010) criteria, I examined the congruence of the theory and practice of the A4GE in relation to the model's operational specificity, range of application, feasibility in practice and discernible impact.

In this final chapter, I outline key findings from this study and discuss them in relation to relevant literature. I also highlight the contribution of this study to theoretical, empirical and methodological knowledge, and I discuss this study's limitations. Lastly, I present the implications of my findings and provide recommendations for research and practice.

7.1 Discussion of key findings of the study

The study consisted of two components, the evaluation case study of the GDE(ST) using the A4GE, and the empirical examination of the theory/practice relationship of the A4GE. In this section, I report the findings from the evaluation case study and discuss the findings from the empirical investigation of the A4GE.

7.1.1 Findings from component one: The evaluation of the GDE(ST)

The findings from the evaluation case study are the claims, concerns, issues and suggestions for improvement that lecturers and graduate teachers from the GDE(ST) expressed about the program. Common claims made by both stakeholder groups included: flexibility of the off-campus program; effectiveness of the program in instructing student teachers on how to develop their teaching programs and lesson plans; and the proficiency of the lecturers teaching into the program. The only concern common to both stakeholder groups related to

the duplication of content within the program. One concern identified by lecturers only was the lack of collaboration among lecturers. Other concerns, only expressed by the science teachers, related to the inappropriateness of content for secondary science teaching. Interestingly, most of the concerns expressed by one group were not discussed by the other, which suggests that different stakeholders had different concerns about the program. Some of the issues common to the two stakeholder groups related to the amount of lecturer support provided throughout the units as well as the perceived relevance and usefulness of the content. Issues that were expressed only by lecturers included the need for mandatory residential schools and the quality of existing program evaluation processes. The issues expressed by the science teachers included learning the content of the HSC syllabuses and the quality of instruction in the units.

Only lecturers participated in negotiations of the nine nominated issues. While I reported some instances where issue resolution occurred, I cautioned that these resolutions were only between lecturers and therefore did not amount to successful negotiation outcomes from the case study.

7.1.2 Findings from component two: The empirical investigation of the A4GE

The findings from the empirical investigation of the A4GE suggest that there was considerable congruence between the theory and practice of the evaluation model. I found that the theory of the A4GE was operationally specific and offered me sufficient guidance to conduct the evaluation using an interpretive, responsive and context-bound approach. The theory also made me aware of my roles and the roles of other participants. In addition, I found that the A4GE theory clarified the range of application of the model and specified the necessary conditions for its implementation. In particular, the model theorized that participants needed to share the view that their personal experiences and those of other stakeholders are essential to the evaluation and development of the program. They also needed to accept the goals of the A4GE and commit to the development of the CCIs and suggestions. While these conditions were not all satisfactorily met in practice, as evident by the high attrition rate and low levels of participant commitment to the evaluation process, the theory of the A4GE was, nonetheless, successful in clarifying the practical conditions needed for its successful implementation. My findings also showed that many of the processes described in the A4GE were feasible in my case study. For instance, I was able to apply the dialectic hermeneutic cycle, use constant comparison to derive the CCIs and suggestions for

improvement from the data, design and use an interactive website to share the findings of the evaluation in a timely manner, and develop a virtual asynchronous negotiation forum to carry out the negotiations. Finally, the study findings revealed that at least some impact can be attributed to the A4GE, because I was able to identify numerous occurrences of process use that were demonstrated by the variety of learning outcomes on the part of the lecturers that the evaluation engendered. Concerning instrumental use, I was not able to provide substantial evidence that use occurred beyond the reported descriptions presented by participants.

Notwithstanding these areas of congruence, the findings from the reported study also revealed some areas of incongruence between the theory and practice of the A4GE in several respects. I needed to compromise on some of the processes of the A4GE; for instance, I did not engage stakeholders in the design of the evaluation, data analysis and reporting.

In what follows, I discuss two aspects of the A4GE which were not feasible in my case study, authentic partnership and negotiations.

In Chapter 6, I presented a discussion of how the evaluation resources and the organizational context of the program might have contributed to this misalignment between the theory and practice of the A4GE. In this section, I further argue that these aspects constitute serious limitations to the implementation of the A4GE but not to its theory. I present my argument in light of the unsustainability of the evaluation model as well as its misalignment with the requirements of current governance systems within Australian universities.

7.1.2.1 Aspect 1 – Authentic partnership

Study participants were not committed to the evaluation in a way that reflected authentic partnership. In terms of contributing their expertise, both stakeholder groups were open to disclosing their experiences with the program. This contrasts with the findings of Kosh (2000) whose evaluation of an elderly nursing home using the 4GE revealed that many participants were particularly reluctant to share their experiences because they "feared exposure and victimization by nursing staff, and their involvement in the 4GE had the potential to make things worse rather than better" (p. 120). With regard to the depth of participation in the evaluation, the extent of participant involvement in technical decision making and evaluation processes was limited to providing feedback about the evaluation procedures as they unfolded. Thus, participants were not partners in the evaluation because they were not involved in the planning of the evaluation and the development of its activities and processes. As I clarified in Chapter 6, limiting the extent of participants' engagement in

the planning and development of evaluative activities was my decision as the evaluator given the context of the evaluation and the limited time and resources available for its implementation. These findings echo those reported by Christie (2003) who examined whether a group of evaluation practitioners faithfully enacted the participatory intent of the models they were using. In comparing these practices to the theoretical formulations described by the models' theorists, Christie found that the extent of stakeholder involvement in practice varied considerably from the models' theories. These findings are also concurrent with those reported in a recent study synthesizing the empirical research literature on stakeholder involvement in program evaluation. In their synthesis, Brandon and Fukunaga (2014) found that, in several studies, evaluation researchers were challenged to implement participatory approaches to evaluation because the "demands on evaluators and program personnel … require considerable time commitment and, in some cases, material resources" (p. 38). This finding resonates with the findings in my study, in which it was challenging for me and too demanding of participants to engage in the evaluation planning and development of the evaluation's processes and activities.

7.1.2.2 Aspect 2 – Negotiation

The findings from my case study revealed that the negotiation process, as conceptualised in the A4GE, was not feasible in the context of my study. Instead, my observations of the interactions that occurred during that stage suggest that negotiations were replaced by hermeneutic dialectic discourses where participants were responding to the prompts (questions and proposals) on the forum by sharing their own stories and referring to those of others. In discussing her findings, Kosh (2000) reported that negotiations were unsuccessful in all three of her reported case studies of the 4GE. She presented several obstacles to successful negotiations, such as the imbalance of power relations and the difference in participants' abilities to articulate their arguments. Similarly, Huebner and Betts (1999) found that negotiation sessions did not generate a great deal of controversy, which caused difficulties in getting the dialogue going. These findings are aligned with those derived from my study, in which I noticed that most participants who were involved in the negotiation session contributed only once to each of the issues they discussed. In contrast, Lay and Papadopoulous (2007) reflected on their application of the 4GE in the context of evaluating the Sure Start program, which aimed to fight child poverty and social exclusion in England, and claimed that their negotiations were successful. Nevertheless, these authors did not

present an elaborate description of the negotiation processes or setting and, therefore, it is difficult to compare my findings to theirs.

7.1.3 Challenges to sustainably implementing the A4GE in the current context of higher education governance

Many factors could have contributed to the observed failure in enacting the two components discussed above in my case study. In Chapter 6, I argued that the limited time and human resources needed for the evaluation and the organizational context of the GDE(ST) were two factors that might have hindered my ability to enact these two components. In this section, I argue that these components constitute serious limitations to the applicability of the A4GE in the general context of higher education governance. Particularly, I argue that the democratic intent of full participation and negotiation theorized in the A4GE is difficult to enact in practice because the A4GE (1) is not sustainable, (2) does not emphasize accountability, and (3) does not specify performance measures.

7.1.3.1 The A4GE is not sustainable

The challenges to implementing the 4GE, and hence A4GE, have been noted by several researchers due to its time-consuming and demanding nature (Lay & Papadopoulos, 2007; Fishman, 1992). Chouinard (2013) argues that one challenge for achieving full participation in evaluation is often attributed to the lack of time and resources necessary for developing the kinds of relationships between the evaluator and other stakeholders that facilitate participation. The A4GE advocates prolonged involvement of participants who need to commit at the outset to three rounds of data collection. Although a valuable feature of the model, this has practical limitations in terms of retaining and sustaining the participation of willing individuals (Stufflebeam & Shinkfield, 2007).

