

A Systematic Approach to Evaluating Teaching and Learning Initiatives in Post-secondary Education

Cindy Ives
Athabasca University

Lynn McAlpine
University of Oxford

Terry Gandell
T.G. Pedagogical Consultant

ABSTRACT

This article describes a research-driven heuristic for the scholarly evaluation of teaching and learning interventions, which is systematic, collaborative, and discipline focused. We offer this guide to educational developers and other instructional support staff who are tracking the impact of interventions in teaching and learning with academic colleagues who lack backgrounds in educational evaluation or social-science research. Grounded in our experience in three different faculties, the framework may be modified to meet the needs of other contexts and disciplines. To aid such modification, we explicitly describe the thinking underlying the key decision-making points. We offer practical advice that may assist academics and academic developers with evaluation processes, thus addressing the scarcity in the literature of comprehensive, programmatic, scholarly, and systematic assessments of innovations in teaching and learning at the university level.

RÉSUMÉ

Dans cet article, nous décrivons une heuristique fondée sur des recherches ciblant une évaluation érudite des interventions en enseignement et en apprentissage tout en étant systématique, collaborative et axée sur cette discipline. Nous offrons ce guide aux concepteurs de programmes pédagogiques ainsi qu'au personnel de soutien pédagogique impliqué dans le suivi des répercussions provenant des interventions en enseignement et en apprentissage assurées par des collègues académiques n'ayant ni l'expérience en l'évaluation pédagogique ni en recherche en science sociale. Fondé sur notre expérience dans trois facultés distinctes, ce cadre peut se modifier, s'adapter afin de répondre aux besoins de d'autres contextes et disciplines. Afin de faciliter une telle modification, nous avons explicitement décrit la logique sur laquelle reposent les points décisionnels clés. Nous offrons des conseils pratiques afin d'assister les académiques et les concepteurs de programmes académiques avec le processus d'évaluation. Ainsi nous adressons la question de rareté d'évaluations compréhensives, programmatiques, érudites et systématiques des innovations en enseignement et en apprentissage au niveau universitaire.

CONTEXT

As Barnett (2000) so pointedly commented, we live in an age of super-complexity, in which demands for change have become a constant in publicly funded higher education systems. Some recent examples are the external calls for accountability in Australia (Robertson, 1998); the creation of the Quality Assurance Agency for Higher Education in the United Kingdom (Randall, 2001); and the requirement of the Bologna Agreement in Europe for universities to re-articulate their programs. Canadian examples include pressures for the incorporation of online learning in post-secondary education (Advisory Committee for Online Learning, 2001); recommendations for reform in Ontario's higher education system (Rae, 2005); and increasing public concern about the ability of post-secondary education to meet the future learning needs of all citizens (Canadian Council on Learning, 2006). These external forces are compelling universities to make substantial changes at the same time as they deal with reduced resources, increased accountability, technological challenges, and more informed students and parents. Thus, evaluation¹ of the impacts of teaching and learning initiatives is of increasing concern to administrators (chairs, heads of departments, and deans), as well as to instructors² and academic developers.

Academic development (or, alternatively, educational or faculty development) is an evolving field that aims to improve teaching policies and practices – on the assumption that such improvements will ultimately enhance student

learning (Brew & Boud, 1996; Candy, 1996). At the individual level, it comprises a range of professional teaching development activities (e.g., workshops, consultations) for faculty members, frequently offered by staff with pedagogical expertise and organized in teaching and learning centres (see, e.g., McAlpine, 2005; McAlpine & Cowan, 2000; Professional and Organizational Development Network in Higher Education, 2007; Saroyan & Amundsen, 2004). At the institutional level, it often includes providing grants and awards for teaching excellence, committee work, and policy development. Moreover, the knowledge that student approaches to learning are influenced by the totality of their experiences (Ramsden, 1992) has stimulated a shift of focus in educational development from the course to the program level. Evaluation activities may involve not just individual professors but entire departments and faculties in system-wide projects. A comprehensive approach to educational development is thus systemic, offering individual-, program-, and institutional-level activities (McAlpine & Saroyan, 2004).

Among academic developers, there is growing recognition of the critical influence of disciplinary variation (Becher & Trowler, 2001), for instance, on knowledge structures (Neumann, 2001), on modes of research (Johnson & Broda, 1996), and on learning tasks and student assessment (Pace & Mittenfiorf, 2004). Because academics are directly associated with the students, learning tasks, and subject matter in specific learning environments, they are well positioned to define what to examine, change, and evaluate (Mentkowski, 1994). Thus, we view our role in working with them as providing a scaffold for jointly exploring the aspects of teaching and learning that can be most meaningfully evaluated in particular contexts (McAlpine et al., 2005) and, during this process, ensuring that evaluation focused principally on teaching is still situated within a learning perspective. Yet another role of the academic developer can be to inspire or facilitate critical reflection on teaching practice (Boud, 1999); although not central to the purpose of the framework described here, reflection on instructor and student conceptions of teaching and learning may occur (Kember, 1997; Land, 2001; Samuelowicz & Bain, 2001; Triggwell, Prosser, & Waterhouse, 1999).

Following Jenkins (1996), we have developed what we call a “discipline-based” approach to faculty development (McAlpine & Saroyan, 2004, p. 218), an approach that has underpinned our work for a number of years and is described in detail elsewhere (McAlpine & Cowan, 2000). Increasingly, our approach is linked to the notion of academic development as a collective task of a learning organization (Candy, 1996). This article describes a research-driven heuristic for scholarly evaluation of teaching and learning that goes beyond course-level analysis to program-level analysis. Varying aspects of this heuristic were developed, implemented, or applied in development activities that took place in faculties of Medicine, Agriculture, and Management; the specific evaluation activity that led to formally documenting this heuristic was understanding the impact on teaching and learning of different technologies in a Faculty of Engineering.

We view institutional leadership activities, of which evaluation is a part, to be not only ongoing and systemic in approach but also incremental in their impact on members of the university community, including students and faculty (Candy, 1996; Fullan, 2006; Land, 2001; McAlpine & Saroyan, 2004). Thus we use, eclectically, multiple models and methods of evaluation, with attention to multiple stakeholders (Calder, 1994; Johnson & Onwuegbuzie, 2004; Stufflebeam, Madaus, & Kellaghan, 2000).

Our systematic, collaborative, and discipline-based approach involves extensive, ongoing discussion and action with academic colleagues. It begins with a discussion to define the nature of the evaluation, the goal of which is to lead to a design that ensures that the problems, questions, and mechanisms for addressing the inquiry are defined in appropriate ways from departmental and disciplinary perspectives. As academic developers, we provide the educational research expertise that our colleagues may lack. Since we are increasingly aware of the impact of teaching approaches on learning (Kember, 1997), we view our work as a vehicle for understanding and improving professional practice, both theirs and ours. The collaboration is only possible when all participants share a common commitment to a scholarly, evidence-based approach to understanding teaching and learning (Boyer, 1990; McAlpine & Saroyan, 2004; Shulman, 2000; Weston & McAlpine, 2001), which is a challenging process in the context of institutional influences on individual teaching practices (McKinney, 2007). More is required of our disciplinary colleagues since they are actively involved in analyzing a process of change while experiencing it – and may not be familiar with the data collection and analysis approaches being used. It also requires more of educational developers in terms of a) being responsive to the concerns, decisions, and practices of those with the most invested, that is, instructors and students (McAlpine, 2005), and b) being prepared to facilitate not just the development of pedagogy but also that of educational inquiry.

When we began the evaluation project described here, we realized that we had been increasingly involved in these kinds of collaborative activities (e.g., Gandell & Steinert, 1999; McAlpine et al., 2005). However, although we were using program (Calder, 1994) and formative (Tessmer, 1998) evaluation methodologies to explore questions about teaching and learning in our context, there were few accessible heuristics³ to provide a scaffold for our work. We were independently drawing on our accumulated tacit knowledge, acquired through training and experience in educational research and formative assessment methods (e.g., Cresswell, 2003; Denzin & Lincoln, 1998; Gall, Borg, & Gall, 1996; Guba & Lincoln, 1989; Leedy & Ormrod, 2001; Weston, 1986; Yin, 1984), a situation that led us to analyze and document the process we had first used in Medicine and Management and then in Engineering to derive a heuristic that could be useful for other evaluation projects. This article, which describes that process, has been written as a guide for educational developers and other instructional support staff who are involved in evaluating the impacts of interventions in teaching and learning with disciplinary colleagues

in collaborative and systematic ways. We focus particularly on issues that arise when working with colleagues who lack backgrounds in educational or program evaluation. Because the heuristic raises questions at different decision-making points about the whys and hows of doing evaluation that may prove effective in post-secondary environments, it may be used to guide and structure an evaluation process. We have found it particularly valuable in our context. Examples provided in this article illustrate how the heuristic has contributed to our understanding of changes in teaching practices and how it has been taken up and disseminated beyond the faculty members for whom it was originally intended.

