THICKENING THE PLOT: COMBINING OBJECTIVES- AND METHODS-ORIENTED APPROACHES IN THE EVALUATION OF A PROVINCIAL SUPERINTENDENTS' QUALIFICATION PROGRAM

Tim Gawley
Wilfrid Laurier University, Brantford Campus
Brantford, Ontario

Abstract: This article discusses the evaluation framework for an ongoing provincial superintendents' qualification program. Combining the standards-based Discrepancy Evaluation Model (DEM) with a methodologically intensive case study method, this evaluation assessed the program across three generally defined focus areas: program alignment, skills commensurability, and comparative training. These areas were grounded in formal legislative and regulatory standards commensurate with the procedures of the DEM, but evaluation findings for each of these areas were strengthened with the addition of the case study method. Theoretical, methodological, and utilization merits of this framework are discussed. Lessons for the evaluation of similar programs are also considered.

Résumé : Cet article traite du cadre d'évaluation relatif à un programme provincial continu d'homologation des chefs de service. Au moyen d'un modèle d'évaluation des divergences, le Discrepancy Evaluation Model (DEM), fondé sur les normes et combiné à une méthode d'étude de cas approfondie, cette évaluation fait l'examen du programme pour trois secteurs d'intervention privilégiés : harmonisation du programme, mesure des compétences, et formation comparative. Ces secteurs se fondent sur des normes officielles de nature législative et réglementaire correspondant aux procédures du DEM, mais les conclusions de l'évaluation pour chacun des secteurs ont été renforcées par l'ajout d'une méthode d'étude de cas. Le bien-fondé théorique, méthodologique, et relatif à l'utilisation de ce cadre est abordé. Les leçons tirées lors de l'évaluation de programmes similaires sont également prises en compte.

Corresponding author: Tim Gawley, Laurier Brantford, Wilfrid Laurier University, 73 George Street, Brantford, ON, N3T 2Y3; <tgawley@wlu.ca>

Across Canada, standards of practice governing the teaching profession ensure that educators are rigorously prepared for the duties and responsibilities of the profession at all levels. To this extent, the Ministry of Education in one Canadian province, along with the provincial governing body that certifies and regulates the teaching profession, makes program reviews and evaluations a priority.[1] When the design and delivery of professional qualifications programs are governed in this manner, traditional standards-based or objectives-based evaluation designs seem like the most reasonable alternatives, because the conceptualization and operationalization of standards are sensibly predetermined by the wording of these acts and regulations. However, the application of qualitatively oriented perspectives in program evaluation helps to identify the undeniable uniqueness of program issues and needs that also aid in the clarification and enhancement of existing program standards. The purpose of this article is to present an objectives-based, multi-method, evaluation framework of a provincial superintendents' qualification program (SQP)[2] that combined components of the standards-based Discrepancy Evaluation Model (DEM) with elements of the more data-intensive case study strategy. Using these options, the evaluation framework assessed the design, delivery, and outcomes of a provincial superintendents' qualification program within the scope of three collaboratively defined areas of focus: program alignment, commensurability, and comparative training. Implications with respect to the evaluation of similar educational administration preparation programs are considered, including the theoretical and methodological merits of this framework, its involvement of stakeholders for enhancing evaluation use, and the value of data triangulation in generating evaluation findings.

THE SUPERINTENDENTS' QUALIFICATION PROGRAM

In 1992 the Ministry of Education changed its method of certifying superintendents by implementing a five-module program to replace the written and oral examinations.[3] Through various delivery agencies, the Ministry oversaw the certification program until the province's governing body for teachers took over the function in 1997. The Superintendents' Qualification Program underwent a program evaluation in 2002 that resulted in some formative revisions. In 2004 the governing body sought external parties to conduct an evaluation of the program's design and delivery during the previous three years of its operation. The SQP is a rigorous program designed to prepare members of the provincial education system for superintendent posts. A

primary goal of the program is to enable superintendents to deal with the contemporary political, economic, and culturally diverse realities of the province's schools. Educational administrators are expected to ensure that school boards are properly executing the policies of the Ministry of Education and that boards are held accountable. Equally important is the superintendent's responsibility for sustaining commitments to administrative or organizational accountability, strong professional development at all levels of the education system, and student achievement.

The SQP is unique among preparation programs for educational administrators. It is the only provincial qualification program in Canada that is required by act of legislation for entry into a superintendent position. In other provinces, educational administrators require a teaching license, extensive periods of experience in the profession as teachers and/or administrators, and, in some cases, a graduate degree. Each of these is a prerequisite for entry into the SQP. This parallels similar programs in the United States where candidates typically require a master's degree, administrative experience, the completion of educational administration courses, licenses, and practicum exercises.

The SQP prepares candidates using a five-module system. The two initial modules of the program range in content from the development of self-awareness (e.g., identifying a candidate's management style or personality type, identifying and developing leadership styles) to the teaching of position-specific skills such as problem-solving, curriculum development, technology use, and leadership skills. The third and fourth modules focus on the comprehension and application of government legislation and the development of visionary, change management, and relationship-building skills. The final module is a leadership practicum in which a candidate is mentored with an existing superintendent. The candidate and mentor then develop a special project through which the candidate can demonstrate a competent application of the knowledge, skills, and aptitudes learned during the previous four program modules.

At the time of the evaluation, the SQP was delivered by four training providers. A fifth provider was added just prior to the evaluation. Although a candidate could attend the SQP with any of these providers, two of the providers serviced English-language public school candidates, one served the Catholic school boards, while a fourth addressed the unique requirements of Francophone candidates. Instructors were hired by these providers to deliver the SQP curriculum and evaluate

candidate performance. Candidate performances were evaluated through oral, written, and performance-based exercises.

