The character of language program evaluation in Australian university language departmental reviews

Abstract

Recent international research studies on Language Program Evaluation (LPE) show key issues that make the evaluation experience a successful one. But, to what extent do Australian language program evaluations reflect these key issues? The reality of LPE in Australia, particularly in tertiary institutions, shows in fact that it is scantily researched and that its confidential nature makes access to evaluation information a difficult task for researchers. In order to explore the current character of LPE in Australian universities, this study uses metaevaluation as a research approach to conduct analysis of evaluation reports. Its main purpose is to shed light on the present state of LPE in Australian higher education by directing attention towards the need to increase research on evaluation in general, and on LPE in particular, for language educators to obtain a thorough picture about its educational usefulness for improving the quality of language programs.

1. Introduction

Typically, governments and higher education institutions use accreditation bodies and quality assurance agencies to review their programs in order to ensure that they are streamlined systems guided by a set of quality standards penetrating all program components. Some examples of these organizations are the newly established Tertiary Education Quality and Standard Agency (TEQSA) in Australia, the Council on Higher Education Accreditation (CHEA) in the USA, and the International Network for Quality Assurance Agencies in Higher Education (INQAAHE), of which most of the quality assurance and accreditation agencies around the world are members. The most common method that these organizations use to warrant quality standards are up to date is Program Evaluation (PE), with which programs are made accountable for the services they provide. While an evaluation normally consists of gathering and analyzing data about a program’s operations, so that some value judgments can be issued, the particular ways in which PE takes place in higher education institutions in different nations, or even within a single national context, respond to a variety of cultural, political and societal factors.

In Australia, the Languages and Cultures Network for Australian Universities (LCNAU) acknowledges Language Program Evaluation (LPE) among the key issues
that need to be addressed in order to create stronger language programs in the tertiary sector. Amongst the guiding principles listed on its website (see www.lcnau.org/about), Principle 5 in particular makes explicit reference to this by stating that LCNAU “fosters systematic review, reflection and monitoring of improvements in program design and pedagogy for university languages programs, and provides a nation-wide focus for continuous sharing of good practice”.

Moreover, in line with one of the stated objectives of LCNAU—“to encourage and enable university research on and through languages other than English” (http://www.lcnau.org/background)—LCNAU Principle 3 asserts that “continuing research is an inalienable dimension of the maintenance and development of the discipline in Australian universities” (www.lcnau.org/about). Accordingly, these principles reflect the significant role that LCNAU attributes to the ongoing practice of LPE in strengthening university language teaching in Australia. In order to ensure that the practical experience of LPE facilitates knowledge that can be translated into meaningful guidelines that inform good practice, it is essential to conduct research on evaluation to learn about what characterizes the current implementation of language program evaluation in the particular context of Australian universities.

This study focuses on the findings of a preliminary project carried out at the University of Melbourne in 2012. At present, it is being developed into a doctoral dissertation, and has been endorsed by LCNAU in acknowledgement of its relevance for the languages profession in Australia. In this study, research on evaluation is conducted through the metaevaluation of language program evaluation reports from two Australian universities, the aim of which is to shed light on the character of current evaluation practices. To this aim, thus, this study presents the following sections: (1) an introduction on program evaluation (PE), the field within which metaevaluation is set; (2) background information on the crucial need to conduct research on evaluation in general and metaevaluation in particular, both in PE and LPE; (3) a description of the methodology used in the study; (4) a summary of the findings; and (5) the conclusions that were reached as a result of both the process of conducting this project as well as of the findings.

2. Pressing calls for research on evaluation

Research on evaluation has surfaced at specific moments in time along the history of evaluation. However, in the past thirty years, studies dealing with how evaluations are actually conducted have been rare. In fact, Cronbach, Ambron, Dornbusch, Hess, Hornik, Phillips, Walker and Weiner had in 1980 identified the need for “analytic studies of the evaluation enterprise” (Cronbach et al. 1980: 359), which in the first years of the 21st century has gained the attention of writers who emphasize the need to increase research on evaluation in a systematic manner (Christie 2003; Henry and Mark 2003; Mark 2008). Relying on the sparseness of research studies on evaluation and also on the anecdotal quality of insights on the practice of evaluation, Henry and Mark (2003: 70) comment:
Prescriptive advice and admonitions about how to do evaluation have been plentiful, filling books, journals, conferences, e-mails, and conversations. But these are generally based on personal experience, observation, and the individual’s sometimes idiosyncratic beliefs and values—and not on carefully gathered evidence that can be described, shared and critiqued.

