Balancing Demands: exploring dilemmas that can arise in evaluative research

Fiona M. Beals M.Ed
Research Assistant
The New Zealand Council for Educational Research

Abstract

Evaluative research is a powerful tool for education researchers. It is through evaluative research that the process of reflection and learning-in-action can occur for many practitioners in the field of education: from policy to the classroom and teacher practice. It has been argued that the most reliable and valid technique of evaluative research is the utilisation of externally contracted researchers. The intention is to enable results to be, to the greatest extent possible, transparent and reflective of the perspectives of all involved in the evaluated programme. However, no matter who performs the research, there are specific dilemmas to be addressed. These dilemmas are directly related to the expectations of the research and attitudes to the research of the many different groups or individuals involved in the evaluated programme and, further, the researcher's strive for useful, valid, and reliable results. This paper will explore the dilemmas encountered in one piece of evaluative research: *Evaluation of the Professional Development for 'The Arts in the New Zealand Curriculum',* finishing with responses to the dilemmas presented. The paper will suggest that the dilemmas of evaluative research should be worked with so that evaluative research can be effective for all levels of the education sector.

Acknowledgments

Introduction

When I wrote this paper, I originally hoped to look at the practical issues that arose in one evaluation involving multiple stakeholders and systems of administration. The paper I will present is not that paper. Instead, as I developed the paper, I found myself asking what came first: the practical, or – the often taboo word in research – the emotional? Doesn't the word 'evaluation' conjure up the types of uneasiness that we experience when we hear people utter words like: "appraisal", "performance review", "audit", or even "ERO"? Gary Anderson (1998, p.136) described evaluative research as a "prevalent activity in contemporary society". Evaluation is a key part of a society which values accountability. The professional development for the Arts curriculum involved the expenditure of $10 million. It was imperative that this expenditure be accountable to New Zealand's taxpayers. In a society that values accountability, we, as educationalists, must accept that at some time what we do will be evaluated. Further, as researchers of today's education system, we must accept that at some time we will evaluate. Because uneasiness can be triggered by the word "evaluation" we must accept that uneasiness is a dilemma in our research; an ethical dilemma that needs to be addressed in the research design.

In most evaluations, the stakeholders' feelings of uneasiness can often be associated with questions like:
- Who is this evaluation *really* benefiting (the contractor, teachers, students, etc.)?
- How is this evaluation going to *affect me* (my job position, the way I do things, my future, etc.)?
- What will happen *after the evaluation* (nothing/something)?
- *Where am I* in the picture?
- *Will the (external) evaluator truly understand* the situation/context?

To address these questions, external evaluators need to work with those they research by acknowledging and addressing the different needs and wants of each stakeholder. In the case of the research to be described in this paper, different needs and wants were addressed either in advance by the contractor, or as they arose by researchers in collaboration with the contractor. Although in this paper I intend to focus on the feelings brought to the research by the researched stakeholders, I will spend a short time looking at the values and intentions of the contractor.[1] This is important, because many of the issues and feelings of uneasiness that researchers need to address and work with are about the participants' perceptions of the contractor's power to withdraw or to support a project. In many instances the perception participants may have of the contractor's intentions may not reflect the actual intentions of the contractor and they may see the evaluation as a personal evaluation of their performance. My discussion of these issues begins with a description of the context for the research.

Introducing the context

During 2001 and 2002, the New Zealand Ministry of Education contracted the New Zealand Council for Educational Research (NZCER) to evaluate the professional development models used across New Zealand in the implementation of the new Arts curriculum. Like earlier curriculum documents the development of the Arts curriculum involved controversial debates. On one level, the controversy surrounded the key learning areas in the curriculum (dance, drama, music, and visual arts). Pedagogically, there were debates about which disciplines were included, how chosen disciplines would be included, and how the inclusion of these disciplines in one document could have adverse effects on the discrete nature of each discipline. Specifically, music and visual arts educators were concerned about their curriculum documents being included in the same document as dance and drama because of concerns about funding, timetabling and the "watering down" of their discipline. Theoretically, scholars in the visual arts debated whether the visual arts only entailed one sense, that of sight (Boyask, 2001). Practically, in the primary and intermediate classroom, teachers worried about teaching and planning for dance and drama in an already "overcrowded" curriculum (Beals, Hipkins, Cameron, and Watson, 2003).