A questionnaire administered as part of the evaluation provides a profile of SQP candidates (N = 152). As of January 2004, information provided by the provincial governing body and SQP providers suggested that over 300 candidates completed at least some of the SQP modules since 2002. Among respondents to the questionnaire, the candidates were split along gender lines with 50.3% of the respondents being female. Almost two-thirds (63.8%) of the respondents were over the age of 40, and 84.8% of the respondents had at least 10 years of experience in the teaching profession. Finally, respondents came from a diversity of positions within the education system. The majority of respondents (53.4%) were principals, 26.7% were either vice-principals or teachers, and 19.9% were superintendents, directors, or in equivalent acting positions. The candidates also represented school boards of all sizes from across the province. No particular board size was overrepresented in the sample.

EVALUATION OF THE SQP: THE DISCREPANCY EVALUATION MODEL

The governing body for teachers in the province is given the authority and responsibility for accrediting additional qualification and pre-service programs. Parameters for the evaluation of these programs are set out in the acts and regulations of the Ministry of Education and the governing body. To evaluate the SQP, this body wanted to determine whether the design and delivery of the program was aligned with the requirements set out by these acts and regulations. Hence, an evaluation model was selected that could analyze the discrepancy between these acts and regulations and the design and delivery of the SQP.
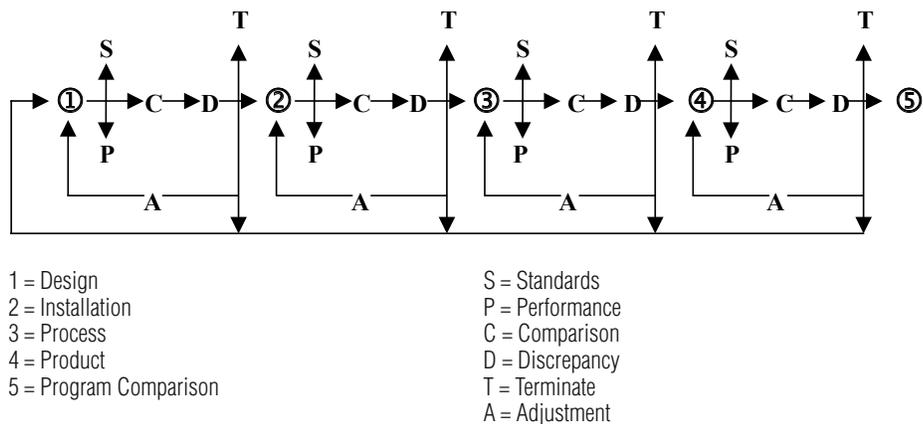
The Discrepancy Evaluation Model is an objectives-based evaluation theory in which a program is evaluated according to the extent to which it achieves its intended objectives (Provus, 1971). Here, the achievement of objectives is determined according to the definition of operationalized standards (Stufflebeam, 2001a, p. 17). Summarized in Figure 1, the model includes five stages of evaluation: design, installation, process, product, and the optional cost-benefit analysis stage (Alkin & Christie, 2004, p. 46; Provus, 1971, 184). The SQP evaluation did not include this optional fifth stage. Within the four stages, the evaluation model includes the definition of standards (S),

the assessment of performance (P), a comparison between standards and performance (C), and finally the determination of standard-performance discrepancies (D). Following the determination of discrepancies, decisions are made on whether to terminate (T) or adjust (A) a program (Provus, 1971, p. 184; Steinmetz, 1983).

The Advisory Committee

The DEM has not been known for its utilization potential until recently. Alkin and Christie (2004, p. 44) situate Provus's (1971) DEM on the "Use" branch of their evaluation theory tree. They see Provus's model as "decision-oriented" in that it is explicitly designed to help key stakeholders make decisions about given programs. In discrepancy evaluation, Provus emphasizes the importance of stakeholder involvement in the generation and support of evaluation standards when he states that there "must be maximum involvement of program staff in every step of the evaluation process. Further it follows that the evaluation staff establish a continual rapport with the program staff based on communication of affect as well as publicly acceptable verbalizations" (Provus, 1971, p. 196). Provus restrains the involvement of staff, however, when he continues by saying that evaluation "is the handmaiden of program development and quiet counselor to administrators, but it operates in accordance with its own rules and on the authority independent of the program unit" (Provus, 1971, p. 186).

Figure 1
The Discrepancy Evaluation Model

1 = Design   S = Standards
2 = Installation   P = Performance
3 = Process   C = Comparison
4 = Product   D = Discrepancy
5 = Program Comparison   T = Terminate
                              A = Adjustment

Adapted from Provus (1971, p. 184).

The provincial governing body, in carrying out this evaluation, was committed to a co-operative transparent endeavour, ensuring that each of the program providers was fully aware of the purpose and form of the evaluation. To that end, the governing body and the evaluators agreed to establish an advisory committee with representatives from all four existing providers and a new SQP provider that would meet with the evaluators, provide feedback, and offer comments and suggestions as the evaluation progressed. Table 1 outlines the main activities of the advisory committee. The evaluation design was not entirely determined by the evaluators, but rather was informed by provider representatives of the advisory committee. The advisory committee maintained its involvement throughout the process by informing which areas of the program would be evaluated and by consulting with the evaluators about methodological strategies. While provider representatives were able to add insight into the evaluation's design, it was the evaluators who would make the final decisions about the actual design and conduct of the evaluation. It was understood that the committee's role was indeed advisory and that the evaluation would be performed by the evaluators. The creation of a stakeholder advisory committee was invaluable in generating an evaluation that was ultimately ethical, feasible, accurate, and useful.