For evaluation to advance, empirical evidence describing how evaluation is conducted is required. If an accumulation of studies is created based less on informal accounts and more on systematic data gathering and interpretation, a more solid foundation of knowledge can be built to “inform judgments about what kinds of evaluation activities are most likely to have what consequences under which conditions” (Henry and Mark 2003: 78). How, then, can research on evaluation be systematized? What research methods should be used in order to create this evidence base?

In the evaluation literature, two trends can be identified offering research frameworks that the researcher can use to study evaluations. First, in 2003, Henry and Mark proposed six equally and independently important types of research conducive to formally narrowing the gap between evaluation theory and practice, metaevaluation being one of them. Metaevaluation in Henry and Mark’s view is defined as the evaluation of an evaluation, and it aims to examine a single evaluation or a series of evaluations against a set of standards.

The second framework is Mark’s taxonomy of methodologies (2008) which researchers can use to study evaluations in the context of today’s fundamental issues in evaluation. According to Mark (2008), research on evaluation, though scarce, has been a recurrent issue historically, and it is currently being debated by theorists such as Alkin (2003) and Donaldson and Lipsey (2006). Mark’s (2008) classification is based on the inquiry modes devised by Mark, Henry and Julnes (1999) which were developed as a result of the authors’ own review of methods used in primary evaluation, namely description, classification, causal analysis and values inquiry. At the same time, these four clusters correspond to the framework grounded in a neorealist theory of evaluation known as Emergent Realist Evaluation (ERE) (Mark, Henry and Julnes 1998). Although this neorealist framework of methods was created for conducting evaluation, Mark (2008) argues that it can be valuable for researchers who seek to give an account of the what and how of evaluation practice.

### 2.1 Metaevaluation in Program Evaluation

Several attempts to bring metaevaluation to the fore have occurred in the past as well as in more recent times. Past examples are the ones illustrated by Cook and Gruder (1978) who advocate the use of metaevaluation to improve the technical quality of evaluations, and Gowin and Millman (1978) who discuss the many advantages to metaevaluating specific evaluations, and point out other benefits that go beyond particular contexts. Metaevaluation, for Gowin and Millman (1978), contributes to four areas of research about evaluation: (1) the practice of evaluation; (2) the
concept of value; (3) the criteria for judging evaluations; and (4) the procedures for conducting metaevaluations.

In the program evaluation literature of recent times, evaluation theorists such as Stufflebeam (2000, 2001) have claimed that, as has been shown in the past, things may and do go wrong in the complex process of an evaluation. Therefore, it is imperative that evaluators apply a review process to their evaluative work and include it as part of every evaluation they conduct. The method assigned to this reviewing exercise is metaevaluation, which Scriven introduced in 1967 and defined as “the evaluation of an evaluation”. Stufflebeam (2001: 185) expands this as:

[...] the process of delineating, obtaining and applying descriptive information and judgmental information about an evaluation’s utility, feasibility, propriety, and accuracy and its systematic nature, competence, integrity/honesty, respectfulness, and social responsibility to guide the evaluation and publicly report its strengths and weaknesses.

Metaevaluation, according to Stufflebeam (2000), is in the interest of not only evaluation producers such as evaluators, but also of users. On the one hand, metaevaluation helps evaluators contribute to the advancement of the profession by improving their evaluation practices. On the other, metaevaluation also assists evaluation users to trust evaluation findings which reveal information about the services provided by a given program.

Proponents of metaevaluation such as Stufflebeam and Shinkfield (2007) thus argue for the conceptualization and use of metaevaluation as a way to ensure that evaluations provide “sound findings and conclusions; that evaluation practices continue to improve and that institutions administer efficient and effective evaluation systems” (Stufflebeam and Shinkfield 2007: 649).