Another level of controversy surrounded the philosophy of the Arts curriculum (Ministry of Education, 1999). On one hand, a modernist philosophy modelled the teaching of the arts in discrete units focusing on the production of an artwork. For example, in the classroom, this meant creating an artwork through the techniques and conventions (or language) of one art discipline in a discrete, timetabled slot. On the other hand, postmodernist philosophy stressed the relationship between arts and larger society. This meant teaching and learning the contexts surrounding an artwork and understanding the role of arts in an ever-changing society. The new curriculum fused together these modernist and postmodernist understandings of arts, requiring teachers and learners to learn to use the language of each discipline to explore and convey their world.

During these controversies, the Ministry of Education contracted the professional development needed for this new curriculum area to six national institutions of teacher training and development. Using research of international "best-practice" in professional development, the Ministry of Education proposed two professional development models: whole school and curriculum leadership. Providers were able to adapt the two models to suit the needs of schools in their region. Consequently, across New Zealand there were sixteen variations of the two Ministry models. Generally in variations of the whole school model, all teaching staff participated in a series of professional development sessions covering the teaching and implementation of all four arts disciplines. This model was well suited to primary and intermediate schools where the mandatory curriculum required teaching all four disciplines from Years 0–8. The second model, that of curriculum leadership, generally involved one or two school staff members receiving

professional development in two disciplines. Participating staff members then transferred their learning to other teaching staff at their school. This model was suited to the specialist teaching styles of intermediate and secondary schools and also primary schools who had existent arts strengths.

Approximately a third of all New Zealand schools participated in the professional development. Across all participating schools, contributing schools were the most over-represented school type and secondary schools the most under-represented school type. The experiences of full primary, contributing, composite, and intermediate schools thus became the foci of investigation in the research. Within the professional development models, the whole school model attracted small- (1–120 students) and medium-size (121–350 students) schools, whereas the curriculum leadership model tended to attract schools with enrolments over 350 students.

Evaluative research coincided with the provision of professional development and was conducted by an external research provider: NZCER. This meant that the Ministry of Education was able to receive an objective evaluation of the two models across all six providers. This evaluation could then be fed into further decisions on research in arts teaching and on the direction of arts education and professional development. The evaluation involved questionnaires, telephone interviews, document analysis, and case study research (Table 1). In this paper, I will use the NZCER research team's experience to show researchers need to work with the feelings of uneasiness that are brought to the research setting.

Table 1
Instruments and Types of Analyses Used

Instrument	Administered by	NZCER Analysis Involved	
		Qualitative	Quantitative
2001 MOE Questionnaire	Providers	*	✔
2002 MOE Questionnaire	Providers	*	✔
2001 Provider Milestone Reports		✔	
2002 Provider Milestone Reports		✔	
2001 Principal Questionnaire	NZCER	*	✔
2002 Principal Interview	NZCER	✔	✔
2001 Teacher Phone Interview	NZCER	✔	✔
2002 Teacher Phone Interview	NZCER	✔	✔
2001 Provider Phone Interview	NZCER	✔	✔
2002 Provider Phone Interview	NZCER	✔	✔
Case Studies	NZCER	✔	

* These instruments collected qualitative data that was then coded into quantitative categories. Shaded rows indicate data that was provided to NZCER by the Ministry of Education and/or individual providers.

Evaluative research: dealing with those uneasy feelings

Whose feelings, whose values?

Evaluative research is especially prone to stakeholder value dilemmas. These types of dilemmas can easily be assumed to be only two-way: between the contractor (or funding provider) and the research provider. However, there are many more stakeholders (or players) in evaluative research. Guba and Lincoln (1989, p. 40) define the stakeholders of evaluative research as "any persons or groups that are put at some risk by the evaluation". In evaluations of educational programmes there are many stakeholders. There are many people who feel uneasy or put at risk. For example, in the Arts PD evaluation, in addition to the Ministry of Education and NZCER, the researched stakeholders included:
• providers of professional development;
• school management staff;

- teaching staff; and
- school communities (particularly students).

With all these stakeholders there is a need to look at where feelings of uneasiness come from in evaluative research. Guba and Lincoln (1989) outline two fundamental dilemmas of any form of evaluative research that can be linked to the researched stakeholders' feelings of uneasiness. The first dilemma concerns the differing stakeholder values brought into the evaluation, and the second dilemma focuses on the political nature of *all* evaluations. Their method, fourth-generation evaluative research, incorporates the values of each stakeholder whilst acknowledging that any evaluation has political consequences. Guba and Lincoln attempt to accomplish this through involving each stakeholder in each stage of the research design. Consequently, no one is left feeling uneasy. This design is well suited to a researcher or community-driven initiative where researchers and stakeholders are able to work together to ensure a desirable outcome.