Table 1
Activities of the Advisory Committee

Representing the interests of providers in the evaluation process
Providing feedback in the initial stages of the evaluation
Acting as a resource during the evaluation for when questions or issues emerge
Sharing ideas between evaluators, providers, and the provincial governing body
Provider representatives acting as liaisons with their organizations
Providing a formal response (feedback) at the end of the evaluation

Applying the DEM in the Assessment of Three SQP Evaluation Focus Areas

In the SQP evaluation, design discrepancies were first analyzed by comparing established design standards and their indicators with the program's actual design. These areas of focus were clarified through meetings with the provincial governing body and later through meetings with an advisory committee. Called alignment, the first area of focus was defined as follows:

1. *Alignment*: the extent to which the SQP as designed is aligned with the acts, regulations, and guidelines that govern the program.

In this stage, it was determined if the program was defined according to existing standards, and if so, to what extent was the actual design of the program in keeping with these design standards (Provus, 1971, p. 199). To measure the SQP's achievements, the evaluation assessed whether these instructional expectations were designed, documented, and disseminated in accordance with five of the governing body's legislative and regulatory standards represented by the eight indicators summarized in Table 2.

Table 2
Three Evaluation Focus Areas, their Indicators (Derived from Ministry and Governing Body Standards) and DEM Stages

DEM	Focus Areas	Indicators
DESIGN	1. Alignment	Program content (presence of all modules and practicum)
		Admission requirements, registration, and record-keeping practices
		Location, schedule, hour requirements, and module structure
		Personnel (vitae submission requirements)
		Program structure (sequence of module delivery)
		Practicum (clarity of expectations, selection criteria for mentors, and use of mentors)
		Program evaluation (commitment to improvement and quality assurance)
		Candidate assessment methods
INSTALLATION	2. Commensurate expectations	General commensurate tasks (instruction in a variety of superintendent-related issues)
		Practicum assessment (availability of practicum information and finding mentors)
		In-course mentoring (presence and usefulness)
		Self-assessment and suitability requirements
		Development of the skill set
		Coverage of the legislative framework
PROCESS PRODUCT	3. Comparative training	Candidate reactions to the program (usefulness of role models and practicum opportunities)
		Behavioural change potential from the program
		• Skill development and utilization
		• Self-efficacy development
		Outcomes: Success in superintendent competitions

Design was not the only issue. When the standards for delivery were defined and assessed, the next step was to examine the actual delivery of the program in accordance with these standards (Provus, 1971, p. 185; Steinmetz, 1983, p. 85–86). The delivery aspect of the DEM was particularly popular in the SQP evaluation for its parallels

with curriculum auditing techniques. Curriculum auditing similarly involves the assessment of whether a curriculum is actually designed and delivered according to formal standards where design refers to "the act of creating a curriculum" while delivery refers to "the act of implementing, supervising, monitoring, or using feedback to improve the curriculum" (English, 2000, p. 3). The evaluators and members of the advisory committee were familiar with this technique, and so the DEM satisfied the preferences of committee members. The evaluation would focus on aspects of the SQP curriculum. However, a simple curriculum audit was not all that the provincial governing body wanted. The evaluators went beyond curriculum alignment to evaluate other potential delivery discrepancies.

The evaluators requested from the provincial governing body and its advisory committee any other concerns that individual providers felt required evaluation. In addition to alignment, the provincial governing body agreed to evaluate the commensurability of skill development in the program with the actual skills acquired by SQP candidates. Also, the governing body agreed that the evaluators should examine how the SQP compares with other high-quality training programs. Moreover, the evaluation of these two focus areas would not concentrate strictly on the legislative standards of the program. Rather, they would also be evaluated according to the comments and experiences of SQP candidates and instructors. These two areas of focus were labeled commensurate expectations and comparative training:

2. *Commensurate Expectations*: the extent to which the program as delivered reflects the expectations of the provincial governing body, SQP candidates, and its instructors.

3. *Comparative Training*: the extent to which the SQP is reflective of current research on high-quality training programs with respect to its ability to encourage successful training-transfer from in-program to the actual performance of superintendent duties and responsibilities.

Here, commensurate expectations reflected the DEM's installation stage whereby evaluators observed the extent to which the SQP was actually delivered in accordance with observed program delivery standards. The indicators for determining whether these delivery expectations were met are summarized in Table 2. Newly minted SQP graduates were invited to provide feedback regarding the program's coverage of issues normally confronting a superintendent such as special education needs, assessment practices, student management,

and staff supervision. The program's practicum, mentoring, and self-assessment components were also evaluated under this second focus area, examining the extent to which they were delivered to simulate activities normally handled by superintendents.

In addition, the evaluation assessed the extent to which the skills learned during the SQP were commensurate with the skills required for the performance of superintendent duties and responsibilities. Nine of the skills selected for analysis were identified in the program's guidelines, and were also identified using Canada's National Occupational Classification–NOC Code 0313: School Principals and Administrators of Elementary and Secondary Education (Human Resources and Skills Development Canada, 2007). The following skills were defined as most relevant in the performance of educational administration duties and responsibilities: interactive skills, problem-solving skills, decision-making skills, organizational skills, evaluation skills, negotiation skills, management skills, teamwork skills, and leadership skills. In addition to these, both candidates and instructors were also asked to assess the extent to which the SQP developed an understanding of the provincial legislative framework.