The A4GE requires a lot of time and effort from participants, who need to contribute to the interviews as well as to the negotiations. After each round of interviews, participants need to browse through the CCIs made by other participants and examine how they are congruent with or different from their own. They are also required to discuss issues thoroughly in the negotiation forum and collectively debate solutions for these issues. In my study, several participants indicated that they simply did not have the time to commit to these demanding tasks of the A4GE. Nevertheless, this situation is not peculiar to my study context and can be expected to arise in any other educational context where workplace demands on participants leave no room for the intensive participation required by the A4GE. The A4GE also requires

significant amounts of resources be implemented correctly. The A4GE requires a lot of time and effort from the evaluator, who needs to perform data collection and data analysis concurrently (because interviews are based on emerging insights during the data collection), build rapport with participants, present useful resources to participants to inform their arguments, and provide training whenever necessary to empower participants in the evaluation. At the same time, the evaluator needs to develop timely and prompt reports to the participants at regular intervals. This requires human and financial resources to carry out immediate interview transcriptions and analyses, resources that are not always available at the evaluator's disposal. The A4GE theory also indicates that the evaluator must provide adequate training for participants to assist them in developing sophisticated constructions about the program. This training not only puts more burdens on the participants and evaluator, but also requires trained professionals to assist the evaluator, particularly if the participants are geographically dispersed and have different training needs. Given the multitude of demands placed before the evaluator and the participants, the A4GE is, at best, qualified as an unsustainable model of regular program evaluation.

Furthermore, not all participants can benefit from the evaluation activities to the extent that justifies and sustains their participation. For instance, in my study, lecturers explicitly recognized the relevance of the evaluation to their everyday practice. This was not true, however, for the science teachers participating in the evaluation. These teachers did not have an obvious stake in the evaluation because they had already graduated from the program. While their input was necessary to the evaluation, their involvement was solely driven by personal intrinsic motives. For example, ST4 (Interview 1) explained that he was happy to participate in the evaluation because he wanted to contribute back to the university he so loved. In those cases where one or more participants do not perceive personal gain from the evaluation, it can be far more challenging to sustain their involvement in the evaluation.

In sum, the A4GE requires time, human and financial resources that are limited in higher education contexts and, therefore, its implementation as a regular program evaluation mechanism is unlikely to be sustainable in such contexts.

7.1.3.2 The A4GE does not emphasize accountability

Educational institutions in Australia must have their programs accredited by the Federal government and/or relevant state authorities. The governance paradigm embraced by the Australian Commonwealth Government is characterized by an emphasis on accountability as

well as outcomes-orientation (Shah, Lewis, & Fitzgerald, 2011). In this paradigm, evaluation is framed as a management tool designed to monitor performance and ensure accountability (Chouinard, 2013). The 4GE, and hence A4GE, are a "countermeasure to assigning responsibility for successes and failure in a program to certain individuals or groups" (Stufflebeam, 2008, p. 1398). While universities can carry out program evaluation any way they prefer, it is common practice within universities to perform evaluations that yield the evidence needed by the government so as to maximize efficiency and avoid redundant expenditures. These evaluation practices usually take the form of standards-based evaluations. Therefore, the A4GE is in stark contrast to the view of evaluation sustained within universities. As such, it is less appealing to institutions as a tool for program evaluation because it is expensive and lies outside the norm of their priorities. Therefore, while the A4GE is sound as an evaluation model, its outcomes are not responsive to accountability mandates that come from within and outside the institutions.

7.1.3.3 The A4GE does not specify performance measures

The A4GE does not predetermine performance measures to evaluate programs. Rather, its criteria are constructed inductively throughout the evaluation based on the CCIs of the various stakeholders. However, the view of evaluation that reigns within higher education institutions is based on measuring performance indicators and providing tangible evidence on these. Therefore, it is difficult to implement the A4GE in a context where the valuable type of knowledge is tracking performance and progress against stated goals. While evaluation scholars argue persuasively for participatory models to program evaluation, there are other tensions from AITSL and TEQSA and similar agencies that drive universities to conform to norms and standards, and provide evidence on indicator-based performance measures. This context presents further limitations to the applicability of the A4GE in the context of current Higher Education governance regimes.

7.2 Contributions of the study

This study contributes knowledge at three levels: theoretical, empirical and methodological. In terms of theoretical knowledge, this study contributes to the development of the PEMED. In relation to empirical knowledge, this study contributes a detailed case study examining the congruence of the practice of the A4GE to the theoretical propositions underlying the model. Lastly, in terms of methodological knowledge, the study contributes a conceptual framework

that builds on Miller's (2010) framework and which can be used to examine the congruence of the theory and practice of program evaluation models.

7.2.1 Contributions to theoretical knowledge

The PEMED was developed in this study as a theoretical tool for clarifying the theoretical propositions of the A4GE. Furthermore, it was used as a research tool to guide the investigation of the model in the context of practice. The PEMED is beneficial because it provides a language though which evaluators and researchers can articulate their perspectives about the dimensions that define a model uniquely, thus providing an explicit and elaborate definition of the theory underlying a program evaluation model. Importantly, the PEMED is not peculiar to the A4GE and can be used to investigate various evaluation models, both singularly and comparatively, by allowing comparisons of various models along the different dimensions. It is, therefore, a tool for reflection about and analysis of the differences and similarities across evaluation models.

Furthermore, the PEMED could potentially provide an agenda for research that is important to the advancement of knowledge on program evaluation. Indeed, the PEMED can be used as a conceptual organizer for clarifying the theories underlying evaluation models. Moreover, the dimensions of the PEMED themselves delineate areas of research that need further interrogation. Additionally, the PEMED provides a structure where the relationships between the various dimensions could be questioned and queried. As such, the PEMED could provide scholars with a scheme to consolidate theoretical gains from empirical investigations, a guide to manoeuvre the field and navigate its landscape, and a tool to identify innovation when it occurs and to expand the knowledge of the profession.

7.2.2 Contributions to empirical knowledge

This study provides empirical knowledge about the A4GE as a model for program evaluation. In particular, the study offers insights into the theoretical tenets of the model and how they were enacted in practice across a well-defined context. The reported evaluation case study also provides specific examples of the processes and outcomes of the A4GE. Additionally, it highlights some challenges to enacting the model in practice. As such, the study adds to the small existing base of empirical studies which use interpretive evaluation models for program evaluation.

7.2.3 Contributions to methodological knowledge

In this study, I used Miller's (2010) criteria in conjunction with the PEMED to develop and apply a conceptual framework for examining the relationship between the theory and practice of one evaluation model. The conceptual framework is a major contribution of this study because it presents guidelines for carrying out research on any program evaluation model. The value of the framework lies in its potential to enable researchers to examine the theory-practice relationship in program evaluation, an understudied area in evaluation scholarship. The framework can, thus provoke deeper learning about evaluation theories and further understanding of the connections between theory and practice.

7.3 Limitations of the study

Three limitations to this study deserve particular attention. The first important limitation is the way the participants were recruited. About 50 members (in total) were invited to take part in the evaluation, which is a reasonable number in the context of a faculty program review. However, only 14 members chose to participate. These participants belonged to two stakeholder groups: science teacher graduates and lecturers. Volunteer respondents who participated in this research might have had a higher stake or interest in the evaluation. Therefore, I suspect that response bias occurred because only those participants who wanted to have a voice in the evaluation actually participated. However, this is the nature of interpretive research and is not attributed to the evaluation design or implementation. Furthermore, some participants clarified that they would rather refrain from discussing some issues when they suspected that their opinions would make them identifiable. Although small sample sizes are common in interpretive research, I believe that a larger sample size could have yielded more comprehensive discussions about these issues. While the main purpose of this study was to empirically investigate the theory of the A4GE rather than to develop extensive evaluation outcomes, I believe that researchers on evaluation can better understand the theory of a program evaluation model when these evaluations are thoroughly and authentically implemented. Nonetheless, recruiting additional participants can potentially introduce additional challenges to the feasibility and management of the evaluation, particularly if it was implemented in a context where time and human resources were scarce.

