I. Proposed Title

Research on Evaluation

II. Guest Editor

Paul R. Brandon
University of Hawaiʻi at Mānoa
College of Education
Curriculum Research & Development Group
1776 University Avenue, Castle Memorial Hall
Honolulu, HI 96822-2463
(808) 956-4928
brandon@hawaii.edu

Paul R. Brandon, PhD, is Professor of Education at Curriculum Research & Development Group, College of Education, University of Hawaiʻi at Mānoa (UHM). Dr. Brandon has been conducting formative and summative educational program and project evaluations (both internal and external), research on program evaluation, and educational research since 1977. He is on the graduate faculty of the Department of Educational Psychology at UHM, where he teaches evaluation courses and chairs dissertation and thesis committees. Since beginning his position at UHM in 1989, he has been the principal investigator (PI) or co-PI for education research or evaluation grants and contracts totaling about $6M, including some for the National Science Foundation and U. S. Department of Education. In addition to conducting his own projects, he serves as the chief consultant to his curriculum-developer colleagues on research and evaluation design, methods, analysis, and reporting. He is Editor-in-Chief of *New Directions for Evaluation* and the author or co-author of nearly 50 refereed journal articles and about 175 evaluation and technical reports. With Nick L. Smith, he was co-editor of the 2008 book, *Fundamental Issues in Evaluation*, and is finishing his tenure as co-editor of the *American Journal of Evaluation* Exemplars section. He has won two national best-evaluation awards from Division H of the American Educational Research Association (AERA) and in 2011 was designated as a distinguished scholar by the AERA Research on Evaluation Special Interest Group.

III. Brief Overview

Research on evaluation (RoE) is systematic inquiry about the inputs into evaluations, the context within which evaluations are conducted, evaluation activities, evaluation outcomes, professional issues, and the implications of the findings on these topics for our understanding of evaluation theory, methods, and practice. In recent years, the breadth and growth of RoE has been well documented; for example, of the *American Journal of Evaluation* (AJE) articles reviewed for the period 1998—2011, 37% were instances of RoE (Brandon, Vallin, & Philippoff, 2012). The breadth of RoE is reflected in 11 systematic reviews of the RoE literature published in AJE since 2004. Summarizing the reviews of the RoE conducted during a period of about 30 years, King and Stevahn (2013) concluded, “Even as we acknowledge the relative limitations of the existing research base, taken together these syntheses provide direction for program evaluators who want their use-oriented practice to be evidence-based” (p. 60). Other manifestations of RoE as a growing endeavor abound: Mark’s (2008) framework for conceptualizing and conducting RoE is increasingly cited in the literature; Cousins and Chouinard (2012) produced a full book summarizing the RoE literature on participatory evaluation; the American Evaluation Association sponsors the RoE Topical Interest Group; and the American Educational Research Association sponsors the RoE Special Interest Group, which annually issues an award honoring a leading RoE researcher. (Of the authors of chapters in this proposed *New Directions for Evaluation* [NDE] RoE issue, two are recipients of the award.) In eight chapters, the authors of the issue proposed here will cover RoE context, practice, methods, effects, capacity building, and training, with some chapters providing the results of original research. This coverage addresses four of the topics (methodological focus, professional practice, theoretical focus, and exemplary studies) given in the list of types of NDE issues in the journal's Proposal Submission Guidelines. Furthermore, the issue will address NDE's call for presenting original RoE (Brandon, 2013).

IV. Justification

Scope

Worthen (1990) stated, "Since the early 1970s, Stufflebeam, Worthen and Sanders, Smith, and others have issued a call for evaluation to be researched to develop an adequate knowledge base to guide evaluation practice" (p. 46). This call for research has continued (Alkin, 2003; Brandon & Ah Sam, in press; Cousins & Chouinard, 2012; Cousins & Earl, 1995; Henry & Mark, 2003; Mark, 2008; Smith, 1993). Numerous suggestions have been made. Smith (1993) suggested that studies of practice can lead to descriptive theories of practice. Mark (2008) encouraged RoE studies about evaluation context, activities, consequences, and professional issues. Stufflebeam and Shinkfield (2007) listed RoE topics including, among others, the effects of stakeholder involvement on evaluation use, the effects of evaluator-stakeholder collaboration on use, the study of politics in evaluations, the effects of evaluations of program implementation, and a number of methodological issues. These topics fall within the four general categories of evaluation practice (including context and politics), evaluation methods, the profession of evaluation, and evaluation theory—the coverage of the NDE issue that we are proposing here.

The contributors to this proposed NDE issue intend to broaden the RoE literature further in eight chapters. The first two chapters will be reports of original research *on* RoE. They will discuss previously published RoE, its implication for evaluators and future RoE, and evaluators perspectives on RoE:

- Ch. 1: Overview of Published Research on Evaluation (Vallin, Philippoff, & Brandon).
- Ch. 2: Evaluators' Perceptions of Published Research on Evaluation (Lewis, Ah Sam, Harrison, & Brandon).

The third through the fifth chapters will address the settings within which RoE happens, how it might be conducted, and its effects:

- Ch. 3: Advancing Research on Evaluation Through the Study of Evaluation Context (Vo & Christie).
- Ch. 4: Using Design Research to Research and Develop Evaluation Practice (Smith).
- Ch. 5: Assessing the Non-Academic Impact of Research on Evaluation (Cousins, Svensson, Szijarto, Pinsent, Andrew, & Sylvestre).

The sixth and seventh chapters will present the results of original RoE studies about learning evaluation:

- Ch. 6: Evaluation Capacity Building of Both Program Personnel and Evaluators in a Stakeholder-Based Evaluation (Lawton, Seraphin, Philippoff, Vallin, Harrison, & Cooper).

- Ch. 7: Evaluation Education: The State and Role of Evaluation Certificate Programs (Lavelle).

The final chapter will wrap up the discussion with prognostications about the future:

- Ch. 8: The Future of Research on Evaluation (Azzam).