Today, as confirmed by Cooksy and Caracelly’s (2009) study exploring the practice of metaevaluation throughout the history of evaluation, while “informal metaevaluation happens all the time” (Wingate 2010: 771), only a few evaluations are metaevaluated systematically. These authors also found that most metaevaluation efforts seemed to have converged after the year 2000, which signals an increased current relevance of conducting metaevaluations. According to Stufflebeam (2001), evaluation needed to reach the level of maturity and development that it currently enjoys as a discipline in order for evaluators to realize the importance of ongoing and systematic assessment of evaluation work.

Mark (2008) additionally suggests the use of metaevaluation as a research approach to contribute to increasing today’s scant evidence base of studies in evaluation. This empirical data is paramount for evaluators to both self-reflect and improve their subsequent evaluations, as well as to create a stronger link between evaluation theory and practice (Henry and Mark 2003). Ultimately, as Mark (2008: 115) remarks, “shouldn’t evaluation itself be open to systematic inquiry, just as the policies, programs, and practices that we evaluate are (Dahler-Larsen 2006)?”
2.1.1 Approaches to conducting metaevaluation

The models offered over the years for conducting metaevaluations are seen through the work of the same authors who argue for the need to metaevaluate evaluation processes. Cook and Gruder (1978), for instance, propose a framework meant to support the then new systematic evaluation research. Identified within the positivistic tradition, their main focus was the quality of methodological procedures in general and particularly on the technical quality in the methods used, which they verified as deficient in the four studies of evaluations that they had conducted. Cook and Gruder’s (1978) framework aims to study only summative evaluations and consists of seven metaevaluation models which they claim do not have to be used independently: (1) a review of an evaluation report; (2) a review of the literature about a specific program; (3) empirical reevaluation of an evaluation; (4) empirical reevaluation of multiple data sets about the same program; (5) consultant metaevaluation; (6) simultaneous secondary analysis of raw data; and (7) multiple independent replications. The concepts of summative evaluation (one that occurs at the end of a program whose goal is to determine the impact of the program), and formative evaluation (which are conducted while the program is running and seeks to help improve the program) were first introduced by Scriven in 1967. Cook and Gruder’s (1978) perception of the scientific superiority of summative evaluations over formative ones clearly reveals how this perception had lingered on after a decade of numerous disappointments in using evaluations based on experimental design and with a focus on the technical quality of their methodologies. This too is an illustrative example of the long-lived quantitative and qualitative methods debate associated with summative and formative types of evaluations respectively which created an important division in the field known as “the paradigm war”.¹

A second proposal for appraising evaluation work through metaevaluation is put forward by Gowin and Millman (1978). It consists of three approaches: (1) standards for evaluation; (2) a checklist of questions for evaluation; and (3) QUEMAC (Question, Event-object, Method, Answer, and Concept: a series of questions seeking to reveal the logical structure of the evaluation). The standards alluded to by Gowin and Millman (1978) anticipate the creation shortly afterwards of a carefully selected list of standards produced by the Joint Committee on Standards for Educational Evaluation (1981) which is now in its third edition (2011). Several checklists have been produced to conduct metaevaluations, such as Stufflebeam’s (1999) checklist based on “The Program Evaluation Standards”, and Scriven’s (2007) “Key Evaluation Checklist”. Stufflebeam and Shinkfield (2007) also provide a framework of eleven generic tasks that metaevaluators can use for heuristically planning their metaevaluations. The procedures for using these tasks are context-dependent and therefore it is the metaevaluator’s job to decide which methods will be more effective within each particular context.

Finally, an additional framework for metaevaluation is proposed by Mark (2008) as a potential approach to conduct research on evaluation. Mark (2008) claims that metaevaluation can involve different kinds of research, and for metaevaluation to
properly belong to a category of research studies, it cannot consist merely of a review of the judgments carried out by an expert. Rather, the types of research studies that Mark groups under metaevaluation are of two kinds: (1) a careful analysis of the apparent consequences of an evaluation, paying special attention to alternative explanations (causal analysis of consequences in relation to the evaluation approach used); and (2) a comparative meta-analysis across several evaluations and examining the relationship between some aspects of evaluation context and the kinds of evaluation activities that were carried out (Mark 2008: 128).