However, what Guba and Lincoln (1989) present is a fantasised situation; the reality is very different in the financially contestable research contract scenario. In reality, evaluative research is contracted by one stakeholder, and the many other stakeholders who play a part in the evaluated programme may feel uneasy or put at risk by the action of evaluation. In this context, the evaluators need to balance the requirements of the contractors with the feelings of the evaluated. In such a context, there is little space to incorporate each stakeholder's views and perspectives. In contracted research, it is very rare to see all stakeholders having an open say in the research design, especially in the research questions and methodology. Furthermore, it is up to the contractor to foresee and build appropriate research methodologies and questions into the contestable *Request for Proposal* (*rfp*). So, in the Arts PD evaluation, what were the values of the contractor?

The contractor's values and the political nature of evaluative research

In the Arts PD evaluation, the contractor's values were clear from the onset. In the *rfp* the Ministry showed that they valued: a multi-perspective approach to research; independent reviews of Ministry-contracted programmes;[1] and research that was informative to all parts of the education sector. Their intentions were to provide an evaluation of a Ministry- contracted professional development initiative that could inform future initiatives and policy development. However, the contractor, a policy advisory Ministry, is inherently political. This brought some practical issues to the research partially around what the research team and the researched stakeholders perceived to be the Ministry's values and intentions and how the data would serve these values and intentions.

The research team needed to take account of the Ministry's values when making practical research decisions such as the use and the balance to be sought between qualitative and quantitative data. On one hand, there was a need to have numbers – statistics help validate the reasons for decisions at the Ministerial level. On the other hand, descriptive data also informed these decisions as well as informing the practice of curriculum developers, providers of professional development, school managers, and school teachers. To ensure that the research was of benefit to these people examples of best practice needed to be provided. Furthermore, the qualitative data also helps to provide insight into future policy decisions (including at the Ministerial level) because it adds contextual richness to the quantitative bones of analysis. However, there was a risk that the research team would fall into an "either/or" trap where the team would assume that the Ministry of Education valued one type of data over another. In the

---

[1] Although the Ministry did not explicitly state in the *rfp* that it valued independent reviews of contracted programmes, it can been inferred by researchers in that a *rfp* was distributed.

interim report the research team did indeed fall into this trap because it perceived that the Ministry valued numbers over descriptions. Despite gathering a mixture of qualitative and quantitative data, the interim report reduced all the qualitative indicators into quantitative numbers. Consequently, this report was one of numbers and not descriptions. Although the report was able to feed into some decisions at the policy level it did not give rich examples of good practice or examples of how the teacher's learning in the professional development was being implemented in classrooms.

Gathering and analysing both qualitative and quantitative information in a report requires the gathering of much data. Different stakeholders bring their own values to the research, and may perceive intentions for the use of this data differently. Researched stakeholders' perceptions of the values or intentions of a contractor also impact on their responses to the different forms of data. Qualitative data allows the researched stakeholder to express their own values in their responses. Even if this data is later reduced to quantitative indicators, researched stakeholders still have some control over how they have responded to the initial question. This is not the case when quantitative data is gathered directly from researched stakeholders. They are restricted in their responses and have less perceived control over these. We found that at times this could result in the stakeholder assuming that we were evaluating them and not the programme.

The ten research questions developed by the Ministry allowed for both qualitative and quantitative data to be gathered. The questions did not evaluate the programme in terms of failures and successes, but rather aimed to identify examples of best practice and to develop a picture about the "how and what" of curriculum implementation in schools. Most questions encouraged a qualitative response which meant participants could express their own feelings and values in the responses. However, one question aimed to quantify the amount of hours used for planning and teaching in each of the Arts disciplines. During the pilot study it became apparent that this quantitative question felt threatening to school management and teaching staff because participants felt they were being held accountable for their participation in the professional development. The research team needed to convey the contractor's actual intentions, allowing participating schools to see that the Ministry was not evaluating them but rather assessing how the professional development programme was impacting on the way Arts was being taught in their school.