Novelty and Timeliness

Evaluation (particularly program evaluation) is an empirical endeavor that uses systematic procedures to gather and provide evidence about the merit, worth, value, significance, and utility of evaluands. As an evidence-collecting and evidence-providing practice, it might be thought that the theory, methods, and practice of evaluation are themselves solidly based on the empirical study of evaluation. Evaluation, and the evaluands it examines, however, come in many sizes, scopes, purposes, and organizational, political, cultural, and societal contexts, making it complex and expensive to examine evaluation empirically. Furthermore, little funding to examine evaluation empirically has been available. In the late 1970s and early 1980s, RoE was conducted at the Northwest Regional Educational Laboratory, led by Nick L. Smith, one of the contributors to this proposed issue, and at the University of California at Los Angeles, led by Marvin Alkin and others. The National Science Foundation has funded two programs for examining evaluation empirically, the Evaluative Research and Evaluation Capacity Building program in the early 2000s and, currently, the Promoting Research on Innovative Methodologies for Evaluation program. Much of the other RoE has been done by individual academics. King and Stevahn (2013) stated that the dearth of RoE might also be due to the relative newness of the discipline and profession; disagreements about the definition, purpose, and methods of professional evaluation; and the practical nature of evaluation, which limits the likelihood of systematic research and theory building on the topic.

Fortunately, RoE has been increasing over the years. Many studies have continued to be ad hoc, driven in part by academic publishing requirements. Some evaluators conduct small add-on RoE studies to their evaluations (Brandon, 2012) such as examining evaluation use, conducting research on stakeholder involvement, or reporting on the degree of success of applying an evaluation approach such as empowerment evaluation. RoE that addresses evaluation methods is also widespread; evaluators often publish the instruments that they have developed and validated in their evaluation studies, and work on experimental design is active and expanding (e.g., Cook, Shadish, & Wong, 2008).

The proposed issue will report on recent efforts on RoE and expand the breadth of the literature. This will be a novel contribution, in that it will be the first NDE issue devoted to the topic; indeed, apart from many narrative case reflections on evaluation practice, little original RoE has been reported in NDE issues. The findings reported in this issue will not only add to our understanding of evaluation practice, methods, theory, and the profession, they will also serve as exemplars of various types of RoE. The proposed issue will represent a timely coming of age of RoE—a manifestation of the possibility of publishing entire volumes on the topic and serving as a motivator for others to do the same in other venues.

Quality

The quality of the proposed issue is demonstrated by its content and by its lineup of authors. The content will address the major components of the evaluation discipline: theory, methods, practice, and professional issues. The chapters will use a variety of research methods, cover both narrow and broad topics, and serve as exemplars for different types of RoE. The authors include seasoned evaluators, including those who initiated early efforts at conducting RoE and have long called for more, researchers whose regular workload includes considerable RoE, and newcomers who have fresh insights as researchers focusing on evaluation topics. Graduate students in particular might find the issue useful as they develop their own research agendas in evaluation.

Audience

Evaluators of all stripes are likely to find the chapters in the proposed issue to be of value. Researchers conducting empirical studies will learn about the existing coverage of RoE topics in the literature, gaps in the coverage, designs that they might use, their peers' preferences for methods in RoE studies, as well as findings of RoE useful in their own practice and teaching. Evaluation practitioners and teachers will be able to add to the body of knowledge about evaluation context, RoE design, evaluation capacity building, evaluation certificate programs, and the effects of RoE outside of academia.

Level of Presentation

The research reported here will be accessible to anyone with a minimal knowledge of evaluation and research methods, yet intriguing to the most sophisticated evaluators and researchers. The methods of the research reported here will include literature reviews, Internet searches, online questionnaire surveys, case studies of evaluation practice, and peer-debriefing, thus making the issue agreeable to evaluators steeped in either quantitative or qualitative approaches.

References

Alkin, M. C. (2003). Introduction to Section 3: Evaluation utilization. In T. Kellaghan & D. L. Stufflebeam D. L. (Eds.), *International handbook of educational evaluation* (pp.189–194). Dordrecht, The Netherlands: Kluwer.

Brandon, P. R. (2012, April). *Ruminations on research on evaluation*. Invited presentation to the Research on Evaluation Special Interest Group at the meeting of the American Educational Research Association, Vancouver, British Columbia.

Brandon, P. R. (2013). Statement of the Editor-in-Chief. In S. Nielsen & D. E. K. Hunter (Eds.), Performance management and evaluation (pp. 1–2). *New Directions for Evaluation, 139*.

Brandon, P. R., & Ah Sam, A. F. (in press). Program evaluation. In. P. Leavy (Ed.), *Oxford handbook of qualitative research.*

Brandon, P. R., Vallin, L. M., & Philippoff, J. (2012, October). *A quantitative summary of research on evaluation published in the* American Journal of Evaluation *from 1998 through 2011.* Paper presented at the meeting of the American Evaluation Association, Minneapolis.

Cook, T. D., Shadish, W. R., & Wong, V. C. (2007). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management, 27*, 724–750.

Cousins, J. B., & Chouinard, J. (2012). *Participatory evaluation up close: A review and integration of research-based knowledge*. Charlotte NC: Information Age Press.

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

Henry, G. T., & Mark, M. M. (2003) Beyond use: Understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation, 24,* 293–314.

King, J. A, & Stevahn, L. (2013). *Interactive evaluation practice: Mastering the interpersonal dynamics of program evaluation.* Thousand Oaks, CA: Sage.
Mark, M. M. (2008). Building a better evidence base for evaluation theory: Beyond general calls to a framework of types of research on evaluation. In N. L. Smith & P. R. Brandon (Eds.), *Fundamental issues in evaluation* (pp. 111–134). New York: Guilford.
Smith, N. L. (1993). Improving evaluation theory through the empirical study of evaluation practice. *Evaluation Practice, 14*, 237–242.
Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation theory, models, and applications*. San Francisco: Jossey-Bass.
Worthen, B. R. (1990). Program evaluation. In H. J. Walberg & G. D. Haertel (Eds.), *The international encyclopedia of educational evaluation* (pp. 42-47). Oxford: Pergamon.