The second important limitation of this research involves the fact that not all aspects of the evaluation model were implemented effectively. Therefore, it was not possible for me to examine and study their practice as planned at the beginning of the study. This is true for the

negotiation forum, which was an amendment to the 4GE. While I originally planned to investigate the usefulness of this addition by using a post-negotiation questionnaire tool, negotiations did not occur, and only one group of stakeholders participated in the forum. Therefore, I could not administer the questionnaire to investigate whether using a virtual asynchronous forum had actually reduced the power differentials often reported in face-to-face negotiations. Similarly, while I was able to report some occurrences of process and instrumental use, these findings were only limited to the immediate impacts of the evaluation. The study does not provide insights into other potential uses that could occur in the long-run. However, this challenge is common in studies on evaluation use (e.g. Alkin & Taut, 2003; Henry & Mark, 2003a; Mark & Henry, 2004) and researchers are yet to determine a way to capture and study long-term use. Furthermore, any reported findings concerning instrumental use were based on information reported by the participants. Whether or not these reported intentions were translated into later actions was not measured in this study. Perhaps future longitudinal studies examining use generated by the A4GE can shed light on this issue and clarify the nature and extent of the evaluation model's impact.

The third and final important limitation to this study related to the ethical issue of confidentiality. The nature of this research necessitated that I indicated the identity of the host institution as well as the academic area involved in this study, which could inflict judgments on the quality of education provided by the institution based on the findings from this research. This ethical dilemma remained unsolved in my study.

7.4 Implications and recommendations

7.4.1 Implications and recommendations for researchers on program evaluation

The conceptual framework used in this study enabled me to empirically investigate the theory of the A4GE. While the framework is only tentative, I believe that it can provide a basis for further research and development. The PEMED proved useful in elucidating the theoretical propositions underlying the A4GE. It was also a useful tool for collecting data about the practice of the model. Nevertheless, I believe that more research will be needed to develop the PEMED further and to capture the complex dynamics that exist between the various dimensions.

The A4GE was successful in promoting process use and generating CCIs. Therefore, the model is promising if it can be used as a tool for program improvement or capacity building. However, while the web tools might have contributed a better communication medium, they

did not overcome the challenges of time and effort demands. Therefore, further research is needed to examine the effectiveness of virtual media in promoting better negotiation environments.

7.4.2 Implications and recommendations for program evaluation practitioners

The findings from my study suggested that full participation in all activities required by the A4GE was not feasible in the context of periodic evaluations of university programs and that negotiations were only carried out as a continuation of the hermeneutic dialectic cycle. I argued that these findings were partly explained by the unsustainability of the A4GE, particularly when implemented in the context of governance prevalent in tertiary education institutions that favour other models for evaluation. Therefore, I believe that future implementations of the A4GE would benefit from considering and integrating more sustainable practices within the model.

King (2011) suggests three courses of action that evaluators can follow to cultivate sustainable participation in accountability-oriented governance regimes. She argues that evaluators would first need to develop more "affordable" versions of participatory practice, as illustrated in Huebner and Betts (1999) who conducted a second round of data collection using a questionnaire instead of conducting interviews. In the questionnaire, participants were asked to indicate their level of agreement on each item as well as how important they thought that item was. Huebner and Betts (1999) reasoned that using questionnaires was less time consuming and less demanding of participants. Another possible way to achieve sustainability is by embedding evaluation practice in regular faculty meetings where critical issues can be addressed in a trusting atmosphere. Additional participation from other stakeholders can also be embedded in regular activities where repeated discussions can take place.

Furthermore, King (2011) suggests that evaluators need to build evaluation capacity for individuals as well as organizations. This can be done by providing adequate training for the different stakeholder groups and building more solid relationships among all parties. In addition, King (2011) advises that evaluators need to "sell" the evaluation to key decision-makers by instructing them about the potential benefits of the model. In my case study, the evaluation was not supported by any authority figure. This might explain the high attrition rate in the evaluation. Support from key decision-makers (such as Academic Boards and the

Head of School) might enhance the perceived weight of the evaluation to bring about changes and, therefore, could increase and enhance participation.

In sum, if more sustainable versions of the A4GE can be developed and implemented, these evaluation practices can be used to complement existing forms of evaluation by promoting depth of understanding without placing excessive demands on participants.

7.5 Summary

In this study, I developed a conceptual framework building on the work of Miller (2010) to examine the relationship between the theory and practice of the A4GE. My investigation revealed that, while there exists considerable congruence between the theoretical propositions of the model and its practice, there are serious obstacles to sustainably implementing the model in the current governance regimes within tertiary education institutions. Given the A4GE's benefits in addressing local program needs, and building individual and organizational capacity, I believe that there is value in conducting further research on the model and on factors that enable its feasibility across educational sectors.

REFERENCES

Alkin, M. (2013). *Evaluation Roots: A Wider Perspective of Theorists' Views and Influences* (Second ed.). Thounsand Oaks, California: SAGE.

Alkin, M., & Christie, C. (2005). Editor's note. *New Directions For Evaluation, Summer* (106), 1-4.

Alkin, M., & Taut, S. (2003). Unbundling evaluation use. *Studies in Educational Evaluation, 29*, 1-12.

Anderson, D., Johnson, R., & Milligan, B. (2000). Quality Assurance and Accreditation in Australian Higher Education: An Assessment of Australian and International Practice. Commonwealth of Australia: Department of Education, Training and Youth Affairs.

Anderson, G. (2006). Assuring quality/resisting quality assurance: Academics' responses to 'quality' in some Australian universities. *Quality in Higher Education, 12*(2), 161-173.

Australian Academy of Science (AAS). (2005). Primary Connections: Linking science with Literacy Retrieved November 1, 2011, from http://www.science.org.au/primaryconnections/index.html

Australian Government. (2013). Higher standards for teacher training courses. Retrieved on April 20, 2014 from http://archive.industry.gov.au/ministerarchive2013/chrisbowen/MediaReleases/Pages/Higherstandardsforteachertrainingcourses.html

Bickman, L. (1987). The functions of program theory. *New Directions for Program Evaluation Spring*(33), 5-18.

Blackmore, J. (2009). Academic pedagogies, quality logics and performative universities: Evaluating teaching and what students want. *Studies in Higher Education, 34*(8), 857-872.

Brandon, P., & Fukunaga, L. (2014). The state of the empirical research literature on stakeholder involvement in program evaluation. *American Journal of Evaluation 35*(1), 26-44.

Campbell, B., & Mark, M. (2006). Toward more effective stakeholder dialogue: Applying theories of negotiation to policy and program evaluation. Journal of Applied Social Psychology, 36(12), 2834–2863.

Charmaz, K. (2011). Grounded theory methods in social justice research. In N. Denzin & Y. Lincoln (Eds.), *The SAGE Handbook of Qualitative Research*. Thousand Oaks, California: Sage.

Chelimsky, E. (1998). The role of experience in formulating theories of evaluation practice. *American Journal of Evaluation, 19*(1), 35-55.

Chouinard, J. A. (2013). The case for participatory evaluation in an era of accountability. *American Journal of Evaluation, 34*(2), 237-253.

Christie, C. (2003). What guides evaluation? A study of how evaluation practice maps onto evaluation theory. *New Directions For Evaluation*(97), 7-35.

Christie, C. (2007). Reported influence of evaluation data on decision makers' actions: An empirical examination. *American Journal of Evaluation, 28*, 8-25.

Chubb, I. (2014). Australia's future depends on a strong science focus today. Retrieved on the 20th March 2014 from http://theconversation.com/australias-future-depends-on-a-strong-science-focus-today-22075

Cohen, L., Manion, L., & Morrison, K. (2007). *Research Methods in Education* (Sixth ed.). New York: Routledge.

Committee for the Review of Teaching and Teacher Education. (2003). Australia's Teachers Australia's Future: Advancing Innovation, Science, Technology and Mathematics. Canberra: Department of Education, Science and Training (DEST).

Contandriopoulos, D., & Brousselle, A. (2012). Evaluation models and evaluation use. *Evaluation, 18*(1), 61-77.

Cooksey, R., & McDonald, G. (2011). *Surviving and thriving in postgraduate research.* Prahran, Victoria: Tilde University Press.