An issue with subjective-based skill development and utilization assessments concerns the validity of these responses. Can self-reports by SQP candidates be considered valid assessments of skill development in the program? To ensure the validity of trainee responses, industrial psychology studies (Fox & Yossi, 1988; Johns, 1994) promote the use of cross-verification between trainee self-reports and the reports of supervisors. Similarly, in the context of the DEM, the assessment of instructional processes also requires the observation of "student activity (SA) and teacher activity (TA) associated with each learning experience" (Provus, 1971, p. 204). Therefore, to improve the validity of SQP candidates' self-reports, their responses were similarly compared with SQP instructor responses.

It was determined that, while the SQP must address the specifics of the superintendent role, the program must also be in keeping with current research and program practice. In other words, did the SQP include the processes for achieving its "enabling" (i.e., the activities needed to produce a goal) and "terminal" (i.e., the achievement of the program's primary goal) objectives necessary in preparing candidates for the superintendent role (Provus, 1971, pp. 55, 204)? In the DEM, the enabling elements of a program are assessed in the process stage. The terminal objectives are assessed in what Provus calls the product stage of evaluation (Provus, 1971, pp. 58, 209). To

evaluate whether the quality of the SQP was in keeping with similar programs, the evaluation included this comparative training area of the program. At this stage, the evaluators were dealing with the program in the absence of the legislative and regulatory standards used in the earlier design and installation phases of the evaluation, focusing only on the standards of career development defined in the SQP Provider's Guide.

To assess the enabling (or process) aspects of the program, it was necessary to focus on the presence of certain learning conditions that enabled SQP participants to pass the program and eventually become successful superintendents. To do this, the program was assessed on its ability to develop self-efficacy among its candidates. Self-efficacy refers to trainees' ability to master a training program so that they may successfully use this training in the performance of their jobs (Haccoun & Saks, 1998). During the training itself, self-efficacy can be increased by the presence and usefulness of verbal encouragement provided by colleagues and instructors before, during, and after training; the presence and usefulness of role models; and opportunities for skills and knowledge application during training (Bandura, 1997). To assess self-efficacy development, candidates and instructors were asked to provide their thoughts on the presence and usefulness of colleague support and instructor support (e.g., verbal encouragement about the candidate's success in the program or as a superintendent), the use of role models in the SQP, and opportunities for SQP candidates to practice the skills and understandings acquired during the program.

To evaluate the terminal outcomes of the program, or in the DEM, the product, candidates were asked about two outcomes: the actual utilization of skills learned through the SPQ at their current jobs and the overall career development of successful SQP candidates. The skill utilization component of the evaluation was an extension of the skill development issues addressed in the commensurate expectations focus area. Candidates were asked to share their thoughts on the extent to which the selected skills and understandings learned during the SQP were actually being utilized in their current positions. It was thought that if these skills and understandings were being taught during the SQP, the acquisition of these skills would prompt changes in the performance of skills in a superintendent position.

To assess career outcomes, successful candidates were asked whether they became superintendents following their SQP participation or whether successful candidates furthered their career mobility fol-

lowing their completion of the program. Career development was measured by asking SQP candidates to state their employment position during the time of their program candidacy, whether they became superintendents following this training, or if they advanced to another higher administrative position following their candidacy. Unlike the cross-validation present in the self-reporting of candidates' skill developments discussed earlier, information was not available to confirm the validity of candidate responses. Therefore, these unsubstantiated self-reports are regarded as proxy responses because the relatively weak validity of survey research, particularly when candidates are asked to convey the totality of their career development in three standardized questions, potentially underestimates the actual career development impacts of the program (e.g., a successful SQP candidate does not pursue a superintendent's position, yet the SQP still provides skills and knowledge that intrinsically enhance her career). In the future, appropriate time and resources could allow evaluators to interview or survey the peers and supervisors of successful SQP candidates who could provide their observations of candidates' skill utilization on the job.

THE CASE STUDY METHOD: PROVIDING MULTI-METHODOLOGICAL "THICKNESS" TO THE SQP EVALUATION

Stufflebeam (2001a, p. 18) hints how models such as the DEM are very popular among administrators who are used to standards-based assessments. The provincial governing body's familiarity with curriculum auditing procedures demonstrated how acceptable the DEM would be. However, when evaluation standards are defined according to acts, regulations, and preconceived guidelines, the DEM, like other objectives-based models, runs the risk of producing an evaluation that is quite narrow in scope. The evaluation of any program is a complex undertaking that potentially goes beyond a literal interpretation of the appropriate legislative or regulatory standards. The SQP evaluation could have simply examined the degree to which the program was aligned with the Ministry's and provincial governing body's expectations. Beyond that, however, is the possibility that an evaluation goes more deeply into the program to determine not only whether the program meets these expectations but also whether it meets the needs and expectations of its candidates. It was determined that the evaluation would move from the assessment of alignment discrepancies to an assessment of outcomes and quality.

To do this, a case study strategy was incorporated into the SQP evaluation framework. A case study involves the evaluation of a

"case," meaning a program in which a person or people are operators, managers, recipients, and so on (Stake, 1995, 2005). The case study approach was selected for several reasons. First, a case study is used in circumstances where a more holistic understanding of the program is desired or required. Second, it enables evaluation researchers to investigate aspects of programs that are not clearly identified in formally defined program standards or boundaries (Yin, 2003, p. 13). Also essential to the case study method is the inclusion of questions that are of primary concern for program stakeholders. The DEM already allowed for the inclusion of stakeholders. The case study's inclusion of stakeholders complements this, as program stakeholders are also an integral part of this decision-making process (Stufflebeam, 2001a, p. 34). Admittedly, the SQP evaluation was not written in the narrative style advocated by certain case study evaluators (Stake, 2004, pp. 94–96; Stufflebeam, 2001a, p. 35). However, the more detailed accounts of the SQP instructors and candidates were extremely helpful in elaborating on standards-based observations found in the collection of regulatory, provider, and governing body documentation. The combination of the standards-based DEM with the case study methodology generated an evaluation framework that was both theoretically and methodologically complementary. By structuring the areas of focus to be evaluated, the DEM provided focus for the evaluation. However, the flexibility of the definitions constructed within each focus area permitted an extension of the evaluation beyond standardized questions about administrative alignment (e.g., providing a simple checklist of standards being met or not) to a more detailed assessment of the SQP.