2.2 Metaevaluation in Language Program Evaluation

Language Program Evaluation is a subset of studies within Program Evaluation, which has mainly evolved around the evaluations conducted in English teaching programs. Research on evaluation practice in the context of languages programs in which languages other than English are taught, both in higher education as well as other contexts, has been scant but is recently increasing, as evidenced by Dassier and Powell (2001), Mathews and Hansen (2004), Birckbichler (2006), Morris (2006), Rivera and Matsuzawa (2007), Elder (2009), Kiely (2009), Gorsuch (2009), Alseweed and Daif-Allah (2012), and others. In particular, the recent study by Shawer (2013) represents an instance of not only a retrospective analysis of an evaluation but of a formal metaevaluation in the field of languages at the tertiary level.

The focus of the studies by Shawer (2013) and Gorsuch (2009), for example, is to document and offer a reflection about how LPE was used in order to deal with issues as serious as the threat of closing languages programs. In these two cases, the internal members of the program conduct a self-evaluation and then assess the evaluation, that is, conduct a metaevaluation, to learn about and to calibrate the effects of the evaluation on the program. Reflecting on evaluation practices and reporting publicly the findings of the evaluation, especially of those evaluations which have yielded positive results, represent a necessary activity that needs to be promoted among higher education professionals in languages programs (Norris 2009; Shawer 2013; Birckbichler 2006; Gorsuch 2009). Assessing the evaluation process of a language program has been highlighted as a means by which to reflect on the influence of evaluation on the improvement of a program (Norris 2009; Elder 2009; Kiely 2009), and also to allow program stakeholders to reflect on how evaluation helped them build on their professional skills (Shawer 2013). Patton (2002: 131) also acknowledges that “for our ongoing professional learning, we need to evaluate our evaluations to find out how they are actually used and become more sophisticated about and adept at doing useful evaluations.”

Norris (2006, 2009) emphasizes the importance of analysis and reflection being included when disseminating the results of evaluation in the tertiary languages sector, as this enables language educators to understand all that language program evaluation can offer to the languages profession.
In sum, though instances of research on evaluation practice and specifically of metaevaluations are still few, researchers such as Norris and others are promoting investigation into how evaluations of language programs are conducted. The aim is to instil a proactive predisposition in language professionals to concentrate their efforts on using rather than just doing program evaluation. Shawer (2013: 2886) explores these metaevaluative studies and confirms that “this category of research marks a move from doing to using evaluation and has been dramatically increasing at the turn of the 21st century”. In his study, Shawer metaevaluates a language program evaluation conducted at King Saud University in Saudi Arabia by assessing the impact of program evaluation on (a) program improvement and (b) on the professional development of the staff. The findings from Shawer’s metaevaluation reveal that the evaluation did improve both the program elements and the faculty and administration program skills. These findings resulted in Shawer (2013: 2883) suggesting the incorporation of metaevaluation as “an integral part of program evaluation to help program stakeholders not only do but also use evaluation”.

In the light of the significant leverage that metaevaluation has gained, by and large, in recent times in the evaluation field and in languages education in particular, this study analyzes language program evaluation reports from two different Australian universities using metaevaluation. In conducting metaevaluation, this study responds to the need to accumulate, examine and interpret empirical evidence from evaluation practices, which can help improve subsequent evaluations. The betterment of evaluations can, in turn, become a reliable source of knowledge that can be used to enlighten good practice in languages programs.

3. Methodology

In this study, Mark’s (2008) taxonomy is used as the basis for carrying out a comparative analysis of two evaluation reports from Australian universities. As outlined above, Mark’s framework consists of what he terms the “Four Inquiry Modes” and for the purpose of this study, the two evaluation reports are metaevaluated through the description and classification modes in order to identify and analyze their character. Information on these reports is thus described and classified in relation to the most relevant issues found in five recent international language program evaluations in higher education, as well as in other educational contexts. These evaluation references constitute an important source of advice on language program evaluation practice, and they highlight the following five issues, which Norris (2009) summarizes and presents as practical recommendations and, which according to Elder (2009), will help either enhance or diminish the value of the evaluation process:

1. Teacher participation and involvement in evaluation leads to gaining a more comprehensive understanding of what happens in language programs.
2. The use of a variety of methods for data collection helps with interpreting data from a variety of perspectives.
3. The contextualization of evaluation findings via triangulation of data sources and methods aids in staving off inaccurate and/or politically-motivated interpretations.