OBJECTIVE

As educational developers, our long-term goal is to encourage pedagogical improvement in support of enhanced student learning. Clearly, it is difficult to relate improved learning to specific teaching interventions (Cronbach, 2000; Haertel, & Means, 2003), but we believe that teaching improvements contribute over time to better learning opportunities and environments for students (McAlpine & Weston, 2000). Although evidence of the impact of our academic development initiatives will take time to accumulate (cf. Fullan & Stiegelbauer, 1991), we evaluate and report regularly as part of our professional commitment to a scholarly approach to our work (McAlpine, 2000; McAlpine & Saroyan, 2004; McKinney, 2007; Shulman, 2000).

Our objective here is to provide guidelines for evaluating pedagogical initiatives, in a heuristic we believe may be applicable to working with faculties. The framework may be used or modified to meet the needs of a variety of contexts or disciplines, since we explicitly describe the thinking underlying the key decision-making points. As in qualitative or mixed-methods research, our framework allows readers to judge its applicability to their particular contexts (Guba & Lincoln, 1989; Yin, 1984). The following questions guided our development of the heuristic.

- What are the important initiating processes and factors involved in our systematic discipline-based evaluations?
- How do we sufficiently clarify evaluation goals? How do we work toward collective agreement on those goals?
- How can we ensure rigour in our evaluations to support the value of our findings for a range of stakeholders?
- How can the process of evaluation be educational in the best sense for us and for our colleagues?

THE HEURISTIC

We use the term “heuristic” to mean a set of questions and guidelines to be used in decision making. Our description of the process is divided into a series of 10 overlapping steps, beginning with building the team, clarify-

ing the need, and moving through the project's design and implementation phases, to analyzing and disseminating the data, and finally to understanding the potential uses of the results. The end result is a documented, systematic evaluation process. Although the steps in the heuristic are analogous to those in social-science research and educational evaluation generally (see Appendix 1 for a table that compares the steps in several models), they have a particular emphasis on supporting faculty members' ownership of the process (Mentkowski, 1994). Program, formative, and other evaluation models describe the steps and tools involved (Calder, 1994; Joint Committee on Standards for Educational Evaluation, 1994; Kirkpatrick, 1998; Stake, 2004; Stufflebeam et al., 2000; Tessmer, 1998; Weston, 1986), and there is a wealth of other resources available to help guide the process (e.g., Cresswell, 2003; Gross Davis, 1994; Rossi & Freeman, 1985; Stufflebeam & Webster, 1994). *The Program Evaluation Standards* (Joint Committee, 1994) provide utility, feasibility, propriety, and accuracy principles in support of "useful, feasible, ethical and sound" (p. xviii) evaluation of educational programs. And yet, none of these explicitly addresses the types of questions evaluators need to ask each other during the process. Thus, we have chosen purposefully (Rossi & Freeman, 1985) from among multiple methods (Stufflebeam, 2000a) to help improve teaching practice in support of an enhanced learning environment for students.

Each of the 10 sections that follow offers a definition of the step and a description of the thinking that underlies that step. Where applicable, we include references to the relevant program evaluation standard set out by the Joint Committee on Standards for Educational Evaluation (Joint Committee, 1994). We have used this heuristic in several disciplines, but since our experience in Engineering was the first in which we explicitly shared it with academic colleagues, we limit our examples to those discussed with them. For each step, the heuristic (see Appendix 2) provides goals, lists questions for evaluators to ask themselves and other stakeholders, and suggests criteria to consider before moving on to the next step. The article has been organized in this way to create a job aid, in the form of guidelines that may assist in the sustainable evaluation of teaching and learning – if they are modified to the particular contexts in which the heuristic is used.

Before continuing, however, we must note an important caveat: the heuristic assumes existing relationships among academic developers and faculty members. In our case, we had been working in a particular Faculty of Engineering for five years and had learned to understand and negotiate its disciplinary community (documented in McAlpine et al., 2005); this is not always the case. So, we wish to emphasize the importance of investing in a period of learning before beginning an evaluation process. This advice, which is critical in establishing relationships of trust and acquiring sufficient knowledge of the particular context, confirms that offered in the literature of naturalistic evaluation (Williams, 1986), systematic evaluation (Rossi & Freeman, 1985), and program evaluation (Joint Committee, 1994).

1. *Building the Team*

The goal of this step is to build an evaluation team that represents the various constituencies (Madaus & Kellaghan, 2000; Patton, 2000; Stufflebeam, 2000b; Utility Standard 1 – Stakeholder identification [Joint Committee, 1994]). For the educational developer, an intimate understanding of the disciplinary culture, as well as the specific history of teaching and learning initiatives, pedagogical attitudes, and relationships among the individuals in the context, is essential to the development of a useful and workable evaluation plan (Utility Standard 2 – Evaluator credibility [Joint Committee, 1994]). By taking time to develop the relationships that will support the evaluation activities throughout the study, academic developers learn to navigate an unfamiliar system and to function effectively within disciplinary norms. To initiate the process, developers need to ask questions that clarify how things get done and by whom (see Appendix 2, Step 1). It is important to avoid perceptions, such as those reported by Wiggins (1998), that external evaluators interfere with the integrity of the teaching and learning system.

Example. In our case, the relationship between instructors and academic developers evolved into a common understanding of the context and a mutual respect for the significance of the work they were doing together to improve teaching and (hopefully) learning in Engineering (McAlpine & Cowan, 2000; McAlpine & Saroyan, 2004). Since this resulted in shared accountability and decision making in the development of the evaluation plan, it was relatively easy to create an evaluation team that included members of several Engineering departments as representatives of the Faculty’s academic priorities. Some members of the team were responsible for driving the evaluation, while others participated in individual studies. We worked with them individually and collectively to explore their needs and to design a discipline-appropriate evaluation process that instructors could later use themselves (Utility Standard 7 – Evaluation impact [Joint Committee, 1994]).

2. *Clarifying the need*

Because needs are sometimes first expressed as complaints or concerns from students, instructors, or administrators, the open communication and trust that evolve from having a team that represents the context are vital for exploring and confirming the actual need. As these concerns emerge, a needs-assessment approach (Gall, Borg, & Gall, 1996; Rossi & Freeman, 1985; Stufflebeam, 2000b) supports the development of consensus on the overall evaluation goals. This process allows the disciplinary community to gain a new awareness about the gaps that need to be addressed, in which departments these gaps are found, and whether they are at the class, course, or program level; with this knowledge, decisions can then be made about the types of evaluation to be done (Utility Standard 3 – Information scope and selection [Joint Committee, 1994]). To determine the needs, evaluators can ask direct questions of the various stakeholders.

Assumptions about the goals of evaluation and the critical elements of the teaching and learning processes must be articulated and shared early in the study. Since evaluation implies comparison, gathering baseline data about the current state of teaching and learning practice, in the unit specifically as well as in the discipline generally, will help in later decision making. These data can be found by directly asking members of the community what their needs are and by searching the literature for relevant evaluation reports (see Appendix 2, Step 2).

Example. Many Engineering instructors had participated in campus-based activities that focused on faculty development, course design, and teaching effectiveness. There was an ongoing perceived need – sometimes expressed as frustration in the Faculty Committee on Teaching and Learning or as questions to academic developers – for assistance in evaluating the impacts of the changes or potential changes promised by the use of technology on their individual teaching practice. How were professors determining the impact of the teaching-improvement initiatives in their courses and initiatives across the Faculty? How was this accomplished in other universities? At the same time, the Faculty was engaged in a five-year planning process and was seeking baseline data for decision making on where to invest resources for future initiatives, especially in technology-supported teaching and learning. For instance, one professor was using the quiz feature of a course-management tool (WebCT) to determine if that feature had any impact on student learning, other professors were curious about the impact of using other technologies, and the Dean wanted to know if resources should be allotted to these and similar initiatives. Needs had to be articulated and evaluation methods designed to address the variety of questions to be answered. As academic consultants to the Faculty, we served as a sounding board for instructors and were able to model the conversations with all stakeholders as a way of clarifying the needs. Consequently, we were able to arrive at a broad, overall agreement on the needs of individuals and the Faculty as a whole.