Coryn, C. L. S., Noakes, L. A., Westine, C. D., & Schröter, D. C. (2011). A Systematic Review of Theory Driven Evaluation Practice From 1990 to 2009. *American Journal of Evaluation, 32*(2), 199-226.

Cousins, B., & Whitmore, E. (1998). Framing participatory evaluation. *New Directions For Evaluation*(80), 5-23.

Creswell, J. (2007). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches* (Second ed.). Thousand Oaks, California: SAGE.

Datta, L.-e. (2003). Important questions, intriguing methods, incomplete answers. *New Directions For Evaluation, 97*, 37-46.

Denzin, N. (1989). *Interpretive Interactionism*. Newbury Park, California: Sage.

Department of Education, Science and Training (DEST). (2002). Review of Teacher and Teacher Education: Strategies to Attract and Retain Teachers of Science, Technology and Mathematics. Canberra: Commonwealth of Australia.

Department of Education, Science and Training (DEST) Committee for the Review of Teaching and Teacher Education. (2003). Australia's Teachers: Australia's Future. Advancing Innovation, Science, Technology and Mathematics. Canberra: Commonwealth of Australia.

Dillman, L. (2013). Comparing evaluation activities across multiple theories of practice. *Evaluation and Program Planning 38* 53-60.

Dinham, S. (2013). The quality teaching movement in Australia encounters difficult terrain: A personal perspective. *Australian Journal of Education, 57*(2), 91–106.

Donaldson, S. (2003). Theory-Driven Program Evaluation in the New Millennium. In S. Donaldson & M. Scriven (Eds.), *Evaluating Social Programs and Problems: Visions for the New Millennium* (pp. 109-141). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Donaldson, S. (2005). Using program theory–driven evaluation science to crack the Da Vinci code. *New Directions For Evaluation, Summer* (106), 65-84.

Donaldson, S., & Lipsey, M. (2006). Roles for Theory in Contemporary Evaluation Practice: Developing Practical Knowledge. In I. Shaw, J. Greene & M. Mark (Eds.), *The SAGE Handbook of Evaluation* (pp. 57-76): SAGE Publications Ltd.

Fetterman, D. (2003). Empowerment Evaluation Strikes a Responsive Chord. In S. Donaldson & M. Scriven (Eds.), *Evaluating Social Programs and Problems: Visions for the New Millennium* (pp. 63-76). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Fetterman, D., & Wandersman, A. (2005). *Empowerment Evaluation: Principles In Practice*. New York, USA: The Guilford Press.

Fishman, D. (1992). Postmodernism comes to programme evaluation: A critical review of guba and Lincoln's fourth generation evaluation. *Evaluation and Programme Planning, 15*, 263-270.

Fitzpatrick, J. (2004). Exemplars as case studies: Reflections on the links between theory, practice and context. *American Journal of Evaluation, 25*(4), 541-559.

Fitzpatrick, J. (2012). An introduction to context and its role in evaluation practice. *New Directions For Evaluation, 135*, 7-24.

Fitzpatrick, J., Sanders, J., & Worthen, B. (2004). *Program Evaluation: Alternative Approaches and Practical Guidelines* (3rd ed.). Boston Massachussets: Pearson Education.

Germuth, A. (2010). Advances in evaluating evaluation theory. *American Journal of Evaluation, 31*(3), 418-420.

Gibbs, G. (2007). *Analyzing Qualitative Data*. Thousand Oaks, California: Sage.

Glaser, B., & Strauss, A. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine Publishing Company.

Goodrum, D., Druhan, A., & Abbs, J. (2012). The Status and Quality of Year 11 and 12 Science in Australian Schools. Canberra: Australian Academy of Science.

Greene, J. (1988). Stakeholder participation and utilization in program evaluation. *Evaluation Review, 12*, 91-116.

Greene, J. (1994). Qualitative Program Evaluation: Practice and Promise. In N. Denzin & Y. Lincoln (Eds.), *Handbook of Qualitative Research* (pp. 530-544). Thousand Oaks, California: SAGE.

Greene, J. (2005a). Context. In S. Mathison (Ed.), *Encyclopedia of Evaluation*. Thousand Oaks, CA: SAGE.

Greene, J. (2005b). A value-engaged approach for evaluating the Bunche–Da Vinci Learning Academy. *New Directions For Evaluation, Summer* (106), 27-45.

Guba, E. (1990). *The Paradigm Dialog*. Newbury Park, CA: SAGE Publications.

Guba, E., & Lincoln, Y. (1989). *Fourth Generation Evaluation*. Newbury Park, CA: SAGE Publications.

Hansen, H. F. (2005). Choosing evaluation models: A discussion on evaluation design. *Evaluation, 11*(4), 447-462.

Hansen, M., Alkin, M., & Wallace, T. L. (2013). Depicting the logic of three evaluation theories. *Evaluation and Program Planning 38*, 34-43.

Henry, G. (2005). In pursuit of social betterment: A proposal to evaluate the Da Vinci Learning Model. *New Directions For Evaluation, Summer* (106), 47-63.

Henry, G., & Mark, M. (2003a). Beyond use: Understanding evaluation's influence on attitudes and action. *American Journal of Evaluation, 24*, 292-314.

Henry, G., & Mark, M. (2003b). Toward an agenda for research on evaluation. *New Directions For Evaluation*(97), 69-80.

Houston, D. (2010). Achievements and consequences of two decades of quality assurance in higher education: A personal view from the edge. *Quality in Higher Education, 16*(2), 177-180.

Huebner, A., & Betts, S. (1999). Examining fourth generation evaluation: Application to positive youth development. *Evaluation, 5*(3), 340-358.

King, J. (2003). The challenge of studying evaluation theory. *New Directions For Evaluation, 97*(57-67).

King, J. (2005). A proposal to build evaluation capacity at the Bunche–Da Vinci Learning Partnership Academy *New Directions For Evaluation, Summer* (106), 85-97.

King, J. (2011). *Cultivating Participatory Evaluation (PE) in a World of Centralized Accountability*. Paper presented at the DeGEval – Gesellschaft für Evaluation, Linz, Austria.

King, L., & Appleton, J. (1999). Fourth generation evaluation of health services: Exploring a methodology that offers equal voices to consumer and professional stakeholder. *Qualitative Health Research, 9*, 698-710.

Kirkhart, K. (2000). Reconceptualizing evaluation use: An integrated theory of influence. *New Directions For Evaluation, 88*, 5-23.

Korthagen, F., Loughran, J., & Russell, T. (2006). Developing fundamental principles for teacher education programs and practices. *Teaching and Teacher Education, 22*, 1020-1041.

Kosh, T. (1996). Implementation of a hermeneutic inquiry in nursing: Philosophy, rigour and representation. *Journal of Advanced Nursing, 24*(1), 174-184.

Kosh, T. (2000). Having a say: Negotiation in fourth generation evaluation. *Journal of Advanced Nursing, 31*(1), 117-125.

Kundin, D. (2010). A conceptual framework for how evaluators make everyday practice decisions. *American Journal of Evaluation, 31*(3), 347-362.

Kvale, S., & Brinkmann, S. (2009). *InterViews: Learning the Craft of Qualitative Research Interviewing* (Second ed.). Thousand Oaks, California: Sage.

Laughlin, R., & Broadbent, J. (1996). Redesigning fourth generation evaluation: An evalaution model for public sector reforms in the UK? *Evaluation, 2*(4), 431-451.

Lay, M., & Papadopoulos, I. (2007). An exploration of fourth generation evaluation in practice. *Evaluation, 13*(4), 495-504.

Lincoln, Y., & Guba, E. (1985). *Naturalistic Inquiry*. Newbury Park, California: Sage.

Lincoln, Y., & Guba, E. (2007). But Is It rigorous? Trustworthiness and authenticity in naturalistic evaluation. *New Directions For Evaluation, Summer*(114), 15-25.

Lincoln, Y., Lynham, S., & Guba, E. (2011). Paradigmatic Controversies, Contradictions, and Emerging Confluences, Revisited. In N. Denzin & Y. Lincoln (Eds.), *The SAGE Handbook of Qualitative Research* (pp. 97-128). Thousand Oaks, California: Sage.

Luskin, R., & Ho, T. (2013). Comparing the intended consequences of three theories of evaluation. *Evaluation and Program Planning, 38*, 61-66.