4. Style variation for reporting of findings helps to: (a) reach diverse audiences and provide them with accurate understanding about language teaching programs; (b) increase the likelihood that evaluation findings are used for improvements (if needed); and (c) raise awareness about the value of LPE.

5. The role of the evaluator, which in Kiely and Rea-Dickins’ view (2005), has evolved from ‘assessment expert’ to a proactive evaluator who will adopt an educative orientation to his or her work, and engage in learning about the language program and its learners, teachers, social circumstances, and educational approach.

These issues have been found to characterize evaluations that are beneficial and meaningful to second language programs, and the analytical framework for this metaevaluation values and incorporates them as guiding questions in order to conduct the comparative analysis of the reports.

4. Findings

In this section, the questions used for the analysis of the reports are provided and the findings that result from the analysis of the responses to the analytical questions are synthesized. The five questions formulated in the wake of the recommended issues that lead to positive LPE experiences are:

1. Who commissions the evaluation?
2. What is the level of teacher participation?
3. What methodologies are used?
4. What strategies are used to report findings?
5. What is the role of the evaluator?

The pseudonyms Greenhill University (GU) and Mount University (MU) are used in place of the universities’ real names to guarantee their anonymity.

Question 1: Who commissions the evaluation?

The first question used in the analysis focuses on who commissions the evaluation or ‘review’ as it is better known in Australia, as well as on determining what motivated the evaluation and who conducted it. In both cases, the review was commissioned by the actual universities. With respect to the impetus for conducting the evaluation, both were set within quality assurance policies implemented at those institutions. However, whereas GU’s report did not contain any statement in connection to a separate purpose for the review, MU’s report does comment on an explicit concern about the program. The reviewers from MU were asked to specifically target program areas that were thought to need improvement in order to propose recommendations.
As for the selection of the reviewers, at GU the reviewers were internal members of the program in question, while at MU members of the review panel were externally-sourced.

**Question 2: What is the level of teacher participation?**

The second question concerned the level of teacher participation in the reviews. At GU, the teaching staff collaborated on the design of the evaluation, the analysis and interpretation of data, as well as on how to address the recommendations that came about as a result of the review. The teaching staff at MU, in contrast, only took part in the review via a one-off group meeting.

**Question 3: What methodologies are used?**

Question three looked at the methodologies used in both reviews. A wider variety of methodologies were utilised in GU’s review than in MU’s. GU employed focus-group discussions, interviews, mapping analysis, benchmarking, questionnaires and internal assessment instruments; MU, in contrast, examined program documents, and held individual and group meetings. Hence, GU’s evaluation exhibited a list of methodologies not only greater in number but also in sophistication.

**Question 4: What strategies are used to report findings?**

Question four dealt with the means used to communicate findings. The GU report displayed greater diversity in strategies employed than the MU report. Apart from the written report, GU’s quality assurance policy requires that any of the changes that the courses undergo as a result of the evaluation are posted on the language program’s website. This way, the students are made aware not only of the possible new plans for the program which affect them directly, but also of the value of the feedback they provided through their participation in the evaluation. GU’s report, moreover, clearly showed that an ongoing exchange of program information takes place between program members and external stakeholders in order to strengthen the link between the study of languages and career opportunities. The decisions and agreements reached during those meetings lead to a consensual proposal for a plan of action which the program develops as a result of those decisions, which are later incorporated in the final marketing plan designed for the program at the end of the evaluation. The array of formats used to report findings in the GU review elucidates, hence, a contrast with MU’s single written report.