Bringing Data Triangulation into the SQP Evaluation

The case study method was judged to be an appropriate complement to the DEM for these reasons, but it was also attractive for its employment of multiple methods that enable evaluators to texture their assessments of program goals, objectives, outcomes, or experiences. In line with this, the case study approach emphasizes the use of triangulation in data analysis. Yin (2003, pp. 98–99) outlines four triangulation protocols: data triangulation (whether observations are replicated by more than one data source), investigator triangulation (whether interpretations are replicated from one evaluator to another), theory triangulation (the comparison of evaluator interpretations of a single piece of data), and methodological triangulation, or the verification of observations by way of multiple research methods (for other explanations of triangulation see Denzin, 1984; Patton, 1987;

Stake, 2004). For the SQP evaluation, the evaluators focused on the achievement of data triangulation, which was employed through the use of four sources: questionnaires, semi-structured interviews, focus groups, and artifact reviews (content analysis). The details of these methods are summarized below. Site visits to each of the SQP providers' facilities were also considered, but not included due to time and resource limitations. Given the popularity of the curriculum auditing function with members of the advisory committee, and given the extolling of data triangulation by certain curriculum audit scholars (English, 1988, p. 46), this strategy received strong support.

With data triangulation, evaluators are encouraged to generate findings about a program that are replicated in multiple sources. When evaluators are able to support an identified observation about a program with two or more data sources, then data "convergence" has been achieved (Yin, 2003, pp. 99–100). That is, an observation is only considered to be an evaluation finding when that observation has been supported by two or more data sources. In the analysis of evaluation data, however, there will also be numerous observations that do not meet the triangulation criteria. Here, it becomes the evaluator's decision to either reject these observations or consider them as serious "rival hypotheses" (Stake, 2004; Yin, 2003). Rival hypotheses may come in the form of unique experiences or issues that require special attention, as their limited observation does not necessarily imply a lack of importance to the program's merit or worth. Nevertheless, only those findings that were substantiated by two or more data sources, and were valuable to addressing any of the three focus areas defined for the evaluation, were included in the final SQP evaluation report.

All four data sources were analyzed for convergent observations. However, some data sources were more effective at capturing certain observations than others. The convergence of the four data sources with SQP findings is summarized in Table 3. To illustrate, the alignment of the SQP's curriculum content with Ministry and governing body standards (see Table 2) was confirmed in the review of document artifacts. While alignment was generally consistent with the expectations of the Ministry and the provincial governing body, there was some lack of cohesiveness among modules with respect to course content and its instruction. Focus group and interview data with SQP instructors suggested how the establishment of more collaboration between module instructors would benefit the program so that curriculum and instructional consistencies could be achieved. In the evaluation of the commensurate expectations area, all four data sources confirmed the provision and delivery of practicum experi-

ences and in-class mentor presentations. With respect to the comparative training area, data from questionnaires, focus groups, and interviews identified that successful program candidates were using the skills and understandings acquired through the SQP, especially for understanding the influence that government legislation has on problem-solving and decision-making strategies.

Table 3
The Convergence of Selected Data Sources with Evaluation Focus Area Findings

Methods	Focus Areas		
	Alignment	Commensurate expectations	Comparative training
1. Questionnaires		✓	✓
2. Interviews	✓	✓	✓
3. Focus groups	✓	✓	✓
4. Artifact reviews	✓	✓	

Data triangulation is an intensive activity. When deciding to use data triangulation in an evaluation, it is advised that evaluators form partnerships in which skill sets can be balanced. Alternatively, an evaluation team can include an evaluator who is well versed in methodological literature and practice. In the SQP evaluation, a partnership was formed between an educational consultant with extensive experience with provincial acts, regulations, governing body documents, and focus group facilitation, and a sociologist with a relatively extensive background in the teaching and application of survey research and interview techniques. The four methods adopted for data collection are summarized here.

1. Questionnaires to SQP Candidates and Instructors[4]

The SQP evaluation included the administration of two mail-out questionnaires, each in English and French versions. An eight-page questionnaire was administered to candidates who successfully completed the SQP within the past three years (excluding current candidates), and a six-page questionnaire was administered to all instructors of the SQP during the past three years (including past and current instructors). The questionnaires were designed and administered using the Tailored Design Method (Dillman, 2000). The mail-out process involved three steps. First, pre-letters were sent to inform potential respondents that a questionnaire would be mailed to them within approximately one week. Questionnaire packages

were then administered to respondents approximately one week later. Each package contained an information letter written on the provincial governing body's letterhead, one questionnaire, and one return envelope with prepaid postage provided. The questionnaire packages were mailed in envelopes with the provincial governing body's insignia visible in the top corner. It was important that the survey's sponsorship be known, hence increasing the legitimacy of the survey. To further ensure a high response rate, follow-up letters were e-mailed to potential respondents three weeks following the mail-out of questionnaire packages. Data collection commenced in late February 2004 and ended in late April.