In the case of the Arts PD evaluation the Ministry of Education's values emphasised research that could move education forward, both at the policy level and at the classroom level. As shown, what was of interest is how the researchers and the researched brought to the research different assumptions of what the Ministry valued. For the research team there was an assumption that the Ministry really valued quantitative more than qualitative data. For the providers, principals, and teachers the tension tended to be around the perceived nature of evaluative research – that is, what are evaluations used for and who uses them? At times providers, principals, and teachers were more guarded in their responses in the evaluation because they perceived that the evaluation was on their own performance.

The evaluated's uneasiness

The providers

For many researched stakeholders, evaluation is not seen as a formative (or feeding back) function of programme provision but as an assessment of a programme's worth (Anderson, 1998). It is an action of accountability. In such a climate it is assumed that programmes are evaluated to be closed off and finished. In such an atmosphere, providers of programmes may feel threatened by external evaluators. In such an environment providers may feel that their

personal performance is the object being evaluated. This feeling can have detrimental consequences for the evaluation. For example, the provider may feel led to certain responses when surveyed and/or interviewed resulting in invalid findings, or the provider may feel the need to influence how participants are involved in the evaluation.

In the Arts PD evaluation, the research was designed to help ensure that providers did not feel that their performance was the object of evaluation. The national provision of two models of professional development and the national focus of the evaluation meant that the focus could be directly placed on the two models of professional development. This was emphasised in the research questions provided by the Ministry of Education which focused on the effects of the programme/s in curriculum implementation and not on a regional comparison of the effects of professional development. While all sixteen variations were documented to give an overall context for the professional development implementation, considerations led us to leave out specific comments made by programme participants on specific providers. Only generalisable comments were used in the evaluation.

Within the application of research, procedures were built into the *rfp* and proposal to reduce provider uneasiness. In all stages of the research, providers were kept informed about the research and what the research intended to achieve. However, this was not sufficient to reduce all provider uneasiness.

In the first phase of the research, providers expressed some uneasiness over the requirements of the research (each provider was required to administer a questionnaire to teachers at the beginning and end of each course). Some saw this requirement as adding an extra burden to their job. Consequently, questionnaires were sometimes administered incorrectly and final return rates for the first year were quite low. Furthermore, two of the six providers commented on the external nature of the evaluation, saying that an internal evaluation would be more suited to their perceived needs. To address these feelings and tensions the research team was given several opportunities by the Ministry of Education to talk to providers about the research process and current findings. It was in these forums that the team was able to talk about the value of external evaluative research and to discuss how factors such as consistency across providers added to research validity.

Acknowledging and incorporating provider concerns into the evaluation is one of the largest "stumbling blocks" to a successful evaluation. However, it is also important that the evaluators incorporate the perspectives and concerns of the evaluated programme's many other participating groups or affected groups.

School management, school teachers, and the school community

In a "simple" programme evaluation only the values of participants would be included. However, simplicity is not a feature in education, particularly in professional development. In education, the teacher's professional development is of concern to school management (who pay or contract into the professional development) and the school community (particularly students, but also including caregivers and families). For a professional development strategy to be successful, it must have some impact on each of these parties. For example, in the Arts professional development: change in school management structures would see policy and long-term strategic plans incorporating, at least, the mandatory requirements of the curriculum; classroom teaching would see teachers being confident in, and taking risks with, the new curriculum; and, finally, school communities would benefit through the increased learning and Arts knowledge of students. Each of these stakeholders have differing views on what "they" hope to achieve

through professional development and what they hope to gain through participating in the research.

Consequently, it was important in the Arts PD evaluation for the contractor to include research questions covering the needs of these stakeholders. It was also critical for the evaluation to include the perspective of all the stakeholders. This included students. The effectiveness of the Arts professional development for curriculum implementation could not be ascertained alone by the quantitative data of a measured increase in planned teaching hours or by the qualitative indicators of principal/teacher experiences. For professional development to be successful it must at some stage engage the students. When evaluators omit this group of stakeholders, they leave the evaluation unfinished; without the final voice of those who are affected by the education of programme participants.

It could be argued that it is too early to attempt to ascertain the effects of professional development on students during a professional development programme. However, teachers said they needed to learn using more effective ways of teaching the Arts (Beals et al., 2003). When case studies are used in an evaluation researchers are able to draw out of models of success and so these were purposely added to the *rfp* to allow this teacher need to be met. The inclusion of student voice was also able to be effectively captured in this context.