V. Outline of the Volume

1. Proposed Table of Contents

Chapter 1. Overview of Published Research on Evaluation
Lisa M. Vallin, Joanna K. Philippoff, and Paul R. Brandon

Chapter 2. Evaluators' Perceptions of Published Research on Evaluation
Nicole Lewis, Anna Ah Sam, George M. Harrison, and Paul R. Brandon

Chapter 3. Advancing Research on Evaluation Through the Study of Evaluation Context
Anne T. Vo and Christina A. Christie

Chapter 4. Using Design Research to Research and Develop Evaluation Practice
Nick L Smith

Chapter 5. Assessing the Non-Academic Impact of Research on Evaluation
J. Bradley Cousins, Kate Svensson, Barbara Szijarto, Céline Pinsent, Caroline Andrew, and John Sylvestre

Chapter 6. Evaluation Capacity Building of Both Program Personnel and Evaluators in a Stakeholder-Based Evaluation
Brian E. Lawton, Kanesa D. Seraphin, Joanna Philippoff, Lisa M. Vallin, George M. Harrison, and Joanne Cooper

Chapter 7. Evaluation Education: The State and Role of Evaluation Certificate Programs
John M. Lavelle

Chapter 8: The Future of Research on Evaluation
Tarek Azzam

2. Abstract

For many years, major figures in evaluation have called for more research on evaluation (RoE). We propose a *New Directions for Evaluation* issue that will broaden the discussion of RoE. The proposed issue will include chapters on

- the results of original research about the extent to which RoE has been published in the *American Journal of Evaluation*, as well as a description of the major characteristics of the research;
- the results of a survey to be conducted on evaluators' knowledge of, attitudes toward, beliefs about, and uses of published RoE;
- research on evaluation context;
- an approach for studying evaluation practice;
- a model for examining the non-academic effects of RoE;
- original RoE capacity building and training (two chapters); and
- the future of RoE.

The research methods reported in the chapters include literature reviews and Internet searches, an online questionnaire survey, and a reflective case narrative. The proposed issue addresses the impetus to include more original RoE in NDE and should help move the conversation on RoE forward.

3. Chapter information, with a separate section for each chapter.

Chapter 1

a. Title: An Overview of the Research on Evaluation Published in the *American Journal of Evaluation*

b. Authors

Lisa M. Vallin
University of Hawaiʻi at Mānoa
College of Education
Curriculum Research & Development Group
1776 University Avenue, Castle Memorial Hall
Honolulu, HI 96822-2463
(808) 956-4920
vallin@hawaii.edu

Joanna K. Philippoff
University of Hawaiʻi at Mānoa
College of Education
Curriculum Research & Development Group
1776 University Avenue, Castle Memorial Hall
Honolulu, HI 96822-2463
(808) 956-4920
philippo@hawaii.edu

Paul R. Brandon
University of Hawaiʻi at Mānoa
College of Education
Curriculum Research & Development Group
1776 University Avenue, Castle Memorial Hall
Honolulu, HI 96822-2463
(808) 956-4920
brandon@hawaii.edu

c. Description

For decades, many contributors to the evaluation literature have called for conducting empirical research on evaluation (RoE) and decried the lack of such studies. However, with some exceptions about particular evaluation domains (e.g., Cousins & Chouinard, 2012), the actual prevalence of RoE studies has not been carefully documented empirically. Our purpose in this chapter is to address this deficit by documenting the number, research topics, designs, data-collection methods, and professional domains of a body of published RoE studies. This chapter builds on the results of an earlier, unpublished review of 586 articles in the *American Journal of Evaluation* (AJE) from 1998 through 2011, in which we identified 219 articles as RoE. We chose to examine AJE because the journal is the flagship publication of the leading evaluation organization. We chose 1998—the year the journal changed its name from *Evaluation Practice* to *American Journal of Evaluation*—as the starting date for our review because people in academic positions, who are the most motivated and have the most time to produce original RoE, are more likely to have chosen the journal for publishing their work after the name change.

To date, we have examined every article in the journal during the selected period except book reviews, editorial notes, and other obviously non-research pieces. We identified the articles that were RoE and classified them by research design. The design categories include single case study; multiple case study; cross-sectional; longitudinal; time series; posttest only, non-randomized; posttest only, randomized; pre-post non-randomized; pre-post randomized; simulation; instrument or method demonstration, development, or validation; oral history; meta-evaluation; non-systematic literature review; systematic literature review; literature review, method unknown; bibliometric study; reflective narrative; and other. This categorization schema reflects a broad definition of RoE; oral histories, for example, might include previously unpublished discussions of evaluation theory, methods, practice, or the profession. Furthermore, reflective narratives typically include the results of evaluators' ruminations on topics or events that they believe can contribute novel, confirmatory, or disconfirmatory results to the literature. We distinguished among reflective narratives and single or multiple case studies by whether they included descriptions of methods and the use of multiple methods (topics that are ignored in reflective narrative studies).

In the second phase of the review, which we will conduct for this chapter, we will further classify the articles according to the domain in which they were conducted (education, public health, criminal justice, and so forth), the research topics that the studies addressed, and the study data-collection methods. We will cross-reference the topics with research designs and discuss the implications of these findings, raising awareness about evaluation purpose within current disciplines and the intersection of these categories.

The substantial number of RoE articles that we identified in AJE confirms our impression that considerable work on the topic has been conducted. Evaluators writing in AJE have found many instances in which they have deemed it fit to share their findings and insights about evaluation. Our results show a slightly increasing number of RoE articles across the years. The increase was greatest in the first half of the first decade of the century and has slowed somewhat since then. This increase might reflect the expanding awareness of AJE as a source for presenting RoE in a reputable journal with a growing impact factor. The relatively large proportion of articles about instruments and methods reflects, we think, how evaluation is a growing discipline, with a steady stream of new instruments and an evolving methodological base.

The designs that we included in the schema for categorizing articles range from carefully designed studies using rigorous methods to personal reflections. The hope of an empiricist seeking strong findings might be that the bulk of the studies would include more of the former than the latter. However, of the 219 articles that we have classified as RoE to date, 73% are reports of single case studies, of evaluation instruments or methods, or reports of the reflections of evaluators about their work. An additional 16% are the reports of literature reviews, multiple-case studies, or oral histories. Very few of the other types of designs are found. These results suggest that the literature is replete with studies, largely descriptive, of evaluation practices and methods without manipulating variables, following cases over time, conducting simulations of evaluation scenarios, and so forth. However, the number of literature reviews, particularly in the most recent time period that we examined, is heartening to us. A strong case can be made that evaluation has moved past an era of the development of new models and approaches. The models and approaches that have appeared in recent years have tended to be variations on the themes of those developed earlier. The literature should have sufficiently matured, with numerous articles on topics, to warrant trying to summarize across it. A cursory review (Brandon, 2012), however, suggests that the findings of some of the reviews tend to be speculative or scattered because the studies that are reviewed are disparate or because the reviewed studies have conflicting findings.