3. *Setting evaluation goals*

Although a discrete step in our process, setting goals flows directly from conversations with stakeholders about the needs emerging from the current state of affairs (Patton, 2000; Stufflebeam, 2000a; Stufflebeam, 2000b). A collaborative decision-making process within the team is intended to ensure that the evaluation goals are valued, achievable, and of potential use. It can also provide guidance for the development of instruments in the later stages of the evaluation (Accuracy Standard 3 – Described purposes and procedures [Joint Committee, 1994]). To gather evidence about possible goals, evaluators – both instructors and educational developers – should ask themselves and their colleagues a series of questions such as those in the heuristic to help structure the collaborative deliberations of the evaluation team (see Appendix 2, Step 3).

Needs clarification (step 2) and goal setting (step 3) are critical for many reasons. The process of extended discussion allows a common language to develop among the members of the evaluation team, a language that evolves from learning about other stakeholders' priorities and expectations as all work toward agreement on the goals (Propriety Standard 2 – Formal agreements [Joint Committee, 1994]). Indeed, engaging a learning orientation is an academic development goal in itself (Kember, 1997; McAlpine & Saroyan, 2004; Samuelowicz & Bain, 2001). In other words, the process of identifying key variables and agreeing on which aspects will be evaluated is focused, as much as possible, on learning as one of the long-term interests. Articulating everyone's expectations in the form of specific learning-oriented goals (McKinney, 2007; Rossi & Freeman, 1985) provides a solid foundation for later stages of the evaluation and helps to ensure that the entire activity is rigorous (Lincoln & Guba, 1986) and scholarly (Shulman, 2000; Weston & McAlpine, 2001). The perception of shared value is key (Utility Standard 4 – Values identification [Joint Committee, 1994]) if other academic colleagues are to invest in the project and if the results are to be effectively used both inside, and outside, the Faculty.

Example. In our case, the team identified three goals for the evaluation project. The first was at the Faculty level: to identify instructors' current uses of and concerns about technology. This would document current practice and provide baseline data for future decision making. The second goal was at the course or program level: to enhance teaching and learning by using technology effectively in specific cases (see step 4). This would respond to needs expressed by individual instructors. The third goal was, again, at the Faculty level: to develop an evaluation guide that could be adapted and re-used in other contexts.

4. *Designing the studies [as in Appendix 1 and 2]*

Using contextual experiences to help focus specific evaluation questions ensures rigour and alignment between the project design and the identified goals (Utility Standard 3 – Information scope and selection [Joint Committee, 1994]). During the design stage, the evaluators work collaboratively to outline the methodology, establish the setting, and select the participants, as well as to confirm the availability of the human, time, and material resources needed to carry out the project (Patton, 2000; Stake, 2000). Williams (1986) offered, for example, a set of questions to help evaluators decide whether a naturalistic approach would be appropriate. The questions that structure the decision-making process in the design step (see Appendix 2, Step 4) help to clarify what will be done and how it will be done.

A key factor for the team to consider in the design phase is how the results will be used (Utility Standard 7 – Evaluation impact [Joint Committee, 1994]). Knowing how the findings will be communicated and their anticipated benefit to other instructors in the Faculty and beyond helps evaluators make appropri-

ate decisions about the methods to be employed in the project. The collaborative process of deciding on design details opens up conversations about students and learning among educational developers and professors. In these conversations, alignment of the expected learning outcomes, teaching strategies, and assessment activities in individual courses can be reviewed, and other faculty development aims (such as course design) can be addressed (Saroyan & Amundsen, 2004).

Example. For each of the three goals, our team generated implementation strategies. For the first goal, we determined that a basic survey sent to all instructors in the Faculty would be most appropriate. The second goal (determining the effective use of technologies to support teaching and learning) was much more complicated and required a multi-pronged approach. To achieve this goal, we decided to conduct several concurrent and complementary evaluation studies in different departments. Although the specific departmental goals varied, they aligned with the higher-level Faculty goal of integrating technology effectively. At this stage we worked with individual instructors to identify and describe the specific technological or pedagogical intervention that would help them answer their questions about student learning in their course. We matched technologies to specific courses; for instance, we examined the use of PC tablets in a design course, where sketching was a part of the requirement. We planned the documentation of each implementation in the best context possible to get the most complete picture of how its use could be most effective. This meant that the broad goal of enhancing teaching and learning by using technology led – through our goal-setting conversations – to the more specific objectives of enhancing teaching and learning in selected courses by integrating particular technologies effectively. For the third goal, we decided to carefully record the steps in our evaluation process as the foundation for building the heuristic reproduced here.

5. *Gaining ethical approval*

This step involves informing the evaluation team about the guidelines for the ethical treatment of participants in educational evaluation and writing an application for approval of the design (Leedy & Ormrod, 2001; McKinney, 2007; Propriety Standard 3 – Rights of Human Subjects [Joint Committee, 1994]). Many faculty are familiar with this process (for a description of the purpose and goals of the evaluation project, an explanation of the study design, participant consent forms, and draft instruments, see *Tri-Council Policy Statement* [Public Works and Government Services Canada 2003]), but this is not always the case. For instance, some disciplines do not typically use human subjects in their research (e.g., Structural Engineering; polymer studies in Chemistry); other disciplines may not consider an ethical review, like those done in more-formal research studies, to be a requirement for getting student feedback on learning and teaching. We believe the questions included in our heuristic will assist with both ethical considerations and instrument design (see Appendix 2, Step 5).

Since students have legitimate concerns about any instructional activities that might compromise their learning, their interests must be protected through the process of informed consent (Propriety Standard 3 – Human interactions [Joint Committee, 1994]). The written application for ethical approval provides instructors with an opportunity to focus their attention on the learners and to confirm that the evaluation project corresponds to their original intentions.

Example. Our collaborators in Engineering were not familiar with conducting inquiries that involved other humans, so they needed time to think about this step.⁴ In our course-level evaluation studies, we focused on specifying how to protect students participating in pedagogical interventions. We took care to ensure that their feedback would be both voluntary and anonymous, whether their responses were collected on paper or online. Similarly, by encouraging instructors to respond voluntarily and anonymously to the survey on technology use, we ensured there would be no implied criticism of those not responding or not using technology in their teaching.

6. *Developing the evaluation instruments*

The process of building detailed and unambiguous instruments (e.g., questionnaires, interview protocols, tests) further clarifies stakeholder expectations of the evaluation and allows confirmation of the project design (Rossi & Freeman, 1985). This step requires careful thinking about how to translate the evaluation goals into data-collection instruments that elicit the information needed to answer the questions (Accuracy Standard 4 – Defensible information sources [Joint Committee, 1994]) (see Appendix 2, Step 6).

This step may be more or less collaborative, depending on the types of data-collection methods that are envisioned. Academic developers may write interview questions, while instructors construct learning-assessment items and, together, they may develop survey questions that assess students' responses to a new teaching strategy. This process offers an opportunity to develop and test measures that will provide data to answer questions about the impact on student learning (Accuracy Standard 5 – Valid information [Joint Committee, 1994]). Pilot testing each measure, collecting sample data, and then analyzing the data to determine if the measures yield appropriate information from which conclusions can be drawn are essential to the success of evaluation studies (Accuracy Standard 6 – Reliable information [Joint Committee, 1994]). Developing instruments also offers educational developers another opportunity to learn more about the teaching and learning context in an unfamiliar discipline.