The sampling lists for SQP candidates and instructors were provided by the four providers. Following the elimination of certain individuals whose addresses were not provided and the elimination of name duplications across providers, a total of 291 questionnaires were initially sent. Instructors were sent 55 of these, and 236 were sent to candidates across each of the providers. Five instructor questionnaires and seven candidate questionnaires did not reach their destinations. Of 50 possible instructor responses, 35 questionnaires were returned for analysis (response rate = 70%). Out of 229 possible candidate responses, 152 questionnaires were returned (response rate = 66.4%). It should be noted that the e-mail follow-up letters increased the instructor response rate by 8 questionnaires and increased the candidate response rate by 26 questionnaires.

The questionnaires addressed the commensurate expectations and comparative training focus areas whereby candidates and instructors were asked to give their thoughts on the extent to which the SQP developed 10 skills and understandings expected in the performance of educational administration positions. Self-efficacy was also addressed in the questionnaires. The candidate questionnaire asked SQP candidates exclusively to assess the extent to which they actually utilized those skills and understandings developed through the SQP at their current jobs. The comparative training area of focus was addressed through measures of career development and skill utilization. Several demographic questions were also requested of candidates, including the gender, age, education, and job tenure of the respondents as well as the respondent's school board district characteristics. In addition to these characteristics, instructors were asked to state the provider for which they instructed. Univariate and bivariate statistics were generated for the analysis, but due to low (and nonrandom) sample sizes and low variance levels, inferential statistics and tests of significance could not be performed.

2. Telephone/In-Person Interviews

Semi-structured telephone interviews of 30 to 40 minutes' duration were held with 54 candidates and 12 instructors from each of the four providers. These totals reflected a 20% stratified random sample from each provider to ensure a balance of gender and providers. The sample came from provider lists of candidates and instructors who had been involved with the program within the past six months. All interviewees were given an initial telephone call to determine their willingness to participate in the interview. This call was followed by a covering letter and a copy of the questions to be used as a basis for the interview. All questions were designed to gather information for all three of the defined focus areas. Of the 66 interviews, 4 were in person; the rest were by telephone.

3. Focus Groups

Ten focus groups were held in various locations throughout the province. Four focus groups included candidates only; four included instructors only; two included both candidates and instructors. In all cases, the program providers were asked to assemble these groups, which were facilitated by a member of the provincial governing body and/or one of the evaluators. One group was successfully handled by teleconference. In all, 20 people took part in the instructor focus groups and 41 in the candidate focus groups.

When the participants were assembled, they were first presented with an overview of the form and purpose of the review and were given copies of the definitions of the three focus areas and their indicators. Participants were told that for the purpose of the focus group exercise, these descriptions were to be seen as those of an ideal SQP. They were asked to match their actual experiences in the SQP against the ideal as presented. They were then asked to formulate responses to the tasks, first as individuals and then working with their respective groups to formulate consensus statements. Only the consensus statements were recorded by the focus group facilitator for insertion into the evaluation report.

4. Artifact Reviews

A small collection of documentation associated with the SQP was also reviewed. This documentation was divided into two groupings: background documentation and provider submissions. In the case of the background documentation, the materials were textually analyzed in part to determine the degree of congruency among them and in

part to ascertain whether or not there were consistent themes or strands that could be used to develop a meaningful context for both the submission by the providers and the data presented through the interviews, the questionnaires, and the focus groups. This background documentation included the acts that establish the authority of the provincial governing body to accredit educational programs responsible for the certification of superintendents, the regulations that detail the qualification of superintendents, and guides that define the general content and evaluation requirements of the SQP. In the case of provider submissions, documentation evaluated here related to the clarity and consistency of admission requirements, record-keeping procedures, length, scheduling and hour requirements, module structures, candidate assessment for eligibility, practicum requirements, and candidate assessments and evaluations.

ENCOURAGING EVALUATION USE

Several evaluation models developed over the past three decades see stakeholder involvement as essential to the achievement of utilization (for example, see Cousins & Earl, 1995; King, 1998; Patton, 1997; Stufflebeam, 2001b; Wholey, 1983). Despite the amount of time and resources devoted to an evaluation, the possibility remains that the evaluation, and its potentially valuable information, will not be used in the final assessment of or decision about a program. Important to ensuring utility is the inclusion of stakeholders in the evaluation. By including stakeholders, evaluators account for the political and cultural context in formulating questions, methods, and assessment approaches (Alkin, 2004, p. 299).

At the end of the SQP evaluation, provider representatives of the advisory committee were invited to provide feedback on the evaluation's findings and recommendations. Provider representatives commented on the general evaluation process, the use of the advisory committee itself, the value of the evaluation, its methodology, findings, and recommendations. Overall, the providers strongly approved of the evaluation's design, particularly the formation of an advisory committee, and would consider many of the recommendations. One provider agency summarized its impressions about the advisory committee when it stated

> The ... goal to develop "a cooperative transparent endeavour" was successfully achieved. This provider, through involvement on the advisory committee, was fully informed and aware of all aspects of the review. The collaborative

and consultative review format also presented the opportunity to clarify issues and offer input regarding the implementation of the review. It also afforded us the opportunity to meet and discuss the program with other providers. The [provider], as the primary provider partner, supports the review format for future use.

Another provider praised the evaluation but was honest about a limitation of the survey in making the distinction between older and recent candidates:

> The [provider] appreciated the opportunity to take part in the … review process set up … early in 2004. The process was well planned and the survey was well organized. Our only quibble is that we would have preferred it if the survey had differentiated between the old and new programs.

Due to the unique perspectives and needs of each provider, certain recommendations were more salient for certain providers, but the provider representatives were highly appreciative of their involvement in the evaluation and honest participants in the process.