Given the dearth of rigorously designed and rigorously conducted studies and concerns about the ability to draw valid conclusions from many of them, backed up with evidence from literature reviews that evaluation studies are difficult to summarize, some might doubt the value in extending our review on RoE to other journals. We suggest, however, that the steps reported in the chapter proposed here could serve as the foundation for a comprehensive effort that might prove beneficial to evaluators in several ways. Were a complete accounting of the RoE in the major evaluation journals to be conducted, with a listing of some of the categories addressed in this paper, a database might emerge that would be a significantly useful resource to evaluators searching for advice and guidance about how to address evaluation issues or seeking conclusions known empirically about evaluation issues. Developing such a database would be a major endeavor, and funding sources for it do not readily come to mind, but having what would essentially be an electronic version of an annotated bibliography might assist the profession and discipline in numerous ways. We will offer our results as the first step in developing this bibliography.

References

Brandon, P. R. (2012, April). *Ruminations on research on evaluation.* Invited presentation to the Research on Evaluation Special Interest Group at the meeting of the American Educational Research Association, Vancouver, British Columbia.

Cousins, J. B., & Chouinard, J. A. (2012). *Participatory evaluation up close: An integration of research-based knowledge.* Greenwich, CT: Information Age.

Chapter 2

a. Title: Evaluators' Perceptions of Published Research on Evaluation

b. Authors

Nicole Lewis
University of Hawaiʻi at Mānoa
College of Education
Department of Educational Psychology
1776 University Avenue, Wist Hall
Honolulu, HI 96822-2463
(808) 956-4296
nicole88@hawaii.edu

Anna Ah Sam
University of Hawaiʻi at Mānoa
Office of Student Equity, Excellence, and Diversity
2600 Campus Road, Queen Liliuokalani Center for Student Services
Honolulu, HI 96822-2463
(808) 956-9217
annaf@hawaii.edu

George M. Harrison
University of Hawaiʻi at Mānoa
College of Education
Curriculum Research & Development Group
1776 University Avenue, Castle Memorial Hall
Honolulu, HI 96822-2463
(808) 956-4919
georgeha@hawaii.edu

Paul R. Brandon
University of Hawaiʻi at Mānoa
College of Education
Curriculum Research & Development Group
1776 University Avenue, Castle Memorial Hall
Honolulu, HI 96822-2463
(808) 956-4920
brandon@hawaii.edu

c. Description:

Research on evaluation (RoE) is a burgeoning endeavor. With some exceptions, however, such as a review of the *American Journal of Evaluation* to identify RoE articles (Brandon, Vallin, & Philippoff, 2012) and a report of the results of an RoE needs assessment survey (Szanyi, Azzam, & Galen, 2012), little research *on* RoE has been conducted. Little is known, for example, about evaluators' knowledge of RoE, their attitudes toward it, their beliefs about its importance, and how much they apply its results in their practice or teaching. In addition, little is known about the extent to which evaluators' knowledge, attitudes, beliefs, and reported use of RoE varies by demographic characteristic.

In this chapter, we will help address this deficit by reporting the methods and results of an online questionnaire survey investigating evaluators' perceptions of RoE. Our intent in this chapter is to improve the field's understanding of how the body of RoE is perceived and potentially put into use by evaluators and teachers of evaluation.

Background and Research Questions

The large and increasing number of RoE articles suggests there is increasing interest in this research endeavor. In a review of all the articles in the *American Journal of Evaluation* (AJE) from 1998 through 2011, Brandon et al. (2012) classified 219 (37.4%) articles as RoE. (An updated and expanded report of this review will be reported in the chapter of this proposed journal issue by Vallin, Philippoff, & Brandon.) The number of RoE articles increased every year except one, with 48% of the 219 articles published in the final five years of the period examined. Most of the research designs reported in the 219 articles identified in Brandon et al. (2012) were one of six design types: *single case studies* (27%); research on the *methods of evaluation*, here considered a "design" in a loose sense (26%), *reflective case narratives*—retrospective essay-like pieces reporting on aspects of evaluation (19%); *systematic literature reviews* (7%); *multiple case studies* (5%); and *oral histories* (5%).With this large and varied set of RoE articles, practitioners and educators have the opportunity to draw from this literature to inform their practice.

Even though the body of RoE is growing, there is a lack of empirical research on evaluators' degree of attention to the research, their attitudes toward it or beliefs about it, and their application of its findings. We seek to address this deficit by conducting a survey study addressing three research questions:

1) To what extent do the respondents report knowing about RoE, and how does this vary within demographic characteristics?
2) What are the respondents' attitudes toward and beliefs about RoE, and how do they vary within demographic characteristics?
3) To what extent do the respondents report applying the results of RoE in their practice, teaching, or understanding of evaluation, and how does this vary within demographic characteristics?

Methods

We will develop a survey questionnaire and distribute it online. The questionnaire will be brief to help ensure a high response rate. The instructions to the participants will include a guiding definition of RoE. To help ensure commonalities in the respondents' view of our conceptualization of RoE, the most common types of designs used in the RoE literature will be described. For each of the three research questions, we will ask the respondents sets of Likert-scale questions (forming three unidimensional scales) about the topics addressed in the questions. Nuances about how to address the topics will emerge as we prepare and pilot-test items. We will include an open-ended question asking for any additional information the respondents might wish to report as a follow-up to their responses to the scale items. We will conclude with demographic questions about gender, years of evaluation experience, the settings in which evaluators work (university, community, or corporate), and so forth.

The authors will pay significant attention to item development and modification. We will prepare draft items for each research question and pilot test the items in cognitive interviewing (think-aloud) exercises in person with at least five participants and online with a group of approximately 50 respondents. Participants will be recruited from among evaluation educators and practicing evaluators in a variety of professional settings. Reviews by potential respondents will help increase the quality of the questionnaire items and the credibility of our results. The items will be revised, as appropriate, based on the results of pilot-testing.

Survey respondents will be recruited via email from the membership of the American Evaluation Association, which currently has approximately 8,000 members, and the Canadian Evaluation Society, which has about 2,000 members. We will limit the respondents to North Americans so as to reduce the heterogeneity of the sample. We will send emails, with a maximum of three follow-ups, during recruitment. Based on the recent example of Szanyi et al. (2012), we expect to have over 1,000 volunteer participants—a sufficient number for acceptably small standard errors of the mean. IRB approval will be obtained from the University of Hawaiʻi review board. The participants will be informed of the purposes, methods, and uses of the study and will be required to consent to the study before proceeding. Responses will be collected using SurveyMonkey or similar software.