Lyons, T., Cooksey, R., Panizzon, D., Parnell, A., & Pegg, J. (2006). Science, ICT and Mathematics Education in Rural and Regional Australia: The SiMERR National Survey. Armidale, The University of New England: National Centre of Science, ICT and Mathematics Education for Rural and Regional Australia.

Mark, M. (2001). Evaluation's future: Furor, futile, or fertile? *American Journal of Evaluation, 22*, 457-479.

Mark, M. (2003). Toward an Integrative View of the Theory and Practice of Program and Policy Evaluation. In S. Donaldson & M. Scriven (Eds.), *Evaluating Social Programs and Problems: Visions for the New Millennium* (pp. 183-204). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Mark, M. (2008). Building a Better Evidence Base for Evaluation Theory: Beyond General Calls to a Framework of Types of Research on Evaluation. In N. Smith & P. Brandon (Eds.), *Fundamental Issues in Evaluation*. New York: The Guilford Press.

Mark, M., Greene, J., & Shaw, I. (2006). Introduction. The Evaluation of Policies, Programs, and Practies. In I. Shaw, J. Greene & M. Mark (Eds.), *The SAGE Handbook of Evaluation*. London: SAGE.

Mark, M., & Henry, G. (2004). The mechanism and outcomes of evaluation influence. *Evaluation, 10*, 35-57.

Markiewicz, A. (2005). 'A balancing act': Resolving multiple stakeholder interests in program evaluation. Evaluation Journal of Australasia, 4(1), 13-21.

Marshall, C., & Rossman, G. (2011). *Designing Qualitative Research* (Fifth ed.). Thousand Oaks, California: Sage.

Merriam, S. (1998). *Qualitative Research and Case Study Applications in Education* (Second ed.). San francisco: Jossey-Bass.

Miles, M., & Huberman, M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook* (Second ed.). Thousand Oaks, California: Sage.

Miller, R. (2010). Developing standards for empirical examinations of evaluation theory. *American Journal of Evaluation, 31*(3), 390-399.

Miller, R., & Campbell, R. (2006). Taking stock of empowerment evaluation: an empirical overview. *American Journal of Evaluation, 27*, 296-319.

New South Wales Government. (2013). Great Teaching, Inspired Learning: A blueprint for action. Retrieved on March 14, 2013 from http://www.schools.nsw.edu.au/news/greatteaching/index.php

Office of the Chief Scientist. (2012). Mathematics, Engineering and Science in the National Interest. Canberra: Commonwealth of Australia.

Owen, J. M. (2006). *Program Evaluation Forms and Approaches* (third edition ed.). Australia: Allen & Unwin.

Patton, M. Q. (1990). *Qualitative Evaluation Methods* (Second ed.). Thousand Oaks, California: Sage.

Patton, M. Q. (1998). Discovering process use. *Evaluation, 4*(2), 225-233.

Patton, M. Q. (2002). *Qualitative Research and Evaluation Methods* (Third ed.). Thousand Oaks, California: Sage.

Patton, M. Q. (2003). Inquiry into appreciative evaluation. *New Directions For Evaluation, winter*(100), 85-98.

Patton, M. Q. (2008). *Utilization-Focused Evaluation* (Fourth ed.). Thousand Oaks, CA: Sage.

Richards, L. (2005). *Handling Qualitative Data*. Thousand Oaks, California: Sage.

Robinson, W. (1951). The logical structure of analytic induction. *American Sociological Review, 16*(6), 812-818.

Rossi, P., Lipsey, M., & Freeman, H. (2004). *Evalaution: A Systematic Approach*. Thousand Oaks, CA: SAGE Publications.

Rossman, G., & Rallis, S. (2012). *Learning in the Field: An Introduction to Qualitative Research*. Thousand Oaks, California: Sage.

Rowlands, J. (2012). Accountability, quality assurance and performativity: the changing role of the academic board. *Quality in Higher Education, 18*(1), 97-111.

Rowlands, J. (2013). The symbolic role of academic boards in university academic quality assurance. *Quality in Higher Education, 19*(2), 142-157.

Schwandt, T. (2003). `Back to the rough ground!' Beyond theory to practice in evaluation. *Evaluation, 9*, 353-364.

Scriven, M. (1969). An introduction to meta-evaluation. *Educational Product Report, 2*(5), 36-38.

Scriven, M. (1991). *Evaluation Thesaurus* (fourth ed.). Newbury Park, CA: Sage Publications.

Scriven, M. (1998). Minimalist theory: The least theory that practice requires. *The American Journal of Evaluation, 19*(1), 57-70.

Scriven, M. (2003). Evaluation Theory and Metatheory. In T. Kellaghan & D. Stufflebeam (Eds.), *International Handbook of Educational Evaluation*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Shadish, W. (1998). Evaluation theory is who we are. *American Journal of Evaluation, 19*(1), 1-19.

Shadish, W., Cook, T., & Leviton, L. (1991). *Foundations of Program Evaluation: Theories of Practice*. Newbury Park, CA: SAGE Publications.

Shah, M., Lewis, I., & Fitzgerald, R. (2011). The renewal of quality assurance in Australian higher education: the challenge of balancing academic rigour, equity and quality outcomes. *Quality in Higher Education, 17*(3), 265-278.

Shenton, A. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information, 22*, 63-75.

Shulha, L., & Cousins, B. (1997). Evaluation use: Theory, research and practice since 1986. *Evaluation Practice, 18*, 195-208.

Skolits, G., Morrow, J. A., & Burr, E. M. (2009). Reconceptualizing Evaluator Roles. *American Journal of Evaluation, 30*(3), 275-295.

Skrtic, T. (1985). Doing naturalistic research into educational organizations. In Y. Lincoln (Ed.), *Organizational Theory and Inquiry: The Paradigm Revolution*. Beverly Hills, California: Sage.

Smith, N. (1993). Improving evaluation theory through the empirical study of evaluation practice. *Evaluation Practice, 14*(3), 237-242.

Smith, N. (2008). *Fundamental Issues in Evaluation*. New York: The Guilford Press.

Smith, N. (2010). Characterizing the evaluand in evaluating theory. *American Journal of Evaluation, 31*(3), 383-389.

Smits, P., & Champagne, F. (2008). An assessment of teh theoretical underpinnings of practical participatory evaluation. *American Journal of Evaluation, 29*(4), 427-442.

Sridharan, S., & Nakaima, A. (2012). Towards an evidence base of theory-driven evaluations: Some questions for proponents of theory-driven evaluation. *Evaluation, 18*(3), 378-395.

Stake, R. (1978). The Case Study Method in Social Inquiry. *Educational Researcher, 7*(2), 5-8.

Stake, R. (1983). *Program Evaluation, Particularly Responsive Evaluation* Netherlands: Springer.

Stake, R. (1995). *The Art of Case Study Research*. Thousand Oaks, California: SAGE

Standing Committee on Educational and Vocational Training. (2007). Top of the Class: Report into the Inquiry into Teacher Education. Canberra: Australian Parliament.

Strauss, A. (1987). *Qualitative Analysis for Social Scientists*. Cambridge, UK: Cambridge University Press.

Stufflebeam, D. (2001). *Evaluation Models: New Directions for Evaluation*. San Francisco, CA: Jossey-Bass.

Stufflebeam, D. (2008). Egon Guba's conceptual journey to constructivist evaluation: A tribute *Qualitative Inquiry, 14*(8), 1386-1400.

Stufflebeam, D., & Shinkfield, A. (2007). *Evaluation Theory, Models, & Applications*. San Francisco, CA: Jossey-Bass.

Tourmen, C. (2009). Evaluators' decision making: The relationship between theory, practice, and experience. *American Journal of Evaluation, 30*(1), 7-30.

Tytler, R. (2007). Re-imagining Science Education: Engaging Students in Science for Australia's Future. Victoria: Australian Council for Educational Research.

University of New England. (2009). Academic Quality Management Policy. Retrieved on March 19, 2014 from http://www.une.edu.au/__data/assets/pdf_file/0010/37657/academicqualitymanagementpolicyandguidelines.pdf

University of New England. (2010). Unit and Course Monitoring Procedures. Retrieved on March 19, 2014 from http://www.une.edu.au/__data/assets/pdf_file/0009/37458/course-monitoring-procedures.pdf

Vo, A. (2013). Visualizing context through theory deconstruction: A content analysis of three bodies of evaluation theory literature. *Evaluation and Program Planning, 38*, 44-52.