Example. Our second goal (to enhance teaching and learning by implementing technology effectively) implied assessing the impact of specific technologies in selected courses. This involved developing several types of data-collection instruments, including tests measuring student learning, questionnaires on student attitudes, and protocols for interviews with students and professors. For two sections of a course in thermodynamics, we initially planned to use

the same instruments and procedures: instructor interviews, student grades, student questionnaires, and usage data from the course-management system. However, because they were accustomed to experimental research in the lab, the instructors wanted to control for all variables affecting student learning and, ultimately, to draw conclusions based on comparisons between classes. After consulting with us about the value of different kinds of data (e.g., quantitative vs. qualitative), the individual instructors decided to adjust the student questionnaires to suit their particular teaching goals and testing methods. This is an example of how context can play an important role in the design and interpretation of evaluation studies. After some discussion, it was agreed that establishing causal relationships based on statistical significance was neither the goal nor an appropriate design for this project (Cronbach, 2000). Together, we concluded that the data would provide patterns indicating trends, rather than showing direct links between teaching practice and student learning. Some of our Engineering colleagues were unfamiliar and uncomfortable with this way of using and interpreting data.

7. *Collecting the data*

This step involves collecting data from various quantitative and qualitative instruments (Rossi & Freeman, 1985). The questions to be asked by evaluators in this step (see Appendix 2, Step 7) relate to how carefully the specified procedures are followed (Accuracy Standard 6 – Reliable information [Joint Committee, 1994]).

The collaborative process of designing the study and clarifying its purpose helps to make faculty members more aware of students' potential reactions to instructional activities. Indeed, students' active participation in the process of data collection may enhance instructors' awareness of the importance of learner perspectives, leading them to seek student feedback more regularly and to adjust their teaching as a result. The evaluation process may in turn enhance instructors' pedagogical understanding and increase their range of options for teaching and assessment strategies (and perhaps lead them to question their previous practices). Earlier team conversations about acceptable methods (e.g., consistency, replicability) for educational inquiry notwithstanding, last-minute changes in classroom delivery may be required. Thus, another benefit of participating in the process may be increased flexibility in instructional practice.

Example. We knew that an anonymous survey was necessary to achieve our first goal, that is, to encourage as many instructors as possible to report what technologies they were using in their teaching. We had questions about the type of survey, how to distribute it, and how to motivate instructors to complete it. After consulting with several representatives of the Faculty, we determined that a paper-based questionnaire, sent with return envelopes to all instructors through campus mail, would respect traditional practice and encourage participation. We wanted the data to be representative of the population

of Engineering instructors so we could have confidence in our findings. We sent one reminder by email but did not further pressure them for responses. A higher than anticipated response rate (48%) to this approach assured us that we had accurately assessed the climate with respect to technology use among instructors. The responses ranged widely in perspective; they were thoughtful and complete, indicating we had collected high-quality data (Patton, 2000). Since the respondents were representative with respect to the reported use of the Learning Management System (73% of respondents said they used it; system data showed 70% used it), we concluded that the results were valid enough to serve as benchmarking data (Ives, Gandell, & Mydlarski, 2004). Not only did most instructors answer all the questions but they also offered numerous specific comments on the issues raised in the survey about their uses of technology. In addition, because the instructions indicated that this survey was the beginning of an ongoing process, they provided feedback and made general suggestions about the wording of the questions, the scales used, and the survey tool itself. We were able to use this feedback as formative assessment to improve the survey and to recommend changes for its future use.

8. Analyzing the data

Because all members of the evaluation team may not be directly involved in collating and analyzing the data, all team members must ask themselves a series of questions about validity/credibility/transferability, reliability/dependability, and interpretability/confirmability (Accuracy Standards 7 – Systematic information, 8 – Analysis of quantitative information, and 9 – Analysis of qualitative information [Joint Committee, 1994]; Lincoln & Guba, 1986) (see Appendix 2, Step 8). Due to the number of steps involved in transcribing, calculating, integrating, and displaying the results of multi-method inquiries (Stufflebeam et al., 2000; Tashakkori & Teddlie, 1998), this step can take much longer than academics anticipate. Thus, it is essential at the outset of the project to both make the timeline clear and get preliminary results back to the evaluation team as quickly as possible for discussion, before other projects and priorities intervene. When studies are done at the class or course level, instructors need timely formative feedback for planning future classes (Utility Standard 6 – Report timeliness and dissemination [Joint Committee, 1994]).

Example. Since our survey of Engineering instructors' technology use and concerns used a five-point Likert-type scale, it was easy to quickly produce descriptive statistics for each question. We were able to do both frequency and correlation analyses to look for patterns in reported beliefs and behaviours. In some cases, in response to feedback on the scales we used, we collapsed the five-point scale to three points to facilitate interpretation, which simplified the tables, charts, and histograms we designed to represent the findings graphically. For open-ended comments, we engaged several coders (graduate students in educational development), who independently assigned categories to the responses. The categories were then compared across the coders as a test

of inter-rater reliability (Leedy & Ormrod, 2001; Tashakkori & Teddlie, 1998); as a result, some comments could be represented quantitatively as response frequencies (Ives, Gandell & Mydlarski, 2004).

9. Interpreting and reporting results

In this step, academic developers work closely with participating faculty members to interpret the results of the evaluation (Accuracy Standard 10 – Justified conclusions [Joint Committee, 1994]; Patton, 2000; Stake, 2000). The findings can then be documented and shared with all stakeholders, and together the members of the evaluation team draft conclusions and recommendations. In order to produce reports (Ives, Gandell, & Mydlarski, 2004) that meet the needs identified at the beginning of the project, this process can be structured to ask questions about how the data relate to the stated evaluation goals (Utility Standard 5 – Report clarity [Joint Committee, 1994]) (see Appendix 2, Step 9).

The continuing importance of collaboration and consultation is evident in this step. Instructors may interpret results in their disciplinary context in ways that are meaningful to them, and academic developers can help them reflect on their teaching practices (Accuracy Standard 11 – Impartial reporting [Joint Committee, 1994]; Boud, 1999; Weston & McAlpine, 2001). Collective interpretations, conclusions, and recommendations may guide practice and decisions at the course, program, and faculty levels. Academic developers can not only explain the difficulties inherent in making causal conclusions in this type of inquiry but also help all stakeholders use the results appropriately, considering the complexities and constraints of the specific contexts (e.g., as formative feedback to improve teaching and learning in the discipline).

Example. In our project, we began by interpreting the results of each facet of the evaluation with the appropriate participants. We shared the findings with them and together discussed the meaning of those findings. For instance, some instructors of the courses using new technologies concluded that students had learned the material better than in previous years and pledged to do more to support learning in future semesters. We then wrote and circulated draft reports for review and feedback. In some cases, there were several conversations and extensive reflection by instructors on the results before final conclusions could be confirmed. Finally, after further discussion with the team, academic developers compiled the full reports (Ives, Gandell & Mydlarski, 2004), integrating the results of all the studies.

10. Disseminating and using the results

Rossi and Freeman (1985) noted that “evaluations are undertaken to influence the actions and activities of individuals and groups” (p. 51). In this step, all members of the evaluation team review the results from their particular perspectives (Utility Standard 7 – Evaluation impact [Joint Committee, 1994]). Instructors may consider how to adapt their teaching to the feedback they have

received from students. Through conversations with academic developers, they may learn not only to integrate new ideas into their teaching practice but also how to continue the process of inquiry into its effectiveness. As well, administrators have access to data sources for consideration in their decisions about resources and instructional priorities. Academic developers have new disciplinary-appropriate expertise for future activities and more information and experience to share with collaborators in other disciplines. All stakeholders review and contribute to the final reports and publications, which are targeted to the various audiences that could benefit from the new knowledge (Utility Standard 6 – Report timeliness and dissemination [Joint Committee, 1994]). The questions that guide decision making in this step focus on how to use and share the results of the evaluation broadly (see Appendix 2, Step 10); for example, individual members of the evaluation team can ask themselves how they might apply the findings to their own practice.

At the end of an evaluation process, the first thing we want to know is, did we reach our goals? The answer is unlikely to yield a simple yes or no, and ongoing discussions will be necessary to determine how the various stakeholders respond to the findings and how they influence or report those findings. Their responses can serve as yet another data source among the many considered for decision making. The process of evaluation and application of results is complex and not necessarily rational, as it requires an understanding of the constraints of individual contexts. However, we believe that a collaborative assessment of the evaluation results helps members of the evaluation team make more comprehensive and useful recommendations about policy and future practice (Calder, 1994).