In the analyses, we will produce descriptive statistics and examine the differences among groups in analyses of variance. We will identify themes in the responses to the open-ended question in content analyses using a grounded-theory approach.

Implications

The findings of our study will have the potential to contribute to the continuing evolution of RoE in significant ways. The proposed study will increase our understanding of how RoE is perceived and used by those practicing and teaching evaluation. Identification and incorporation of this knowledge into future RoE is likely to lead to greater effects of RoE, including, but not limited to, development and publication of studies that are more likely to be read by evaluators, an increase in the number of RoE studies that are read by evaluators, and a proliferation in the application of RoE. The responses will also likely reveal the extent to which RoE is being used by practitioners and whether this use depends on demographic and professional characteristics, which may inform future RoE endeavors seeking to appeal to broad or to specific audiences. In addition to positioning researchers to be able to increase the effects of RoE, the findings are also likely to signify additional research questions that will be important to consider in establishing a research agenda for research on RoE.

References

Brandon, P. R., Vallin, L. M., & Philippoff, J. (2012, October). *A quantitative summary of research on evaluation published in the* American Journal of Evaluation *from 1998 through 2011*. Paper presented at the meeting of the American Evaluation Association, Minneapolis.

Szanyi, M., Azzam, T., & Galen, M. (2012). Research on evaluation: A needs assessment. *The Canadian Journal of Program Evaluation, 27*, 39–64.

Chapter 3

a. Title: Advancing Research on Evaluation Through the Study of Evaluation Context

b. Authors

Anne T. Vo
University of California, Los Angeles
Graduate School of Education & Information Studies
Social Research Methodology Division
Box 951521, Mathematical Sciences 8142B

Los Angeles, CA 90095-1521
annevo@ucla.edu
(310) 845-6779

Christina A. Christie
University of California, Los Angeles
Graduate School of Education & Information Studies
Social Research Methodology Division
Box 951521, Moore Hall 2022B
Los Angeles, CA 90095-1521
tina.christie@ucla.edu
(310) 825-0432

c. Description

Research on evaluation (RoE) is a sub-field within the evaluation discipline that takes the systematic examination of the field's theories, methods, and practices as its focus. Systematic studies of evaluation practice, however, are particularly challenging, perhaps because practice is often contextually bound. It follows, then, that we need an improved understanding of context to better understand practice. But, the notion of evaluation context itself needs to be unpacked.

A few operational definitions of "evaluation context" have been offered in the literature. The clearest is perhaps Greene's (2005), who referred to context as "the setting within which the evaluand (the program, policy, or product being evaluated) and thus the evaluation are situated" (p. 83). Earlier writings on this topic tended to be more specific and explored particular aspects of context that affected evaluation use, including stakeholder engagement and decision-making (see, for example, Patton, 1976; Stufflebeam, 1968, 1971; Stake, 1974, 1980; Weiss, 1972, 1973; Wholey, 1979, 1987).

While there has been longstanding interest in evaluation context, conceptual exploration and systematic investigation of context are recent developments. Rog (2012) and Vo (2013), for example, have independently developed frameworks for understanding the complexity of context (see also Alkin, 2012). Fitzpatrick, Christie, and Mark's (2009) collection of interviews with evaluators reflecting upon the ways in which they have responded to and addressed challenges that surfaced in practice is an example of such interest. Azzam's (2010) investigation of evaluators' responsiveness to various stakeholder groups and their expectations is yet another example of the field's desire to better understand context's influence on practice (see also Chouinard & Cousins, 2009; Christie, 2003; Rog, 2009).

Although there seems to be universal agreement in the field that context matters, we have yet to fully understand how it matters. For this reason, issues explored in this chapter are guided by these questions: What are some of the challenges associated with studying context and how might they be addressed? What are the implications for future RoE studies when context becomes the focus of an investigative endeavor? We argue that several obstacles contribute to the challenge of understanding context, one of which has to do with how embedded it is in evaluation practice and that one of the ways in which this challenge can be addressed is perhaps through the methodological choices made when studying context. We also outline some implications for approaching the study of context in this way on the RoE enterprise below.

Perhaps one of the greatest challenges encountered in the practice of evaluation is that the evaluand can be just about anything—a program, policy, consumer product, curriculum, etc. In the same vein, no two evaluands—even if they are of the same kind—are alike, perhaps because no two sets of contextual conditions can be exactly the same. Thus, when the need to conduct evaluation also requires explicit consideration for context, what quickly becomes apparent are not only reasons why evaluation is a challenging task to undertake, but perhaps more importantly how closely connected context is to evaluation practice, as well as systematic examinations of it. For example, context affects various aspects of an evaluation, including decisions about design and method choice, ways in which findings are communicated, as well as its perceived credibility (Julnes & Rog, 2008). Each of these elements could be an object of study in its own right, and each will pose a set of unique challenges during the investigative process. Issues about an evaluation's perceived credibility, for instance, are particularly challenging to tease apart in a systematic fashion because not only are they deeply embedded within context, but they are also intimately related to issues of culture—that is, the culture of stakeholders, evaluators, as well as the local and broader communities from which these individuals come (Mertens, 2008). Thus, the question then becomes, How do we best deal with a construct that is so complex and is often conflated with other equally complex aspects of evaluation?

Like many areas of evaluation, research on context is a nascent area of scholarship within the evaluation field. As such, one of the ways in which the challenge described above could be addressed is through the initial use of non-inferential research methods to understand the nature and influence of context on practice. Use of descriptive methods to gain a deeper understanding of context is critical because we are only starting to develop an understanding of this area. Inference-based methods at such an early stage of the inquiry process may lead to interesting findings, but would be slightly premature because they are more likely to be revised as context is better understood.

With this in mind, Rog's (2012) and Vo's (2013) context frameworks inspire a number of possible investigations. One such study, which could be carried out using either framework, might include a document analysis of evaluation reports where the unit of analysis would be the extent to which operational dimensions of context (e.g., scale, budget, etc.) were considered in the written evaluation. Subsequent investigations may dig deeper and attempt to shed light on other contextual dimensions—such as beliefs, values, assumptions, and norms of individuals and groups who are involved with the evaluand. Concept mapping, network analysis, and observation and interview methods could be helpful for such an endeavor.