West, M. (2012). STEM education and the workplace *Office of the Chief scientist Occasional Paper Series*. Canberra: Commonwealth of Australia.

Winchester, H. (2007). *Academic board and the academy: Seizing the moment*. Paper presented at the Evolution and Renewal in quality Assurance, Hobart. www.unisa.edu.au/cha/mgmt&gov/winchester_presentation/winchesterauqf2007.pdf

Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The Program Evaluation Standards: A Guide for Evaluators and Evaluation Users*. Thousand Oaks, California: SAGE Publications.

Yin, R. (2009). *Case Study Research: Design and Methods* Thousand Oaks, California: SAGE.

APPENDIX A – DESCRIPTION OF THE INTERACTIVE WEBSITE

In the reported case study, I designed an interactive website to enhance transparency in the evaluation processes, and to enable participants to interact with the data and develop their own constructions using the insights derived from their analysis of data. The website can be accessed using the following link: http://www.adaptedfourthgenerationevaluation.com/. The website is password protected and only participants in the study were able to access the site using personal usernames and passwords. A guest username and password can be provided on request.

The main page of the website features ten tabs: home page, tool exploration page, glossary of terms, data and summaries, forum, feedback, ethics declaration, downloads and resources, contact us and logout function.

The content from the home page served to remind the participants of the purpose of the website and provided guidance regarding how to navigate the site. On the tool exploration page (see Figure a), I provided explanations about an interactive tool that I designed to enable participants to browse through the data and findings.

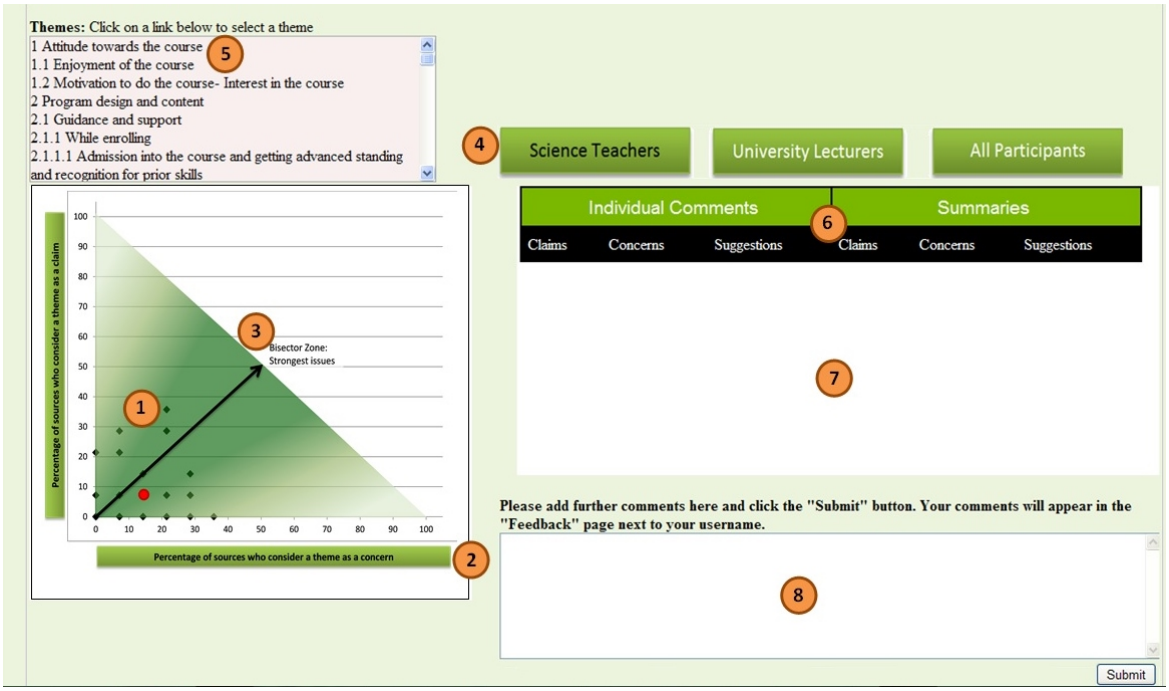


Figure a – Website: The tool exploration page

The tool consisted of two parts: a graph and a viewing pane. The graph was introduced as a visual representation that gives immediate indication, for any selected group of stakeholders, about whether a theme was a claim, a concern or an issue. The viewing pane displays the data or summaries of data related to each theme. To learn about the function of each item in the tool, participants could click on the orange circles and read the explanations on a new window. These explanations are summarized in Table a.

Table a – Explanation of the Components of the Exploration Tool

Components	Description
1. The plane field	The plane field helps in identifying whether a theme is a claim, a concern or an issue. The black dot points on the plane field represent the “themes”. Depending on your selection of a cohort, and for each theme selected, one of

	these dots will turn red. The place of the red dot on the graph indicates whether the theme is a claim (red dot is on the y axis), a concern, (red dot is on the x axis) or an issue (red dot is anywhere else). The other black dots indicate the location of the other non-selected themes on the plane field.
2. The X and Y axes	Every theme is represented by a dot point that has two coordinates, x and y. X corresponds to the percentage of participants who thought negatively about that theme and Y corresponds to the percentage of participants who thought positively about that theme. A data point (theme) that is found on the x-axis corresponds to a concern, that is, all participants have expressed negative thoughts about it. The greater the number of people who expressed their concern about that theme, the greater the value of x is. Similarly, data points on the y-axis correspond to claims. The greater the number of people who talked about a claim, the greater the value of y is.
3. The Bisector and Peripheral Zones	Issues are themes that participants disagree about. The more evenly split the participants are in their views, the stronger the issue. Strong issues exist in the dark green area, with any points on or near the black bisector arrow showing the strongest issues. Weaker issues exist in the light green areas; the lighter the area, the weaker the issue.
4. The Cohort Selection Pane	You can choose to access the data and summaries of the data related to each individual stakeholder group or to the combined set stakeholder groups.
5. The List of Themes	The list entertains a certain hierarchy for the themes as indicated by the numerical scale used. Only themes of the lowest order have data attached to them. No red dot or data point will appear when the higher-order themes are selected. The higher order themes are used as conceptual categories to organize the list.
6. Data and Summaries of Data	To access the data or summaries of the data, choose a group (science teacher, university lecturers or all participants). Next, click on a theme from the themes list. The corresponding data point is highlighted in red on the graph. Then, depending on whether you want to look at the data or at the summaries of data, you can navigate the various tabs and access the quotes related to claims, concerns, or suggestions.
7. Display Pane	Data and summaries of data are displayed in the “display pane”. If for a particular selection no data exist, the pane will indicate that.
8. Comments Box	You can leave your comments in this box. Your comments will be visible to all participants on the “Feedback” tab on the website but you will remain anonymous. Once a comment is placed in this box, it cannot be edited or deleted. If you don’t want to share your comments with other participants, but would like to share them with me, you can always send them to me on nrizk@myune.edu.au.

Once participants have understood how the interactive tool works, they can navigate to the “Data and summaries” tab to browse through the actual data and summaries of data.

The sequence of browsing the data was as follows: first, the user choose a stakeholder group (i.e., lecturers, teachers or both); then the user decides on a theme (i.e., a topic); the user then chooses to browse the actual data or the summaries of the data; and finally, the user chooses to display the claims, concerns or suggestions. A possible outcome of this series of decisions is presented in Figure b.