Example. Our evaluation sponsor, the Dean, concluded that the Faculty's three evaluation goals for this project were met. Instructors' concerns about the uses of technologies for teaching and learning were identified, documented, and shared with the disciplinary community (goal 1). Several specific pedagogical and technological interventions were assessed from both student and instructor perspectives (goal 2). Participating faculty members expressed enthusiasm for continuing their initiatives and for adjusting them in light of student feedback. Our recommendations for ongoing integration of technology in the Faculty teaching and learning plan were accepted. And we produced an evaluation heuristic that was accessible to the Faculty and reusable (goal 3). The usefulness of our results to the Faculty thus met not only the general quality indicator of "active utilization" (Lincoln & Guba, 1986), by contributing to organizational decision making, but also the three standards of utilization of evaluation results proposed by Rossi and Freeman (1985), that is, direct utility, conceptual use, and persuasive use (pp. 387–388). Since we had several audiences in mind at the beginning of our project – administrators, participating instructors, other instructors in Engineering (locally and at other universities), instructors in other disciplines – we provided excerpts from our final report for different groups with specific needs for information

(e.g., the Dean, the five-year planning committee, the Committee on Teaching and Learning, participating instructors, and other on-campus units with interests in the evaluation of technology in teaching and learning). We designed general dissemination strategies (e.g., presentations, posters, Web pages, other publications) to share our results as broadly as possible (Ives, Gandell & Mydlarski, 2004). We realized that educational-development colleagues might also benefit from the practice-based framework that evolved out of our experience; to this end, we contribute this detailed articulation of the evaluation process and the accompanying heuristic.

DOES THIS PROCESS WORK?

Patton (2000) suggested that successful evaluations are useful, practical, ethical, and accurate. Our experience suggests that our heuristic may be fruitful in supporting long-term pedagogical improvement. What evidence do we have that the evaluation team provided data that are being used by academic administrators for planning, by individual instructors for teaching improvement, and by academic developers in the form of needs assessment for future academic development activities, especially given the complexity of such a multi-faceted project (Accuracy Standard 12 – Metaevaluation [Joint Committee, 1994])?

In Engineering, the interpretive analysis is ongoing, serving individual participant instructors and the Faculty in general as formative assessment of teaching and learning in Engineering. The project provided a comprehensive analysis of specific instructional uses of technology in Engineering pedagogy, examining a range of technologies (e.g., tablet computers, PDAs, WebCT) and serving as a baseline for future development. Since our evaluation, several of the participants have made changes to their courses based on the results of their particular studies and are evaluating the impacts over time. Faculty administrators have instituted Faculty-wide technology initiatives designed to enhance student learning, including introducing a student laptop program to support student learning outcomes and promoting discipline-specific WebCT training. Individual participants in the evaluation project occupy leadership positions in the Faculty and are well positioned to influence future developments by sharing their experiences (Boud, 1999; McKinney, 2007).

From our perspective, academic developers are still working with instructors and Faculty administrators as they make decisions about technology implementation and integration activities. We and our successors have continued the conversations about learning and teaching with Engineering colleagues through a renewed commitment to the faculty development initiative that inspired our project (McAlpine & Saroyan, 2004). Although these are long-term initiatives, they build on the results we documented in our reports. The challenge of continuous improvement in student learning outcomes remains, but enhanced capacity for undertaking and evaluating innovative practices in the Faculty is established (Fullan, 2006). As Rossi and Freeman (1985) pointed out, evaluations designed to inform decision making may also have indirect or delayed effects.

The more general contribution of the project – the systematic evaluation heuristic (our third goal) – offers a framework and guidelines for future evaluation projects in the Faculty and beyond. It combines our experience and practice with educational inquiry guidelines in a way that highlights the factors of most value to those without formal training in educational evaluation or social-science research methods. Our evaluation tools, including the heuristic, are available in electronic form on an accessible website for Engineering professors to continue to use. We have worked with academic colleagues in Engineering who wished to reuse these tools and have helped adapt them for use in other Faculties as well, so we know they are helpful. For example, the university's Faculty of Continuing Education used the survey of instructor concerns about technology to gather data to help plan an e-learning initiative. As well, the university's teaching technology services group has adapted several of the course-level evaluation instruments for use with instructors in various disciplines who are testing such new technologies as classroom recording systems, personal response systems, and podcasting. Some of our Engineering colleagues are involved in these efforts, offering leadership and new expertise to the rest of the university community.

Although the results of these individual teaching and learning enhancement initiatives are not yet available, the systematic evaluation is contributing to the ability of our colleagues across the university to both assess the impact of their work and share the results of their practice, thereby advancing the scholarship of the teaching and learning community (McKinney, 2007; Weston & McAlpine, 2001).

CONCLUSION

In this article, we have emphasized three potential contributions of systematic, collaborative, and discipline-based evaluation. The process provides

- (a) a framework for tracking the impact of specific interventions in teaching in a formative assessment approach;
- (b) opportunities to initiate and continue conversations about teaching and learning within the disciplinary context; and
- (c) a focus on evidence-based decision making about teaching priorities within a specific academic unit and beyond.

Note that we are not trying to give the impression that change is straightforward or totally rational and that these are commandments to be followed. Our systematic evaluation initiatives were the product of a collaborative inquiry (Bray, Lee, Smith, & Yorks, 2000; Propriety Standard 1 – Service orientation [Joint Committee, 1994]) process with our disciplinary partners. The process was characterized by rigour – in the design, in the conduct of the inquiry using social science techniques, in the collection of data, and in the

integrative analysis. Our approach was discipline based but not discipline specific (McAlpine & Saroyan, 2004), and in providing an evaluation heuristic that may be adapted by Faculties and departments at our university and beyond, we have addressed a critical gap in the literature of the evaluation of teaching and learning.

In recent years, researchers and educational developers have noted a scarcity of comprehensive, programmatic, scholarly, and systematic assessments of innovations in teaching and learning at the university level (e.g., Ives, 2002; Sheard & Markham, 2005; Wankat et al., 2002). To address this, they and others have proposed a number of contextually grounded participative evaluation strategies that are similar in principle to what we do. For example, the following are recommended: multidisciplinary collaboration (Wankat et al., 2002), practitioner-centred research (Webber, Bourner, & O'Hara, 2003), scholarship of teaching approaches (Ives, McWhaw, & De Simone, 2005; McKinney, 2007; Wankat et al., 2002), action research (Dobson, McCracken, & Hunter, 2001; Rowley et al, 2004), action science/action inquiry (Argyris, Putnam, & Smith, 1985; Argyris & Schön, 1974), and design-based research (Design-Based Research Collective, 2003; Wang & Hannafin, 2005).

Our approach, which uses elements of formative, decision-oriented, responsive, and empowerment models of educational evaluation (Stufflebeam et al., 2000, pp. 26–30), shares these assumptions. This detailed description of our process offers insight and practical advice for those attempting systematic, discipline-based educational evaluation studies. Furthermore, the heuristic makes explicit underlying assumptions and asks specific questions not described in methodology texts or research reports of evaluation studies.⁵ It offers a scaffold for structuring collaborative evaluation projects, which may assist academics and educational developers with the process and help them ensure a scholarly (valid and reliable) approach. It explicitly describes the thinking and questions around which conversation develops among academic developers and academics as they collaboratively design evaluation studies to assess the impact of interventions in teaching and learning approaches. In particular, we focused on the distinctive activities that are involved when working with disciplines that do not use human subjects, including gaining ethical approval. In fact, developing the heuristic has made us aware that our notion of our roles as educational developers has expanded. We served at various times throughout the evaluation process as methodological experts and trainers, negotiators, facilitators of change, consultants, critics, and judges (Patton, 2000). As a result, we now realize that we are engaged in supporting not just those who wish to better understand or improve teaching and learning but also those who want to better understand and use social-science inquiry methods in the evaluation of learning and teaching. The heuristic provides a framework to do this – to engage in conversations about tracking impact, about interpreting data, about using evidence to support decisions on teaching and learning priorities. This scholarly approach will, we hope, resonate with our academic colleagues in a variety of disciplines. ♣