Shifting the investigative lens on to context has conceptual, methodological, and disciplinary implications for the evaluation field. For example, we can reasonably anticipate that continued RoE context will likely lead to the refinement of existing (Rog, 2012; Vo, 2013) as well as the development of new conceptual frameworks of context. Development and refinement of more sophisticated descriptive methods could also be expected. Finally, as these bodies of conceptual and methodological literatures continue to grow and inform the RoE enterprise, the field will likely move toward increased intra- and interdisciplinary collaboration.

Due to the paucity of RoE studies, issues related to the study of context have yet to be examined in the fashion proposed here. As such, this chapter highlights and anticipates possible obstacles, solutions, and directions for future RoE work. Perhaps more importantly, it also outlines the ways in which RoE can be a tangible contribution to the field's practitioners as well as scholars.

References

Alkin, M.C. (2013). Context-sensitive evaluation. In M.C. Alkin (Ed.), *Evaluation roots: A wider perspective of theorists' views and influences* (2nd edition, pp. 283–292). Thousand Oaks, CA: Sage.

Azzam, T. (2010). Evaluator responsiveness to stakeholders. *American Journal of Evaluation, 31*, 45–65.

Chouinard, J. A., & Cousins, J. B. (2009). A review and synthesis of current research on cross-cultural evaluation. *American Journal of Evaluation, 30,* 457–494.

Christie, C. A. (2003). What guides evaluation? A study of how evaluation practice maps onto evaluation theory. In C. A. Christie (Ed.), The practice–theory relationship in evaluation. *New Directions for Evaluation, 97,* 7–36.

Fitzpatrick, J. L., Christie, C. A., & Mark, M. M. (2009). *Evaluation in action: Interviews with expert evaluators*. Thousand Oaks, CA: *Sage.*

Greene, J. C. (2005). Context. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 82–84). Thousand Oaks, CA: Sage.

Julnes, G. & Rog, D. J. (2009). Evaluation methods for producing actionable evidence: Contextual influences on adequacy and appropriateness of method choice. In S. I. Donaldson, C. A. Christie, & M. M. Mark (Eds.), *What counts as credible evidence in applied research and evaluation practice* (pp. 96–131). Thousand Oaks, CA: Sage.

Mertens, D. M. (2008). Stakeholder representation in culturally complex communities: Insights from the transformative paradigm. In N. L. Smith & P. R. Brandon (Eds.), *Fundamental issues in evaluation* (pp. 41–60). New York, NY: Guilford.

Patton, M. Q. (1976). *Utilization-focused evaluation.* Beverly Hills, CA: Sage.

Rog, D. J. (2009, November). *Toward context-sensitive evaluation practice.* Presidential address to the meeting of the American Evaluation Association, Orlando, FL.

Rog, D. J. (2012). When background becomes foreground: Toward context-sensitive evaluation practice. *New Directions for Evaluation, 135,* 25–40.

Stake, R. E. (1974). *Program evaluation, particularly responsive evaluation (Occasional Paper No. 5).* Kalamazoo: Western Michigan University Evaluation Center.

Stake, R. E. (1980). Program evaluation, particularly responsive evaluation. In W. B. Dockrell & D. Hamilton (Eds.), *Rethinking educational research.* London: Hodeder & Stoughton.

Stufflebeam, D. L. (1968). *Evaluation as enlightenment for decision making.* Columbus, Ohio: Ohio State University Evaluation Center.

Stufflebeam, D. L. (1971). The relevance of the CIPP evaluation model for educational accountability. *Journal of Research and Development in Education, 5,* 19–25.

Vo, A. T. (2013). Visualizing context through theory deconstruction: A content analysis of three bodies of evaluation theory literature. *Evaluation and Program Planning, 38*, 44–52.

Weiss, C. H. (1972). *Evaluation research: Methods for assessing program effectiveness.* Englewood Cliffs, NJ: Prentice-Hall.

Weiss, C. H. (1973). Where politics and evaluation research meet. *Evaluation, 1,* 37–45.

Wholey, J. S. (1979). *Evaluation: Promise and performance*. Washington, DC: *Urban Institute.*

Wholey, J. S. (1987). Evaluability assessment: Developing program theory. *New Directions for Evaluation*, *33,* 77–92.

Chapter 4

a. Title
Using Design Research to Research and Develop Evaluation Practice

b. Author
Nick L. Smith
Syracuse University
330 Huntington Hall – IDD&E
School of Education
Syracuse University
Syracuse, NY 13244-2340
(315) 443-3703
nlsmith@syr.edu

c. Description

In contrast to current approaches to research on evaluation (RoE), I present in this chapter an alternative conceptualization of the nature of evaluation practice and then propose the methods of design research as a suitable way to both study and develop evaluation practice under this alternative view. The following points will be elaborated on in the full chapter.

An overview of current RoE illustrates that most of the studies employ such methods as large-scale, self-report surveys of evaluators about their practices, reviews of published evaluation studies, critiques of collections of evaluation reports, and similar approaches. There are benefits to these types of studies: They provide descriptive information about common practices, provoke questions about practice, provide for characterization of the profession, and provide for some level of public accountability. However, these types of studies have several limitations, including inadequate coverage of actual practice, reliance on unsubstantiated evaluator self-report, and perhaps most seriously, the inherent inability of the results to inform improvements in actual practice.

Current research is generally predicated on the view of evaluation practice as a kind of intervention in which researchers ask about common practices (shared approaches), emphasizing ingredients (the tools evaluators use), and looking for general effects (common likely outcomes). Many of those conducting RoE are themselves evaluators, and they seem to adopt the approach that researching practice is the same as evaluating a program: What are its components? How is it implemented? What are its effects? An alternative is to view evaluation practice not as an intervention to be evaluated, but as a performance to be investigated. Under this second view, evaluation practice is a case-based performance in which actions are contextualized, process is emphasized, and specific effects are of concern.

I believe that this view of evaluation practice as a case-based performance rather than a generalized intervention is evidenced in the large number of reflective self-report narratives produced by evaluation practitioners: Practitioners implicitly view evaluation practice as a case-based performance. A review of these reflective narratives begins to uncover what is needed if researchers are to use case-based studies to improve individual evaluator performance.