**Question 5: What is the role of the evaluator?**

Question number five explored the role of the reviewers, which in GU’s evaluation consisted of engaging proactively in several evaluation activities as well as developing relationships with students and external stakeholders. The role of the evaluators in MU’s evaluation differs from GU’s in that in the former, the evaluation panel has a limited level of involvement with program staff. This level of involvement is determined by and aligned with what the terms of reference state the purpose of the
evaluation is: to address a specific concern. The purpose of the review thus defines the role of the evaluators; while at GU the evaluators acquire a deep and broad knowledge about the program through their active involvement, the evaluators at MU involve themselves in the evaluation only to the degree that it serves to address that concern. This limitation causes the scope of the evaluator’s engagement with program members to narrow to a minimum. But the question remains: “Can outsiders evaluate effectively without the cooperation and indeed the collaboration of insider staff?” (Weir and Roberts 1994: 23).

In sum, these findings show clear contrastive features between the two reviews, which lead to the characterization of GU’s review as formative and developmental, and MU’s review as summative and strategic. In fact, this outcome reflects the long-debated choice in PE between summative and formative types of evaluations (see section 2.1.1 above), which has equally influenced the development of LPE (Lynch 1996). Traditionally, summative evaluation occurs at the end of a program and seeks to make definite statements about its value or worth. It focuses on outcomes and performance and it helps administrators determine whether the program should continue, grow or be closed down. In contrast, process rather than outcomes is the focus of formative evaluations. Formative evaluation’s central interest is development and it usually takes place while the program is implemented. Although no in-depth discussion about this important distinction can be offered in this study for lack of space, it should be said that this dichotomy, as well as other ones such as product vs process and quantitative vs qualitative, have come to be seen as dimensions of evaluation rather than divides. As experiences in evaluation have shown, “they may be complementary rather than mutually exclusive. [...] All available perspectives may prove valuable for the evaluation of a given program. How they are utilized in a particular setting will naturally depend on the various educational philosophies of the administrators, teachers and students in the program” (Brown 1989: 229). This perspective goes hand-in-hand with the idea that there is not one correct way to conduct an evaluation but that “it is essential that evaluations be conducted in a principled, systematic and explicit manner” (Alderson 1992: 275). Weir and Roberts (1994: 5) also see as beneficial the use of both formative and what they call ‘accountability purposes’ for evaluation: “The information gathered through formative evaluations can too be used to address accountability demands in the sense that summative data can be used to examine the value of a program or aspect of a program from a bureaucratic perspective”. However, they also warn about the limitations of each of these purposes and suggest that on the one hand, insider and outsider perspectives are included in formative evaluations to avoid bias; and on the other, that the political nature of evaluations for accountability purposes, in particular those externally imposed, are counterbalanced by “(1) initiating evaluation activity at an early stage in the development of the project or program, and (2) specifying an on-going role for an evaluation consultant” (Kiely and Rea-Dickins 2005: 66).
Table 1: Summary of findings

<table>
<thead>
<tr>
<th>ANALYTICAL QUESTIONS</th>
<th>GREENHILL UNIVERSITY</th>
<th>MOUNT UNIVERSITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Who commissions the evaluation?</td>
<td>- University - Purpose: Quality assurance policy - Reviewers: internal members of program</td>
<td>- University - Purpose: Quality assurance policy + specified concern - Reviewers: external review panel</td>
</tr>
<tr>
<td>2. What is the level of teacher participation in the evaluation?</td>
<td>- High-level of participation: Designing evaluation Analysing and interpreting data Plan of action from recommendations</td>
<td>- Minimal participation: One group meeting together with rest of program staff</td>
</tr>
<tr>
<td>3. What methodologies are used to gather data?</td>
<td>- Wide variety: Focus-group discussions Interviews Mapping analysis Benchmarking Questionnaires Other internal assessment instruments.</td>
<td>- Limited variety: Examination of documents on program operations prior to review Individual meetings Group meetings</td>
</tr>
<tr>
<td>4. What strategies are used to communicate findings?</td>
<td>- Various formats Written report Program’s website Engagement of external stakeholders in program’s events</td>
<td>- Written report</td>
</tr>
<tr>
<td>5. What is the role of the evaluator?</td>
<td>- Proactive role: (Understand program operations) Deeply involved in evaluation of activities and developing relationships with students and external stakeholders</td>
<td>- Limited role: (Task completion) Limited engagement with stakeholders</td>
</tr>
</tbody>
</table>