Patton (1997, p. 90) states how the collaboration between evaluators and stakeholders encourages the use of evaluation findings. However, there is also the potential for process use when stakeholders are able to become more intimately involved in the evaluation practice itself. Also helpful here is the act of teaching stakeholders how to conduct future evaluations. As noted above, one of the providers advocated the use of this evaluation framework for future use. Following the completion of the SQP evaluation, the provincial governing body hired the evaluators to perform a two-day training seminar on the steps and techniques applied in the SQP evaluation. Program developers and accreditation officers from across the province were given presentations on standards and indicator development and data-collection methods. Following these presentations, participants were given opportunities to apply the SQP evaluation framework to realistic evaluation scenarios. Feedback from the evaluation exercise demonstrated much appreciation for these application-oriented sessions.

The majority of participants were enthusiastic about the opportunity to learn evaluation techniques. However, some others were lukewarm to the idea, notably the more experienced program developers and officers who seemed comfortable with the standards-based notions of evaluation and quality assessment, but were uneasy about the case

study aspects of the SQP evaluation framework, which they perceived as ambiguous and legally precarious. For them, realities such as program effectiveness, efficiency, and quality were to remain closely tied to the articulation of Ministry and governing body standards. Others were more open to the addition of the case study approach, recognizing how it provided unique experience-based definitions of the program's design, delivery, outcomes, and quality. Still others even believed that their knowledge of evaluation was extensive enough and found the evaluation seminar redundant. These individuals were the exceptions, but it was an interesting observation that in a professional setting where participants are highly educated and experienced, not everyone preferred this self-evaluation exercise. This suggests that self-evaluation initiatives be adopted with care in professionalized contexts or that evaluators thoroughly consider their audience(s) in advance of self-evaluation opportunities.

CONCLUSION

This evaluation framework was the first of its kind for the provincial governing body and is expected to be influential in similar evaluations of educational administration preparation programs in the province. The prior evaluation of the SQP was judged to have lacked a theoretically sound and methodologically rigorous evaluation framework. The evaluation framework outlined here provided an alternative that combined two seemingly different theoretical and methodological approaches: the DEM and the case study. The DEM provided a focused objectives-based foundation in the formulation of the three focus areas: alignment, commensurate expectations, and comparative training. This especially fulfilled the conceptual needs of the provincial governing body and SQP program providers who requested that the evaluation identify potential alignment discrepancies. To assess the installation, process, and product of the SQP, the two other areas of focus were developed. It was the case study's provision of multi-methods and data triangulation that allowed the evaluators to contextually assess the SQP's installation along with its enabling and terminal outcomes beyond the acts and regulations of the province and the SQP governing body.

In conclusion, other lessons have been learned from the combination of objectives- and methods-based evaluation approaches that future evaluators of similar educational administration qualification programs could consider:

1. Defining Standards with Flexibility

For the evaluation of qualification programs such as the SQP, it is worth considering how an evaluation can transcend the assessment of regulatory alignments to explore how it can address program issues more extensively. Clear definitions of existing standards provide concrete directions for the evaluation. This is required for the effectiveness of discrepancy evaluation or any similar standards-based evaluation model. However, the inclusion of stakeholders in the evaluation enables the framework to be guided by more flexible, yet standards-related, areas of focus. A methods-based strategy such as the case study can take the evaluation beyond standards-based analyses to a more textured assessment of program design, delivery, and outcomes.

2. Balancing Multiple Program Administrative Norms with Data Triangulation

Equally important to developing strong areas of evaluation focus is data collection. The case study's strength is its openness to multiple data sources. A multi-methodological evaluation is a daunting task; it requires the presence of at least one evaluator with expertise in a selection of research methods. When the appropriate resources are available, the performance of data triangulation can strengthen the validity of findings and subsequent recommendations. The verification of findings according to their presence in at least two data sources provides an efficient strategy for data analysis and reporting. This also generates a vast amount of quantitative and qualitative data pertaining to program issues and outcomes.

The strength of the case study method is also a potential weakness because the multi-method evaluation, with its gathering of multiple perspectives, can lose focus on key areas of interest (Stufflebeam, 2001a, pp. 35–36). This can also be a concern for policy traditionalists who demonstrate reservations about multiple program realities that deviate from well-defined regulations. This concern can be abated through the combined use of a methods-based approach such as the case study with a standards-focused framework such as the DEM. However, caution is also advised for future evaluators of similar qualification programs. The training seminar that followed the SQP evaluation involved one enclave of program officers whose opinions leaned toward a traditional standards-based approach. They conveyed the hypothetical legal ramifications of evaluating government-regulated programs with methods that permitted multiple program realities

with respect to program quality. Another enclave was relatively liberal in accepting qualitative methods and the multiple program realities associated with them. This dichotomy deviated slightly from the unity seen on the advisory committee. Weiss (1993) reminds evaluators of the political aspects of program evaluation. The SQP evaluation demonstrates how methodological choices have philosophical as well as political consequences that can emerge at any phase of an evaluation from its initial design to its final promotion for future use. Before determining a methodological strategy, it is prudent to be familiar with the administrative or organizational cultures from which stakeholders or evaluation learners are coming.

3. Establishing and Maintaining Evaluator and Stakeholder Cooperation

As the previous lesson hints, it is an advantage for evaluators and program stakeholders to take a cooperative approach in evaluating the program. The use of an advisory committee representing all immediately relevant stakeholders is a constructive option for building cooperation from the earliest stages. The formation of an advisory committee secures the legitimacy of the evaluation's focus, design, and outcomes. However, in an evaluation such as that conducted of the SQP, evaluators must also ensure that stakeholders understand the advisory role of the stakeholder committee and that the evaluators ultimately steer all aspects of the evaluation.