There is available a group of designs that enable researchers to move from case-based knowledge to generalized performance solutions, known as design research or design-based research. Design research combines scientific investigation with systematic development and implementation of solutions to address practical problems. It comprises an iterative process of design, development, testing, reflection, and redesign, used within stages of protocol development, field testing, user studies, and dissemination. Studies are conducted in applied settings in order to develop usable and effective solutions. The research is also structured to produce theoretical knowledge for use in understanding commonalities among problems of practice and general principles underlying their solution.

In concluding the chapter, I will examine the ways in which design research can be used in the simultaneous research and development of evaluation practice, including likely costs, benefits, and limitations. Design research provides a means of developing better case-based practice while at the same time developing improved generic models of best practice. Instead of only descriptively studying evaluation as it is currently done, or comparing and evaluating existing approaches, researchers should also consider forms of research that enable us to develop new forms of case-based evaluation while at the same time developing deeper knowledge about evaluation practice more generally. Current approaches emphasize descriptive research on existing practice; design research emphasizes developmental research to create future practice. Together, they offer research *on* evaluation practice and research *for* evaluation practice.

Chapter 5

a. Title: Assessing the Non-Academic Impact of Research on Evaluation

b. Authors

J. Bradley Cousins
Centre for Research on Educational and Community Services
University of Ottawa
136 Jean Jacques Lussier
Ottawa, ON, Canada, K1N 6N5
bcousins@uottawa.ca
+1(613) 562-5800 ext 4036

Kate Svensson
Centre for Research on Educational and Community Services
University of Ottawa
136 Jean Jacques Lussier
Ottawa, ON, Canada, K1N 6N5
kate_svensson@rogers.com
+1 (613) 700-1601

Barbara Szijarto
Centre for Research on Educational and Community Services
University of Ottawa
136 Jean Jacques Lussier
Ottawa, ON, Canada, K1N 6N5
barbszijarto@gmail.com
+1 (613)562-5800 ext 4036

Céline Pinsent
Pinrae Research Associates
56 Castillou,
Gatineau, Québec, Canada J8T 5S6
celine@pinrae.ca
+1 (613) 266-5574

Caroline Andrew
Centre on Governance,
Faculty of Social Sciences
University of Ottawa
120 University, Social Sciences Building
Ottawa, Ontario, Canada, K1N 6N5
candrew@uottawa.ca
+1 (613) 562-5800, ext. 2755

John Sylvestre
School of Psychology
University of Ottawa
136 Jean Jacques Lussier, Vanier Hall
Ottawa, ON, Canada, K1N 6N5
jsylvest@uottawa.ca
+1 (613) 562-5800, ext. 4307

c. **Description**

Calls for more and better research on evaluation (RoE) have been sounded for some 20 years now (e.g., Smith, 1993), and the recent appearance of several significant integrations of empirical research suggests that interest in RoE is on the rise (e.g., Brandon & Singh, 2009; Cousins & Chouinard, 2012; Labin, Duffy, Meyers, Wandersman, & Lesesne, 2012). While many empirical studies on evaluation and syntheses of such studies result in implications for evaluation policy and practice, to date there has been little attention devoted to tracking the non-academic impact of RoE. The present chapter will present a strategy for assessing the impact of RoE on evaluation policy and practice.

In our recent review of empirical research on participatory evaluation (Cousins & Chouinard, 2012), we identified 121 studies published in peer-reviewed outlets over the past fifteen years. Regardless of methodological approach, virtually all of the studies generated implications for participatory evaluation practice, typically in the form of lessons learned. While such wisdom can and should be used to inform evaluation theory (Smith, 1993), such lessons are most often directly relevant to practicing evaluators and in some cases evaluation policymakers. Lessons of this sort, however, are not restricted to research on participatory evaluation. Despite the availability of a growing body of practical advice arising from RoE, to what extent has the field capitalized on this advice?

When we talk about evaluation practice, we are most often referring to the practice of doing evaluation, but this is not necessarily the only form of practice. We can also think about the practice of evaluation capacity building and training and what research has to offer in this regard (e.g., Labin et al., 2012). Another dimension of practice is evaluation policy. According to

Trochim (2009), “An evaluation policy is any rule or principle that a group or organization uses to guide its decisions and actions when doing evaluation” (p. 32). Recognizing that policy has profound effects on evaluation practice, Trochim makes a cogent argument that evaluation policy needs to be informed by practice, and, in doing so, he privileges a role for RoE. Research, according to Trochim, can evaluate approaches to policy development and inform taxonomies and audit methodologies. Despite the compelling arguments made, what evidence do we have that RoE has been effective in shaping policy debates and assessing evaluation policy effectiveness? Perhaps more importantly, how would we know?

In applied research in the social sciences, implications for policy and practice are becoming increasingly valued by research funding agencies and granting councils (e.g., Davies, Nutley & Walters, 2005; Huberman, 1994; National Science Foundation, 2013; Wixted & Beaudry, 2012), and such bodies are commonly interested in tracking the non-academic impact of research. Davies et al. (2005) reported on the results of a symposium sponsored by the United Kingdom’s Social and Economic Research Council designed to explore alternative ways of assessing the non-academic impact of social sciences research. They discussed the possibilities of tracking ‘forward’ (assessing the path of research leading to impacts) or ‘backward’ (tracking research after a change has already occurred) to inform research impact assessment. Essential to these approaches is the conscious development of mapping processes through which impact is expected to occur, thereby informing the identification of indicators and sources for data collection.

More recently, this same group developed a compendium of knowledge mobilization models (Davies, Powell, Ward & Smith, 2011). Considered were a variety of “push” (researcher-originating) and “pull” (user-originating) models of knowledge transfer and exchange. In reviewing the choices available, there are compelling reasons to believe that models that favor heightened interaction and exchange between researcher and user communities are likely to be more effective than push or pull models. These models feature such elements as knowledge brokers, change agents, in-person assistance, and other strategies to enhance interaction between knowledge producer and user communities. Given the range of roles and possibilities for exchange and interaction, the evaluation of such approaches is complex and challenging (Ward, House & Hamer, 2009).

Recently, we considered seriously the problem of the evaluation of knowledge mobilization, or “knowledge for action” (KfA) (Cousins et al., 2013). We concluded that effective approaches to the evaluation of KfA need to take into account the interaction between knowledge producer and knowledge user communities and the need to track knowledge transfer, exchange, and co-creation. Given these considerations, a class of evaluation approaches that consider program theory emerged as a reasonable choice. One such approach—contribution analysis (Mayne, 2001, 2011)—seems particularly well-suited to the challenge.