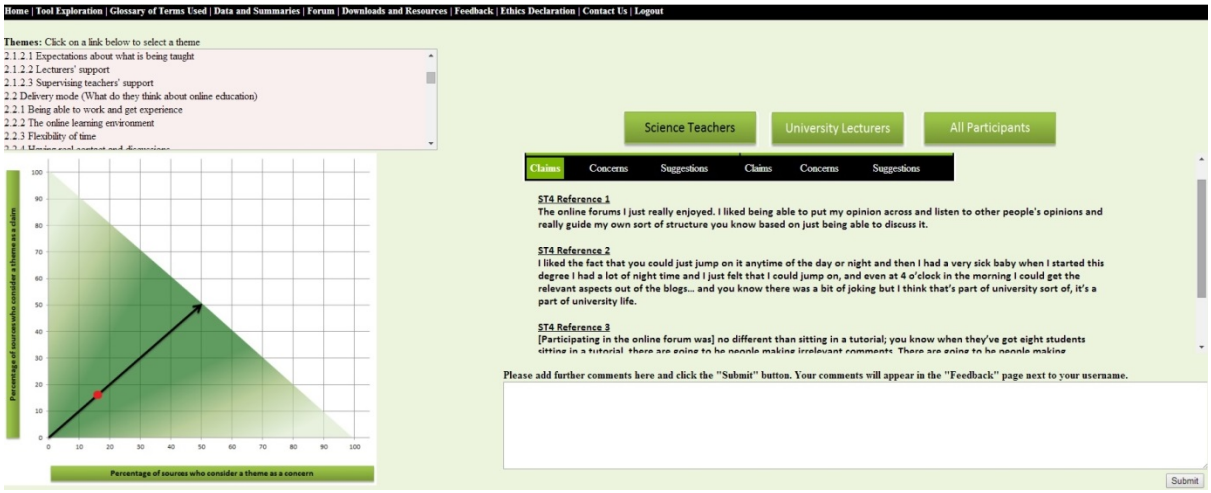


Figure b – Snapshot of the Data Exploration Tool

My rationale for this sequencing was to enable participants to focus on the data derived from the individual or combined stakeholder groups and to be able to carry out their own analyses of the data. I did not include the issues on the website because I did not want to impose my own interpretation of the issues. It was not until the negotiation session that I revealed my construction of the issues, which occurred after participants had two chances to interact with the data; once after each interview round.

The tab labelled “Glossary of terms” served to remind participants about the definition of claims, concerns, suggestions, issues and themes. The website also features: the “Forum” tab, a “Feedback” tab, where comments of participants from the comments text box were displayed; an “Ethics declaration” tab; and a “Contact us” tab. Users could log out from the website by clicking on the “log out” tab. I also included a “Downloads and resources” tab on the website. This page provided resources for informants to enrich their interpretations of the data and included the AITSL professional standards for teachers, an overview of the GDE(ST), and a list of the themes discussed during the interviews.

Prior to publishing the website, I trialled it using dummy data with the help of three lecturers and two teachers who were not involved in the study. Using their feedback, I made some adaptations to make the website more appealing and user-friendly.

APPENDIX B – MANDATORY UNITS FOR THE GDE(ST)

Unit name	Introductory blurb
Aboriginal Education	This unit provides students with the knowledge and skills to teach in Aboriginal and cross-cultural educational contexts. Students explore the legacies of colonisation and racism in Australia and examine current legislation policies and strategies that seek to redress the effects of these in educational contexts. Students study issues of power and authority in teaching-learning relationships, in school-community dynamics, in lesson planning and implementation, and in linguistic and cultural diversity. Through the use of and responses to multimedia texts and by preparing for and/or engaging in and leading interactive seminars, students both explore and demonstrate their knowledge of and skills in Aboriginal Education.
Curriculum and the Social Context of School	This unit is concerned with critical inquiry in schooling. Students will engage with conceptual frameworks which draw upon critical social and curriculum theory concerned with cultural and linguistic diversity, social class, racism, poverty, ethics, Indigeneity, place, and globalisation. They will analyse curriculum policies and teaching practices which advance social justice goals, or support the status quo, or disadvantage pupils. Students will also be encouraged to identify, analyse and develop social actions and educational strategies for the advancement of equality, equity, social justice and human rights in schooling.
Literacies in Context	The unit introduces students to the explicit teaching of language and literacy in context across all subject areas of the curriculum. It examines literacy demands, requirements and teaching strategies relevant to all curriculum areas. Special attention is given to students' personal literacy and their obligation to achieve and maintain the highest standards of literacy as part of their professional practice.
ICT in Education	This unit is designed to allow students to demonstrate the learning outcomes related to the requirements of the NSW Institute of Teachers in the area of Information and Communication Technologies (ICT). These learning outcomes encompass the areas of multimedia, social networks and communications, strategies for class and faculty administration, strategies for using information from electronic media, the ethical use of electronic information, and the educational use of software.
Planning for Effective Learning	This unit introduces pre-service teachers to issues that are important in the organisation and management of learning. The various aspects of the teaching cycle are studied - planning, designing learning activities, and assessment and reporting. The focus is on planning for diverse classrooms. Stages in development are introduced, and strategies to promote effective learning are

	examined. Assessment tasks are designed to address Professional Teaching Standards frameworks, especially those related to knowledge of pedagogy, knowledge of students and how they learn, skills in planning, assessing and reporting for effective learning, and ability to communicate effectively with students.
Inclusive and Special Education	This unit aims to develop preservice teachers' understanding of educational strategies appropriate to students with exceptional abilities in regular classrooms. The unit includes the mandatory special education content required for teachers employed in (New South Wales) public schools and is an introduction to the field of inclusive and special education.
Classroom Behaviour Management	This unit introduces pre-service teachers to issues that are important in the organisation and management of positive learning environments in the school. Strategies to promote positive behaviours are examined for both their practical relevance and theoretical purpose to allow all students to participate fully in educational activities. Assessment tasks are designed to address the application of an appropriate range of prevention, intervention and response strategies in managing classroom and individual student behaviour, risk assessment and risk management, as well as assessment of learning to better plan these strategies.
Teaching for Cultural Diversity-NESB Students	This unit is concerned with the study of the issues and attitudes around the education of culturally and linguistically diverse students in Australian schools, and with the development of understanding of appropriate policies and effective teaching practices to meet NESB students' needs. Students will examine the impact of cultural assumptions and biases in teaching NESB students. Students will investigate the rationale for current multicultural education, anti-racism, citizenship, human rights, anti-discrimination and ESL policies, and for related programs, teaching practices and resources. They will also learn about values education and building partnerships with parents and community.
Science Education 7-10: Foundation for Teaching	The unit introduces pre-service teachers to the philosophical and theoretical underpinnings of Junior Secondary Science as a continuum from primary studies of the NSW K-6 Science and Technology syllabus or equivalent. It involves developing: an understanding of junior science syllabuses to gain an appreciation of the relevance and application of key scientific concepts and skills involved; an awareness of the occupational health and safety issues associated with the science laboratory; a demonstrable understanding of classroom management strategies relevant to the scientific context; lesson sequences that incorporate interpretive approaches to teaching and learning; and an appreciation of the role of assessment to inform learning activities and teaching practice.
Science Education 7-10: Teaching and Learning	This unit covers specific aspects of teaching the compulsory component of Secondary Science with focus on the NSW BoS 7-10 Science Syllabus or equivalent. Topics presented include: the nature of science and its relevance to modern society; strategies to

	guide the selection of pedagogies appropriate to a range of scenarios for teaching junior secondary science units; the design of summative and formative assessment tasks to inform and develop individualised learning; identification of the role of mathematics and literacy in science education and their integration into learning strategies; and communicating student progress to students and parents/caregivers
Science Education 11-12: Advanced Pedagogy	This unit covers specific aspects of Senior Secondary Science teaching for non-compulsory Years 11 and 12 science study with the focus on elective courses defined by the NSW Stage 6 BoS science syllabi, and with reference to other state system requirements. Pre-service senior secondary science teachers are assisted with: identification of the place of history and philosophy in the study of science; the development of techniques of investigation identification, selection and appraisal regarding their suitability for inclusion in learning sequences; the identification of syllabus components and alternate course structures which may be incorporated into faculty programs to address specific student-school-community goals and needs within statutory authority mandates; mechanisms to assist students understanding of the nature of mandatory assessment tasks and the reporting of staged achievement (bands); and the construction and implementation of faculty programs for incorporation into classroom learning and teaching sequences as lesson plans.
Science Education 11-12: Plan, Assess and Report	This unit has a focus on the planning of units and incorporation of compulsory student assessment tasks which address specific skills mandated by Stage 6 syllabi for successful completion of Preliminary and HSC courses or equivalent. Further, the unit addresses broader issues around the profession. Students are required to develop skills and knowledge in relation to: accreditation and the range of professional development opportunities available; legislative requirements of science teachers and implications to teaching and learning; understanding of school structure, administration and dynamics; and social contexts in science teaching.