REFERENCES

- Advisory Committee for Online Learning. (2001). *The e-learning e-evolution in colleges and universities: A pan-Canadian challenge*. Ottawa: Council of Ministers of Education, Canada, and Industry Canada.
- Argyris, C., Putnam, R., & Smith, M. C. (1985). *Action science: Concepts, methods, and skills for research and intervention*. San Francisco: Jossey-Bass.
- Argyris, C., & Schön, D. (1974). *Theory in practice: Increasing professional effectiveness*. San Francisco: Jossey-Bass.
- Barnett, R. (2000). University knowledge in an age of supercomplexity. *Higher Education*, 40, 409–422.
- Becher, T., & Trowler, P. (2001). *Academic tribes and territories* (2nd ed.). Buckingham, UK: Open University Press.
- Biggs, J. (2001). The reflective institution: Assuring and enhancing the quality of teaching and learning. *Higher Education*, 42, 221–238.
- Boud, D. (1999). Situating academic development in professional work: Using peer learning. *International Journal for Academic Development*, 4(1), 3–10.
- Boyer, E. L. (1990). *Scholarship reconsidered: Priorities of the professoriate*. Princeton, NJ: Carnegie Foundation for the Advancement of Teaching.
- Bray, J. N., Lee, J., Smith, L. L., & Yorks, L. (2000). *Collaborative inquiry in practice: Action, reflection and making meaning*. Thousand Oaks, CA: Sage.
- Brew, A. & Boud, D. (1996). Preparing for new academic roles: A holistic approach to development. *International Journal for Academic Development*, 1(2), 17–25.
- Calder, J. (1994). *Programme evaluation and quality: A comprehensive guide to setting up an evaluation system*. London: Kogan Page.
- Canadian Council on Learning. (2006). *Report on Learning in Canada 2006, Canadian postsecondary education: A positive record – An uncertain future*. Ottawa: Author. Retrieved April 13, 2008, from <http://www.cclca.ca/CCL/Reports/PostSecondaryEducation/Archives2006/index.htm>
- Candy, P. (1996). Promoting lifelong learning: Academic developers and the university as a learning organization. *International Journal for Academic Development*, 1(1), 7–18.
- Chatterji, M. (2004). Evidence on “what works”: An argument for extended-term mixed-method (ETMM) evaluation designs. *Educational Researcher*, 33(9), 3–13.
- Cresswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed approaches*. Thousand Oaks, CA: Sage.

Cronbach, L. J. (2000). Course improvement through evaluation. In D. L. Stufflebeam,

G. F. Madaus, & T. Kellaghan (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation* (2nd ed., pp. 235–247). Boston: Kluwer Academic Publishers.

Denzin, N. K., & Lincoln, Y. S. (Eds.). (1998). *Strategies of qualitative inquiry*. Thousand Oaks, CA: Sage.

Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, 32(1), 5–8.

Dobson, M., McCracken, J. & Hunter, W. (2001). Evaluating Technology-Supported Teaching & Learning: A Catalyst to Organizational Change. *Interactive Learning Environments*, 9(2), 143–170.

Evaluation Center of Western Michigan University. Evaluation Checklists website. Retrieved August 26, 2006, from <http://www.wmich.edu/evalctr/checklists/>

Fullan, M. G. (2006). The future of educational change: System thinkers in action. *Journal for Educational Change*, 7, 113–122.

Fullan, M. G., & Stiegelbauer, S. (1991). *The new meaning of educational change* (2nd ed.). New York: Teachers College Press.

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational research: An introduction* (6th ed.). White Plains, NY: Longman.

Gandell, T., & Steinert, Y. (1999). *Faculty development in information technology for the basis of medicine: First year report*. Montreal, QC: McGill University, Faculty of Medicine.

Gross Davis, B. (1994). Demystifying assessment: Learning from the field of evaluation. In J. S. Stark & A. Thomas (Eds.), *Assessment and program evaluation* (pp. 45–57). Needham Heights, MA: Simon & Schuster Custom Publishing.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.

Haertel, G. D., & Means, B. (Eds.). (2003). *Evaluating educational technology: Effective research designs for improving learning*. New York: Teachers College Press.

Ives, C. (2002). *Designing and developing an educational systems design model for technology integration in universities*. Unpublished doctoral dissertation. Concordia University, Montreal, QC.

Ives, C., Gandell, T., & Mydlarski, L. (2004). *Systematic evaluation of the use of technologies in enhancing teaching and learning in engineering*. Unpublished report, McGill University, Centre for University Teaching & Learning, and Faculty of Engineering, Montreal, QC.

Ives, C., McWhaw, K., & De Simone, C. (2005). Reflections of researchers involved in the evaluation of pedagogical technological innovations in a university setting. *Canadian Journal of Higher Education*, 35(1), 61–84.

Ives, C., Mydlarski, L., Gandell, T., Gruzleski, J., Frost, D., Kirk, A., et al. (2004, October). *Systematic evaluation of the use of technologies in enhancing teaching and learning in engineering*. Poster session presented at the EDUCAUSE conference, Denver, CO.

Jenkins, A. (1996). Discipline-based educational development. *International Journal for Academic Development*, 1(1), 50–62.

Johnson, R., & Onwuegbuzie, K. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14–26.

Johnson, S. & Broda, J. (1996). Supporting educational researchers of the future. *Educational Review*, 48(3), 269–281.

Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards: How to assess evaluations of educational programs* (2nd ed.). Thousand Oaks, CA: Sage.

Kember, D. (1997). A reconceptualisation of the research into university academics' conceptions of teaching. *Learning and Instruction*, 7(3), 255–275.

Kirkpatrick, D. (1998). *Evaluating training programs: The four levels* (2nd ed.). San Francisco: Berrett-Koehler.

Land, R. (2001). Agency, context and change in academic development. *International Journal for Academic Development*, 6(1), 4–20.

Leedy, P. D., & Ormrod, J. E. (2001). *Practical research: Planning and design* (7th ed.). Upper Saddle River, NJ: Prentice-Hall.

Levin-Rosalis, M. (2003). Evaluation and research: Differences and similarities. *Canadian Journal of Program Evaluation*, 18(2), 1–31.

Lincoln, Y. S., & Guba, E. G. (1986). But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. In D. D. Williams (Ed.), *Naturalistic evaluation* (pp. 73–84). New Directions for Program Evaluation, no. 30. San Francisco: Jossey-Bass.

Madaus, G. F., & Kellaghan, T. (2000). Models, metaphors and definitions in evaluation. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation* (2nd ed., pp. 19–31). Boston: Kluwer Academic Publishers.

McAlpine, L. (2005). The impact of academic development: Questioning my evaluation practices. *Educational Developments*, 6(1), 5–8.

McAlpine, L., & Cowan, S. (Eds.). (2000). *Reflections on teaching and learning: 30 years at McGill*. Montreal, QC: McGill University, Centre for University Teaching and Learning.

McAlpine, L., Gandell, T., Winer, L., Gruzelski, J., Mydlarski, L., Nicell, J., et al. (2005). A collective approach towards enhancing undergraduate engineering education. *European Journal of Engineering Education*, 30(3), 377–384.

McAlpine, L., & Saroyan, A. (2004). Toward a comprehensive framework of faculty development. In A. Saroyan & C. Amundsen (Eds.), *Rethinking teaching in higher education: From a course design workshop to a faculty development framework* (pp. 207–232). Sterling, VA: Stylus.

McAlpine, L. & Weston, C. (2000). Reflection: Issues related to improving professors' teaching and students' learning. *Instructional Science* 28, 363–385,

McKinney, K. (2007). *Enhancing learning through the scholarship of teaching and learning*. Bolton, MA: Anker.

Mentkowski, M. (1994). Creating a context where institutional assessment yields educational improvement. In J. S. Stark & A. Thomas (Eds.), *Assessment and program evaluation* (pp. 251–268). Needham Heights, MA: Simon & Schuster Custom Publishing.

Neumann, R. (2001). Disciplinary differences and university teaching. *Studies in Higher Education*, 26(2), 135–146.

Pace, D., & Middendorf, J. (Eds.). (2004). *Decoding the disciplines: Helping students learn disciplinary ways of thinking*. New Directions for Teaching and Learning, no. 98. San Francisco: Jossey-Bass.

Patton, M. Q. (2000). Utilization-focused evaluation. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation* (2nd ed., pp. 425–438). Boston: Kluwer Academic Publishers.

Professional and Organizational Development Network in Higher Education. (2007). What is faculty development? Retrieved May 20, 2008, from http://www.podnetwork.org/faculty_development/definitions.htm

Public Works and Government Services Canada. (2003). Tri-Council policy statement: Ethical conduct for research involving humans. Retrieved May 20, 2008, from http://pre.ethics.gc.ca/english/pdf/TCPS%20June2003_E.pdf

Rae, B. (2005). Ontario: A leader in learning. Report and recommendations. Toronto: Government of Ontario. Retrieved April 13, 2008, from <http://www.edu.gov.on.ca/eng/document/reports/postsec.pdf>

Ramsden, P. (1992). *Learning to teach in higher education*. London: Routledge.