Addressing the dilemmas: a simple solution to a complex problem

I have spent some time discussing the feelings of uneasiness that can be brought to evaluative research by those participating in the evaluated programme. In some way the words "being evaluated" sum up the problems found in evaluative research. Although the programme is the object of evaluation, the subjects of the programme (the providers, school managers, and teachers) feel they are the ones being evaluated. This is enhanced when research methodologies that gather data from participants are used. The dilemma then is in the researcher's hands – how do we address the feelings of uneasiness found in evaluative research?

Talking about evaluative research is a significant method in addressing the feelings of uneasiness experienced by the participants in evaluative research. "We" as researchers, who are already knowledgeable in the practice and theory of valid and reliable research, often assume those people who also play a part in the research process know what "we" are going on about. That is, we assume participants know that it is the programme that is being evaluated, not them. This may not be the case. Even providers of programmes may feel uneasy about an evaluation when they feel that they are the objects of evaluation.

What was discovered in the Arts PD evaluation was that bringing programme providers "on board" in the evaluation meant discussing what evaluative research is, and how evaluative research could feed into the programme delivery. In short, it was educating the providers about evaluative research, and allowing them to know that they were not personally being evaluated, rather it was the programme. There were several specific strategies used by the research team in the methodology to ensure that providers would feel comfortable and included in the research. These were:
- initial introductions personally made by members of the research team to co-ordinator's representative of each provider;
- email updates of the research;
- on-going emails and telephone conversations about research methodology with each co-ordinator;
- advanced notice of any research requirement (together with the purpose for this requirement);

- telephone interviewing used as the main source of data gathering (as recommended in the *rfp*); and
- use of any opportunity given by Ministry of Education to talk to providers about the research.

This same message needed to be given to schools, both to school managers and to school teachers. When gathering quantitative data the research team learnt that the purposes this data served needed to be communicated to participants. The research team found that participants felt more comfortable offering quantitative data when it had been explained that the data gathered was to help generate a national picture of curriculum implementation.

In the past, allowing those researched to know the purposes for the gathering of data was questioned for reasons of validity (Cook and Campbell, 1979). Today, where it could now be argued that the education sector is over-researched, communication of intentions is needed, not only for ethical reasons of informed consent, but to allay uneasiness and allow for a genuine authenticity of participant response.

Conclusion

Guba and Ross (1989, p. 3) claim that "evaluation is an investment in people and in progress". In the context of today's education system evaluative research is a powerful tool. Through carefully constructed evaluations, educators are able to reflect on themselves and on their practice. It is the role of contractors and external evaluators to ensure that evaluation is not used for the purposes of personal appraisal but that the results are made available for personal reflection on how a programme has affected their practice. Such an evaluation allows for the incorporation of stakeholder values, concerns, issues, and differences. Such an evaluation allows for positive change.

The Arts PD evaluation is one example of a balanced evaluation where the research design attempted to allow reflection and appraisal. In attempting to achieve this the research team at NZCER recognised that an uneasiness does, at times, exist between the contractor and researched stakeholders. In this paper I have discussed the importance of educating those participating in the evaluation about the nature and purpose of research. As evaluation informs more and more political decisions concerning education in our society it is important for researchers to grasp the potential of this type of research and to use it, and report it, and most importantly, teach it, in a way that allows for development.

Endnotes

1   Unfortunately, the space limitations of this paper mean that I will be avoiding discussion of the values brought to the research by the contracted researchers.

References

Anderson. G. (1989). Fundamentals of Educational Research (2nd ed.). London: Falmer Press.

Beals, F., Hipkins, R., Cameron, M., & Watson, S. (2003). An Evaluation of the Professional Development to Support Arts in the New Zealand Curriculum (Evaluation Report). Wellington: New Zealand Council for Educational Research.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis for field settings.* Chicago: Rand McNally College Publishing Co.

Boyask, R. (2001, 6–9 December, 2001). *Understanding the visible in art – the regime of visual arts in the New Zealand Curriculum.* Paper presented at the New Zealand Association for Research in Education Annual Conference, Christchurch, New Zealand.

Guba, E., & Lincoln, Y. S. (1989). *Fourth Generation Evaluation*. Newbury Park, California: Sage Publications.

Ministry of Education. (1999, 15 April 2003). *The Arts in the New Zealand Curriculum: a background paper*, [HTML Document]. Ministry of Education. Available: http://www.minedu.govt.nz/print_doc.cfm?layout=document&documentid=5660&data=l&goto= 00-02&fromprint=y [2003, 8 August].