APPENDIX C – SAMPLE INTERVIEW WITH UL8 FROM THE FIRST ROUND

The study you are taking part of seeks to explore the potential of a particular approach to evaluating the Graduate diploma in education for science teachers. Throughout the evaluation process, I am hoping to interview you twice. Today will be our first interview and hopefully after 3 months we'll have another interview. Afterwards, you among other participants will be invited to take part in an online forum to discuss your ideas about the program. The interviews will be audio recorded for the sake of transcription and later analysis. The interview transcript will be sent to you soon after this interview to check the validity of its content. Any information that will be disclosed here today will remain confidential and later on, during the write-up of my dissertation, this information will be de-identified and used anonymously.

In this first interview, I want to know about your experiences with the GDE at UNE. I should clarify that I'll be using the word program synonymously with course which is a collection of units that could be either of theoretical or practical natures or both. Do you have any questions that you would like to address before we begin the interview?

1. How would you describe your work experience in the program?
2. Can you tell me about the challenges you face during your work in the program?
3. If you were to describe the most important elements of this program, what would they be? What do you like most about this program?
4. What do you like least? Could you tell me more about this? If you could change anything about the program, now that there are course restructuring, what would it be? (cross curricular perspectives)
5. I would like us now to talk about evaluation. What kind of evaluation do you think would be most appropriate to evaluate the GDE at UNE and why? Who should be involved in the evaluation, and what should be the direct and indirect outcomes of the evaluative process?
6. Are there any questions that I didn't ask that you think are important?

---

   a. Lack of individual vision of how the units articulate with the course. Duplication? Missing out on something? There is no course map in terms of the content. Is this important? Why?
      a. This issue becomes particularly important when you're talking about professional experience and sending students to apply what they learned during your courses? Do you play any role during the professional experiences of the students?
   b. UNE is moving towards having everything online while the technology is not there to support it
   c. Having the residential school optional
   d. You'll have to finalize the design of your units 12 months ahead
   e. Trimesterization
   f. Time is a major constraint to juggle the various tasks that UL are asked to do
   g. No proper evaluation exists of the course/program. Do you think this is important? Why?

h. The students are working towards the assignments as evidenced by the nature and the timing of the questions that are asked on Moodle (which is fine for some people as they believe the assignment content is enough to engage the students with the content, whereas others believe that this is not the most appropriate way to learn, working towards the assignments)
i. Based on previous interviews, I noticed there were mixed views about the perceived roles of course coordinators (managers who deal with the mechanics of the course, look into the structure of the course and its design…). From your point of view, what does the c.c. do? What do you think they should be doing?

APPENDIX D – INFORMATION SHEET AND CONSENT FORM

INFORMATION SHEET FOR PARTICIPANTS

NOTE: This page remains with you for your records.

Research Project: EMPIRICAL INVESTIGATION OF AN ADAPTED FOURTH GENERATION EVALUATION: THE CASE OF EVALUATING A SECONDARY SCIENCE TEACHER PREPARATION PROGRAM

Dear Sir or Madam,

My name is Nadia Rizk and I'm a PhD researcher in science education at the University of New England, working with Professor Neil Taylor, Dr Frances Quinn, and Associate Professor Terry Lyons. As you are involved with the UNE Science Education program, I am writing to invite you to participate in a study to evaluate the UNE secondary science teacher preparation program. The study is being undertaken as part of my PhD research project and will also assess the particular model used for this evaluation.

Researchers:

Ms. Nadia Rizk - nrizk@myune.edu.au, ph: 02 6773 3874; 04 8734 9561

Professor Neil Taylor (Principal supervisor) – ntaylor6@une.edu.au, ph: 02 6772 3984

Dr. Frances Quinn (co-supervisor) - fquinn@une.edu.au, ph: 02 6773 3411

Associate Professor Terry Lyons (co-supervisor) – terry.lyons@qut.edu.au, ph: 02 6773 2983

Your views are highly valued and we would greatly appreciate your help to improve the quality of our program. As part of the study I would like to conduct two interviews with you via phone, Skype or in person, whichever is most convenient for you. The interviews - which will be about 3 months apart - will look at your perceptions of the science teacher preparation programs offered at the University of New England. I will also ask your views on this evaluation approach itself. Later in the study you may also be invited to take part in an online discussion forum where you and other participants discuss issues that have emerged from the research. However, this forum is optional - as are all elements of the study – so please don't be put off if you can only commit to the interviews. These activities will be spread over a period of 12 months, but will only amount to a commitment on your part of around 3.5 hours.

All interviews will be audio-recorded for the sake of transcription and analysis. The interview transcripts will be sent to you soon after each interview for you to check. Your responses will be de-identified and remain confidential so that only you and I will know which comments are attributable to you; other members of the research team will not be able to identify

participants from their responses. Later in the reporting of the data I may quote your statements but you will not be identifiable and all quotes will be used anonymously. Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission. Hard copy data from this research will be stored for 5 years in a locked filing cabinet and electronic data will be password protected on a computer in the School of Education. At the end of that period hard copy data will be destroyed by shredding and electronic data will be permanently deleted.

Upon completion of the evaluation process, you will receive a report which includes the (anonymous) outcomes of the evaluation. This report, although copyrighted, could come in handy as it informs you about the various issues and concerns that various participants expressed in relation to the science teacher preparation program.

Your decision whether or not to participate will not prejudice your future relations with the University of New England. If you decide to participate, you are free to withdraw your consent and to discontinue participation at any time without penalty.

If you have any questions, feel free to contact me by phone on 0487-349-561 or by email on nrizk@myune.edu.au, and I will be more than happy to answer them.

This project has been approved by the Human Research Ethics Committee of the University of New England (Approval No. HE12/072 Valid to 01/5/2013).
Should you have any complaints concerning the manner in which this research is conducted, please contact the Research Ethics Officer at the following address:
Research Services

University of New England

Armidale, NSW 2351.

Telephone: (02) 6773 3449 Facsimile (02) 6773 3543

Email: ethics@une.edu.au

Thank you for considering this request. If you agree to be interviewed for this project, please complete the Consent Form and return by email or post.

Kind regards,

Nadia Rizk

CONSENT FORM

NOTE: Please return this consent form to Nadia Rizk by email or post if you agree to participate.

Research Project: A MODEL FOR EVALUATING THE WORTH AND MERIT OF SCIENCE TEACHER PREPARATION PROGRAMS

PLEASE NOTE THAT YOU ARE MAKING A DECISION WHETHER OR NOT TO PARTICIPATE IN EACH STAGE OF THE STUDY. YOUR CONSENT INDICATES THAT, HAVING READ THE INFORMATION PROVIDED ABOVE, YOU HAVE AGREED TO PARTICIPATE. YOU ARE FREE TO WITHDRAW FROM THE STUDY AT ANY TIME WITHOUT PENALTY. IF YOU WISH TO PARTICIPATE IN THE INTERVIEWS ONLY (AND NOT THE ONLINE NEGOTIATION SESSION), THEN PLEASE TICK 'NO' FOR POINTS 3 AND 4 BELOW, AND 'YES' FOR THE OTHER POINTS.

"I agree to take part in the research project specified above. The project has been explained to me and I have read the Information Sheet, which I keep for my records. I understand that agreeing to take part means that (tick 'Yes' for those elements which apply):

(Double click the boxes to "check")

1.	I agree to be interviewed by the researcher on two occasions	☐ Yes ☐ No
2.	I agree to allow the interview to be audio-recorded	☐ Yes ☐ No
3.	I agree to take part of the virtual negotiation session (optional)	☐ Yes ☐ No
4.	I agree to be quoted in subsequent publications (but not identified)	☐ Yes ☐ No
5.	I understand that I will be given a transcript of all the interviews concerning me for my approval before it is included in the write up of the research.	☐ Yes ☐ No
6.	I understand that my participation is voluntary, that I can choose not to participate in part or all of the project, and that I can withdraw at any stage of the project without being penalised or disadvantaged in any way.	☐ Yes ☐ No
7.	I understand that any data that the researcher extracts from the interview for use in reports or published findings will not, under any circumstances, contain names or identifying characteristics.	☐ Yes ☐ No

Participant's name: .....

Signature *(for posted hard-copy consent forms)* **.....**

Date .....