Randall, J. (2001). Academic review in the United Kingdom. In D. Dunkerley & W. Wong (Eds.), *Global perspectives on quality in higher education* (pp. 57–69). Aldershot, UK: Ashgate.

Robertson, M. (1998). Benchmarking teaching performance in universities: Issues of control, policy, theory and best practice. In J. Forest (Ed.), *University teaching: International perspectives* (pp. 275–303). New York: Garland.

Rossi, P. H., & Freeman, H. E. (1985). *Evaluation: A systematic approach* (3rd ed.). Beverly Hills, CA: Sage.

Rowley, J., Ray, K., Proud, D., Banwell, L., Spink, S., Thomas, R., et al. (2004). Using action research to investigate the use of digital information resources in further education. *Journal of Further and Higher Education*, 29(3), 235–246.

Samuelowicz, K & Bain, J.D., (2001). Revisiting academics' beliefs about teaching and learning. *Higher Education*, 41, 299–325.

Saroyan, A., & Amundsen, C. (2004). *Rethinking teaching in higher education: From a course design workshop to a faculty development framework*. Sterling, VA: Stylus.

Sheard, J., & Markham, S. (2005). Web-based learning environments: Developing a framework for evaluation. *Assessment & Evaluation in Higher Education*, 30(4), 353–368.

Shulman, L. S. (2000). From Minsk to Pinsk: Why a scholarship of teaching and learning? *Journal of the Scholarship of Teaching and Learning*, 1(1), 48–52.

Smith, S. (1991). *Report of the Commission of Inquiry on Canadian University Education*. Ottawa, ON: Association of Universities and Colleges of Canada.

Stake, R. E. (2000). Program evaluation, particularly responsive evaluation. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation* (2nd ed., pp. 343–362). Boston: Kluwer Academic Publishers.

Stake, R. E. (2004). *Standards-based & responsive evaluation*. Thousand Oaks, CA: Sage.

Stufflebeam, D. L. (2000a). Foundational models for 21st century program evaluation. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation* (2nd ed., pp. 33–83). Boston: Kluwer Academic Publishers.

Stufflebeam, D. L. (2000b). The CIPP model for evaluation. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation* (2nd ed., pp. 279–317). Boston: Kluwer Academic Publishers.

Stufflebeam, D. L., Madaus, G. F., & Kellaghan, T. (Eds.). (2000). *Evaluation models: Viewpoints on educational and human services evaluation* (2nd ed.). Boston: Kluwer Academic Publishers.

Stufflebeam, D. L., & Webster, W. J. (1994). An analysis of alternative approaches to evaluation. In J. S. Stark & A. Thomas, *Assessment and program evaluation* (pp. 331–347). Needham Heights, MA: Simon & Schuster Custom Publishing.

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.

Teichler, U. (2003). Changing concepts of excellence in Europe in the wake of globalization. In E. De Corte (Ed.), *Excellence in higher education* (pp. 33–51). London: Portland Press.

Tessmer, M. (1998). *Planning and conducting formative evaluations: Improving the quality of education and training*. London: Kogan Page.

Trigwell, K., Prosser, M., & Waterhouse, F. (1999). Relations between teachers' approaches to teaching and students' approaches to learning. *Higher Education*, 37, 57–70.

Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development*, 53(4), 5–23.

Wankat, P. C., Felder, R. M., Smith, K. A., & Oreovica, F. S. (2002). The scholarship of teaching and learning in engineering. In M. T. Huber & S. Morreale (Eds.), *Disciplinary styles in the scholarship of teaching and learning: Exploring common ground*. Washington, DC: AAHE/Carnegie Foundation for the Advancement of Teaching.

Webber, T., Bourner, T., & O'Hara, S. (2003). Practitioner-centred research in academic development in higher education. In H. Eggins & R. Macdonald (Eds.), *The scholarship of academic development* (pp. 117–128) Buckingham, UK: SRHE and Open University Press.

Weston, C. (1986). Formative evaluation of instructional materials: An overview of approaches. *Canadian Journal of Educational Communications*, 15(1), 5–17.

Weston, C., & McAlpine, L. (2001). Integrating the scholarship of teaching into disciplines. In C. Kreber (Ed.), *Scholarship revisited: Perspectives on the scholarship of teaching* (pp. 89–98). New Directions in Teaching and Learning, no. 86. San Francisco: Jossey-Bass.

Wiggins, G. (1998). *Educative assessment. Designing assessment to inform and to improve performance*. San Francisco: Jossey-Bass.

Williams, D. D. (1986). When is naturalistic evaluation appropriate? In D. D. Williams (Ed.), *Naturalistic evaluation* (pp. 85–92). New Directions for Program Evaluation, no. 30. San Francisco: Jossey-Bass.

Yin, R. K. (1984). *Case study research: Design and methods*. Thousand Oaks, CA: Sage.

APPENDIX 1

Comparison of the Steps in Our Heuristic with the Stages of Evaluation Described by Other Models

Steps in Heuristic	Joint Committee (1994) Key Tasks	Calder (1994)	Gall, Borg, and Gall (1996)	Responsive evaluation (Stake 2000)
Building the team	Decide whether to evaluate		Identify stakeholders	Identify users of results
Clarifying the need	Define evaluation problem	Identify an area of concern	Clarify reasons	Determine focus
Setting evaluation goals			Decide	
Designing the studies*	Design	Decide whether to proceed	Identify questions, procedures Evaluation design / timeline	Methods, measurement, and design decisions
Gaining ethical approval				
Developing instruments				
Collecting data	Collect information	Investigate identified issues	Collecting and analyzing	
Analyzing data	Analyze information	Analyze findings		Analysis and interpretation
Interpreting and reporting results	Report evaluation	Interpret findings	Report	
Disseminating / using results		Disseminate findings and recommendations Review responses Implement actions		Dissemination
*includes resources	Budget		Resources	
	Contract			
*includes management	Management, staff		Management	
	Develop evaluation policies			

APPENDIX 2

Systematic Evaluation Heuristic for Discipline-based Collaborative Examination of Teaching and Learning

Sample questions for evaluators at each step	Criteria to attend to before moving on to the next step
1. Building the team	
Goals: build representative team; develop relationships; explore disciplinary culture	
Who will be involved? What roles do they play (e.g., advisers, information providers, disrupters)?	Build on what has already been learned Clarify, articulate roles of key participants
What is the history of teaching and learning initiatives? Who has been involved?	Identify formal and informal channels of communication
What are the preferred styles of communication and collaboration? What has been successful so far? What has been tried but has not worked?	Identify general resources available to the process
How do decisions get made? Who participates? Who does not participate?	
2. Clarifying the need	
Goals: build consensus; identify current practices; determine gaps	
What are the general and specific concerns of instructors?	Involve a broad range of stakeholders
What evidence exists of student concerns?	Specifically articulate the problems and issues to be addressed and why
What aspects of the course or program cause students most difficulty? What might help?	Share findings with stakeholders for buy-in and for consensus on relevance and importance of the problems
What aspects of student learning or performance would we like to improve or change?	
What other information would we like to have about teaching and learning in our context?	