CHARACTER OF REVIEW | FORMATIVE, DEVELOPMENTAL | SUMMATIVE, STRATEGIC |

5. Conclusions

The analysis of the two reports using a framework of international research on language evaluation (see Norris 2009) has shown a clear link between the method of implementation of each evaluation and their resulting dissimilar character. How can we explain this diversity? Due to the autonomous nature of universities in Australia, as Lo Bianco and Slaughter (2009) claim, each operation tends to be highly individualized (Nettelbeck, Byron, Clyne, Elder, Hajek, Levy, McLaren, Möllering and Wigglesworth.
thus lacking a coordination which could, otherwise, be used to reinforce an on-going practice of evaluation adapted to the specific needs of language programs. Moreover, the notable difference between both reports conveys an absence of a common framework to carry out reviews in language departments in Australia. It seems that, language program evaluations in the tertiary sector currently take place with neither any internal theorizing practice prior to the review, nor any after it. Additionally, difficulty of access to language review reports due to confidentiality, both in the Australian context and internationally, prevents researchers from discussing LPE openly and more transparently. The underdevelopment of LPE in Australia and the lack of availability of review reports seem to go hand in hand with the lack of a well-established culture of research on language study at university level, as Lo Bianco and Slaughter (2009: 38) point out. These authors also argue that Australian university language teaching requires the formulation and provision of a language policy supported by the government which would help mitigate the current crisis in Australian language departments, as acknowledged in the National Languages Summit of 2007 (Group of Eight 2008). On another note, a number of initiatives represent a clear move to revitalize the notion of evaluation. These include LCNAU, which, as the peak body for tertiary languages in Australia, states the fundamental need to pursue “systematic review of program design and pedagogy” (www.lcnau.org/about). Similarly, scholars such as Scarino and Liddicoat (2009) emphasize the use of PE as an integrated component of language programs, in line with Norris’ view (2009) of a new educative role for evaluation that can help guide good practice.

To conclude, although two reports represent too small a sample to make generalizations about the character of all language program evaluations across the university sector in Australia, the findings from this metaevaluative study allow us to formulate key issues in LPE. Further research on metaevaluation which follows on from this study is already underway. A cross-examination of additional evaluation reports from multiple Australian university language programs will help attain a two-fold goal:

1. A focus on ‘doing’ metaevaluation (emphasizing the research remit of metaevaluation) in order to (a) empirically and systematically study program evaluation in Australian university language programs, thereby addressing the fundamental need to create evidence-based knowledge to help evaluators improve their evaluative work; and (b) contribute to increased research on Australian university language teaching, as emphasized by Lo Bianco and Slaughter (2009).

2. A focus on ‘using’ metaevaluation (emphasizing the practical remit of metaevaluation) to (a) inform Australian language educators about the character of LPE in Australian higher education; and (b) support evaluation practices in second language education in general and Australian higher education languages programs in particular, and offer practical recommendations for conducting evaluations which can guide good practice.
Notes

1. Also known as the ‘paradigm dialog’ (Lynch 1996: 9), it refers to the philosophical implications of conducting evaluation research under a positivistic or a naturalistic approach. In brief, the positivistic tradition focuses on rigorous experimental methodologies and quantitative data as the most reliable means that can yield valid and accurate results, while naturalistic positions emphasize the use of qualitative methodologies on the premise that these can best reveal the complexity of social phenomena. However, the response to the shortcomings of the orthodox adherence to either side of the debate has given way to more eclectic views and even the proposition of a new paradigm under Emergent Realist Evaluation (ERE) (Mark, Henry and Julnes 1999). ERE’s authors claim that the philosophical stance of the researchers must not become an obstacle in their work, and advocate the use of multiple methods depending on the needs of the particular evaluation. A broader and more inclusive view of evaluation has also been acknowledged in LPE by authors such as Lynch (1996) and Weir and Roberts (1994: 132) who argue that “the purpose of evaluation should override such quasi-ideological preferences in favour of principles of utility and relevance (Patton 1986)”.

2. Meta-analysis—though often used as a synonym of metaevaluation, as it seems to be in this case—really consists of the aggregation and analysis of quantitative data from multiple research studies investigating the same phenomena (Cronbach et al. 1980; Stufflebeam 2001; Wingate 2010).

References