4. Enhancing Use Potential

Finally, the incorporation of a feedback stage and training seminar at the conclusion of the evaluation enables stakeholders to engage the evaluation and add their insights about the framework. It provides a learning component during which stakeholders can interact with, and reflect on, different program evaluation perspectives while debating an evaluation framework's use potential. The SQP training seminar was generally successful in this regard, but something for future evaluators to consider is the salience that such activities have for those who participate. Sharing evaluation knowledge for the betterment of future programs works best when people are willing and interested participants in accomplishing that end. Ideal to the success of these evaluation use initiatives is the presence of a strong learning organization wherein the possession of appropriate capacities and commitments by its individuals makes learning about evaluation, and its use, salient to all (Cousins, Goh, Clark, & Lee, 2004). In the absence of these conditions, however, initiatives by evaluators to

conduct learning practices for the future use of evaluation nevertheless provide those first steps in the integration of evaluation into the organization's culture.

ACKNOWLEDGEMENTS

NOTES

1. In this article, the organization that is responsible for the certification and regulation of the teaching profession in this province will be referred to as the "provincial governing body" or "governing body." Provider agencies of the program will be called "providers."

2. It was the wish of the provincial governing body responsible for the superintendents' qualification program that the program remain anonymous. Therefore, the name of the program has been altered for this article.

3. In the province of interest, the position of superintendent is synonymous with the position of director in some other Canadian provinces, and is responsible for overseeing specialized school board functions such as curriculum, human resources, administration, finance, or facilities. In the province of interest, superintendents answer to a director of education. In other provinces, this director would be called a superintendent or chief superintendent of schools.

4. At the provincial governing body's request, all data collection instruments and specifically stated findings are to remain confidential.

REFERENCES

Alkin, M.C. (2004). Context-adapted utilization: A personal journey. In M.C. Alkin (Ed.), *Evaluation roots: Tracing theorists' views and influences* (pp. 293–303). Thousand Oaks, CA: Sage.

Alkin, M.C., & Christie, C.A. (2004). An evaluation theory tree. In M.C. Alkin (Ed.), *Evaluation roots: Tracing theorists' views and influences* (pp. 12–66). Thousand Oaks, CA: Sage.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Cousins, J.B., Goh, S.C., Clark S., & Lee, L.E. (2004). Integrating evaluative inquiry into the organizational culture: A review and synthesis of the knowledge. *Canadian Journal of Program Evaluation, 19*(2), 99–141.

Cousins, J., & Earl, L. (Eds.). (1995). *Participatory evaluation in education: Studies in evaluation use and organizational learning*. London: Falmer.

Denzin, N. (1984). *The research act*. Englewood Cliffs, NJ: Prentice Hall.

Dillman, D. (2000). *Mail and internet surveys: The Tailored Design Method*. New York: John Wiley.

English, F. (1988). *Curriculum auditing*. Lancaster, PA: Technomic.

English, F. (2000). *What to teach and test. Developing, aligning and auditing the curriculum. Millennium edition*. Thousand Oaks, CA: Corwin.

Fox, S., & Yossi, D. (1988). Validity of self-assessment: A field evaluation. *Personnel Psychology, 41*, 581–592.

Haccoun, R.R., & Saks, A.M. (1998). Training in the 21st century: Some lessons from the last one. *Canadian Psychology, 39*(1–2), 33–51.

Human Resources and Skills Development Canada. (2007). *National Occupational Classification/Classification nationale des professions*. Retrieved May 20, 2007, from <http://www23.hrdc-drhc.gc.ca/2001/e/generic/welcome.shtml>.

Johns, G. (1994). How often were you absent? A review of the use of self-report absence data. *Journal of Applied Psychology, 79*(4), 574–591.

King, J. (1998). Making sense of participatory evaluation on practice. *New Directions for Evaluation, 80*, 57–67.

Patton, M.Q. (1987). *How to use qualitative methods in evaluation*. Newbury Park, CA: Sage.

Patton, M.Q. (1997). *Utilization-focused evaluation. The new century text* (3rd ed.). Thousand Oaks, CA: Sage.

Provus, M. (1971). *Discrepancy evaluation*. Berkeley, CA: McCutchan.

Stake, R.E. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.

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

Stake, R.E. (2005). Qualitative case studies. In N.K. Denzin & Y.S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed., pp. 443–466). Thousand Oaks: Sage.

Steinmetz, A. (1983). The discrepancy evaluation model. In G.F. Madeus, M.S. Scriven, & D.L. Stufflebeam (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation* (pp. 79–99). Boston: Kluwer-Nijhoff.

Stufflebeam, D.L. (2001a). Evaluation models. *New Directions for Evaluation, 89*(Spring). San Francisco: Jossey-Bass.

Stufflebeam, D.L. (2001b). Interdisciplinary PhD programming in evaluation. *American Journal of Evaluation, 21*, 293–314.

Weiss, C. (1993). Where politics and evaluation research meet. *Evaluation Practice, 14*(1), 93–106.

Wholey, J. (1983). *Evaluation and effective public management*. Boston: Little, Brown.

Yin, R.K. (2003). *Case study research: Design and methods. Sage applied social research methods series, 5*. Thousand Oaks, CA: Sage.

Tim Gawley is an assistant professor of Leadership at Wilfrid Laurier University's Brantford Campus, where he teaches research methods and evaluation research for Laurier Brantford's multidisciplinary Contemporary Studies, Criminology, and Leadership programs. He currently collaborates with St. Leonard's Community Services of Brantford on a number of evaluation projects. Other interests include interpretive theories, qualitative methods, and the sociology of occupations and professions in everyday life, with a recent focus on educational administrators.