Sample questions for evaluators at each step	Criteria to attend to before moving on to the next step
<p>3. Setting evaluation goals</p> <p>Goals: focus on learning; determine level of evaluation; develop shared values</p> <p>Once we agree on the concerns of the unit as a whole, how do we articulate and validate our decisions in terms of goals?</p> <p>What specific questions would participating instructors like to see answered by the evaluation? (For example, what instructional strategies have the most impact on student learning?)</p> <p>What kinds of evidence would be useful to us or to our colleagues? How do we anticipate using the results? (For example, what recommended practices do we hope will emerge?)</p>	<p>State accepted goals explicitly</p> <p>Establish consensus about the proposed level of evaluation</p> <p>Ensure conceptions of the educational research process are realistic, so individuals do not have expectations that cannot be met</p>
<p>4. Designing the studies</p> <p>Goals: determine appropriate methods; describe methodology; identify specific resources</p> <p>What are the characteristics of the participating learners (so that design decisions are appropriate)?</p> <p>What is the instructional context of the evaluation study (e.g., classroom, laboratory, or online; single learning activity; or spread throughout the semester)?</p> <p>What data are necessary to make decisions about future activities? What data will help to capture student learning and impact (e.g., student grades, satisfaction surveys)?</p> <p>Do the measures selected provide the necessary data to sufficiently answer the questions? (For example, is a questionnaire adequate, or would interviews or focus groups help collect better information?)</p> <p>What data will be collected, by whom, and for what purpose (e.g., quantitative, qualitative, or a combination)?</p> <p>What resources are necessary and where can they be found?</p>	<p>Ensure goals are achievable, given resources available (if not, redefine them)</p> <p>Confirm data to be collected will provide sufficient information to answer the questions</p> <p>Prepare back-up plans to cope with unforeseen difficulties in data collection and analysis</p>

Sample questions for evaluators at each step	Criteria to attend to before moving on to the next step
<p>5. Gaining ethical approval</p> <p>Goals: describe study for others (transparency); protect learners</p> <p>What are possible consequences to learners of the intervention not working out as intended?</p> <p>What questions get at the information we are looking for (to help design draft instruments)?</p> <p>What are we going to do with the findings? How are we going to report back to the stakeholders (to help design dissemination plans)?</p>	<p>Follow proposal submission procedures as specified by the relevant ethical review board</p> <p>Allocate sufficient time for the approval process before the study begins (this can sometimes take months)</p> <p>Work out study design in detail</p> <p>Clearly describe methodology in the application</p>
<p>6. Developing the evaluation instruments</p> <p>Goals: continue teaching and learning conversations; anticipate and prepare for problems</p> <p>Who is the best source of relevant information (e.g., students, instructors)?</p> <p>What is the best way to capture evidence of the desired results (e.g., survey of attitudes, test of learning, interviews, log files)?</p> <p>What is an appropriate way to motivate participants to take part in the data collection?</p> <p>How can the desired information be collected without wasting valuable learning or teaching time but still respecting the integrity of the participants?</p>	<p>anticipate and prepare for problems</p> <p>Focus questions so the answers will be interpretable</p> <p>Pilot-test instruments with members of the target group(s)</p> <p>Analyze results from pilot tests</p> <p>Revise questions based on these results</p>

<p>Sample questions for evaluators at each step</p>	<p>Criteria to attend to before moving on to the next step</p>
<p>7. Collecting the data</p>	<p>Goals: increase awareness of learner perspectives; ensure fairness and ethical conduct</p> <p>Do all participants get the same instructions? Ensure participating students feel comfortable</p> <p>Are the instruments readily available and readable? Are all the instruments administered according to the specified timeline? Collect and store data to maintain confidentiality Allow flexibility in case of unintended consequences on students</p> <p>Are all other accepted procedures followed exactly? Are participants properly informed and do they give consent freely and without coercion? Focus attention on each research method equitably Follow specified procedures</p> <p>Do all participants have the same opportunities to learn or to benefit from the instructional activities being evaluated? Do particular contexts require a change in the actual delivery of the planned innovation? What impact might this have on the study and/or on the students' learning?</p>
<p>8. Analyzing the data</p>	<p>Goals: demonstrate value of feedback; represent findings so they are meaningful</p> <p>Are the data complete and accurate? Assess validity and reliability of data Illustrate findings with graphic representations of data</p> <p>Is there evidence of problems with the collection process or the instruments? Complete and report initial data analysis to collaborators in a timely manner</p> <p>How will the data be represented (charts, frequencies) to make them meaningful to the stakeholders? Represent qualitative data in multiple ways</p>

<p>Sample questions for evaluators at each step</p>	<p>Criteria to attend to before moving on to the next step</p>
<p>9. Interpreting and reporting results</p>	<p>Goals: compare various interpretations; remember the concept of formative assessment; identify audiences for various findings</p> <p>What do the results mean? Can they be interpreted in more than one way? Interpret results collaboratively Address expectations of causal relationships</p> <p>Do qualitative and quantitative results support (or contradict) each other? (For example, are there correlations among results?) Explicitly articulate limitations of study/studies Relate conclusions and findings specifically to goals</p> <p>Is there sufficient evidence to draw conclusions regarding impact or change? Can multiple measures be used to more strongly support evidence of change or impact? Report findings that address the needs of a range of audiences</p> <p>Do the data collected answer the research questions? What variables can be identified that may have affected the results? What limitations were there and what impact did they have on the interpretation of data?</p> <p>What conclusions and recommendations can be drawn?</p>
<p>10. Disseminating / using results</p>	<p>Goals: explore implications broadly</p> <p>Did this evaluation answer our questions? How can we use this information? Consider and address all potential audiences Provide answers to research questions in reports and presentations</p> <p>Who else might benefit from the information gathered or from the general conclusions drawn? Articulate new questions as goals for future research</p> <p>Who are the possible audiences for the results? How do we construct reports for different potential audiences? How might other audiences be reached? What new questions arise from this report?</p>

CONTACT INFORMATION

Cindy Ives
 Centre for Learning Design and Development
 Athabasca University
 1 University Drive
 Athabasca AB
 T9S 3A3
 cindy@athabascau.ca

Cindy Ives has worked in higher education for many years in a variety of contexts and locations. Instructor, faculty developer, researcher, distance educator, academic administrator, and evaluation consultant, she is currently director of the Centre for Learning Design and Development at Athabasca University. There, she leads the teams responsible for the design, development, production, and evaluation of distance education courses and programs at Canada's Open University.

Lynn McAlpine is a professor of Higher Education Development at the University of Oxford. She was formerly at McGill University in Canada. Her current research is directed at understanding the experiences of doctoral students, post-doctoral fellows, and pre-tenure academics as they construct their academic identities.

Terry Gandell is a pedagogical consultant in private practice. She was formerly an assistant professor at McGill University and at Bishop University and a special education teacher for the English Montreal School Board, in Quebec, Canada. Terry works with individuals and organizations on strategic planning, curriculum and staff development, and program evaluation to help enhance teaching and learning in a variety of contexts.

ACKNOWLEDGEMENT

The authors acknowledge the collaboration of their colleagues during the evaluation studies that inspired this article, and they thank the reviewers for their suggestions for improvement.

NOTES

1. The term "evaluation" has several connotations. In this article, we focus on evaluation research about the effectiveness of organized teaching and learning supports for student learning (Calder, 1994; Stufflebeam et al., 2000), rather than on evaluation as an assessment of student learning or on student evaluations of teaching. Although some scholars distinguish research from evaluation (Levin-Rosalis, 2003), we view evaluation as a form of social science research (Chatterji, 2004; Rossi & Freeman, 1985). What

makes evaluation distinctive is its origin: problem-oriented, driven more by the needs emerging within the context than by questions or gaps in the discipline (Teichler, 2003). Nevertheless, at its best, it is scholarly; it uses a range of data collection, display, and analysis strategies; and it is rigorous and open to critique. Context is a critical factor in evaluation (Chatterji, 2004), as it is in all social science research. Furthermore, in evaluation (as in some curiosity-driven research), collecting and analyzing data over time can be significant, supporting its interpretation in formative as well as summative contexts.

2. We use the terms “instructor,” “faculty member,” “academic,” and “professor” interchangeably in this article to refer to those staff assigned responsibility for teaching and learning activities organized as courses in our university.
3. Although the checklists provided on the website of The Evaluation Center of Western Michigan University (2006) are useful background resources for our academic development work, they are very detailed, generic, and not discipline based. These features make them cumbersome for working directly with academic colleagues unfamiliar with educational evaluation.
4. Also reported in Wankat et al. (2002).
5. The only other example we were able to find of a generic evaluation framework was produced at the University of Calgary in the late 1990s. In an effort to document and evaluate technology implementation efforts by individual instructors, academic developers produced the *Formative Evaluation Planning Guide* (Dobson, McCracken, & Hunter, 2001) to help them assess their technological innovations. The guide explicitly describes the roles of participants in the process, the possible types of studies, and the data collection and analysis tools available for faculty conducting evaluations. However, its focus is individual, rather than programmatic, and it is described as a tool for evaluating technology specifically, rather than pedagogy more generally. The program evaluation standards and guidelines produced by the Joint Committee on Standards for Educational Evaluation (Joint Committee, 1994) are useful for informing the design and assessment of evaluation projects, but they do not explicitly address the questions we wanted to ask our disciplinary colleagues during the process.