Contribution analysis is an evaluation approach where an explicitly delineated theory-of-change (ToC) guides the identification and collection of evidence in support of telling the causal performance story. In contribution analysis, the emphasis is on assessing *contribution* rather than *attribution*. Similar to logic models, the ToC includes elements and suggested relationships among them, a flow of activities, outputs and results. Of particular interest in the ToC are the mechanisms through which causal effects are expected to take place and, most importantly, the assumptions (and risks) that need to be in place for expected causal relations to hold.

In the present paper we draw on guidance from contribution analysis to develop a ToC for RoE. The ToC includes considerations of RoE knowledge creation, linkage mechanisms, user community, and downstream consequences. The knowledge creation element draws on Mark’s

(2008) conceptual framework for empirical inquiry on evaluation, specifically, in terms of research problem identification (focus on context, methodology, consequences, or professional issues) and inquiry modes (description, classification, causal analysis, values inquiry). The paper describes and justifies the ToC and concludes with guidance for its development and use for the evaluation of the non-academic impact of RoE.

References

Brandon, P. R., & Singh, M. (2009). The strength of the methodological warrants for the findings of research on program evaluation use. *American Journal of Evaluation, 30*, 123–157.

Cousins, J. B., & Chouinard, J. (2012). *Participatory evaluation up close: A review and integration of research-based knowledge*. Charlotte NC: Information Age.

Cousins, J. B., Svensson, K., Szijarto, B., Lemyre, L., Kristjansson, E., & Sylvestre, J. (2013, June). *Evaluating knowledge mobilization: The case for a contribution analysis approach.* Paper presented at the meeting of the Canadian Evaluation Society, Toronto.

Davies, H., Nutley, S., & Walter, I. (2005, May). *Assessing the impact of social science research: conceptual, methodological and practical issues.* Paper presented at the Economics and Social Research Council Symposium on Assessing Non-Academic Impact of Research, London.

Davies, H., Powell, A., Ward, V., & Smith, S. (2011). *Supporting NHS Scotland in developing a new knowledge-to-action model*. St Andrews: Research Unit for Research Utilisation.

Huberman, A. M. (1994). Research utilization: The state of the art. *Knowledge and Policy, 7*, 13–33.

Labin, S., Duffy, J. L., Meyers, D. C., Wandersman, A., & Lesesne, C. A. (2012). A research synthesis of the evaluation capacity building literature. *American Journal of Evaluation, 33*, 307–338.

Mark, M. M. (2008). Building a better evidence base for evaluation theory: Beyond general calls to a framework of types of research on evaluation. In N. L. Smith & P. R. Brandon (Eds.), *Fundamental issues in evaluation* (pp. 111–134). New York: Guilford.

Mayne, J. 2001. Addressing attribution through contribution analysis: using performance measures sensibly. *Canadian Journal of Program Evaluation, 16*, 1–24.

Mayne, J. (2011). Contribution analysis: Addressing cause and effect. In Schwartz, R., Forss, K., & Marra, M., (Eds.), *Evaluating the complex* (pp. 53–96). New Brunswick, NJ: Transaction.

National Science Foundation. (2013, January). *Proposal and award policies and procedures guide*. Washington, DC: Author. Retrieved from http://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/index.jsp .

Smith, N. L. (1993). Improving evaluation theory through the empirical study of evaluation practice. *Evaluation Practice, 14*, 237–242.

Trochim, W. M. K. (2009). Evaluation policy and evaluation practice. In W. M. K. Trochim, M. M. Mark, & L. J. Cooksie (Eds.), *Evaluation policy and evaluation practice. New Directions for Evaluation, 123* (pp. 13–32). San Francisco: Jossey-Bass.

Ward, V., House, A., & Hamer, S. (2009). Developing a framework for transferring knowledge into action: a thematic analysis of the literature. *Journal of Health Services Research and Policy, 5*, 267–279.

Wixted, B., & Beaudry, C. (2012). *"Capturing the impacts" of research.* Discussion paper. Ottawa: Social Sciences and Humanities Research Council of Canada.

Chapter 6

a. Title: Evaluation Capacity Building of Both Program Personnel and Evaluators in a Stake-holder-Based Evaluation

b. Authors

Brian E. Lawton
University of Hawaiʻi at Mānoa
College of Education
Curriculum Research & Development Group
1776 University Avenue, Castle Memorial Hall
Honolulu, HI 96822-2463
(808) 956-4921
blawton@hawaii.edu

Kanesa Duncan Seraphin
University of Hawaiʻi at Mānoa
College of Education
Curriculum Research & Development Group
1776 University Avenue, Castle Memorial Hall
Honolulu, HI 96822-2463
(808) 956-4439
kanesa@hawaii.edu

Lisa M. Vallin
University of Hawaiʻi at Mānoa
College of Education
Curriculum Research & Development Group
1776 University Avenue, Castle Memorial Hall
Honolulu, HI 96822-2463
(808) 956-4920
vallin@hawaii.edu

Joanna K. Philippoff
University of Hawaiʻi at Mānoa
College of Education
Curriculum Research & Development Group
1776 University Avenue, Castle Memorial Hall
Honolulu, HI 96822-2463
(808) 956-4920
philippo@hawaii.edu

George M. Harrison
University of Hawaiʻi at Mānoa
College of Education
Curriculum Research & Development Group

1776 University Avenue, Castle Memorial Hall
Honolulu, HI 96822-2463
(808) 956-4919
georgeha@hawaii.edu

Joanne Cooper
University of Hawai‘i at Mānoa
College of Education (emeritus)
1776 University Avenue, Wist Hall
Honolulu, HI 96822-2463
jcooper@hawaii.edu

c. Description:

Taylor-Powell and Boyd (2008) defined evaluation capacity building (ECB) as

> …strengthening and sustaining an organization's capacity to (1) design, implement, and manage effective evaluation projects; (2) access, build, and use evaluative knowledge and skills; (3) cultivate a spirit of continuous organizational learning, improvement, and accountability; and (4) create awareness and support for program evaluation and self-evaluation as a performance improvement strategy (p. 56).

Evaluators and researchers on evaluation have generally discussed ECB as a front-end process in which the evaluators' specify methodological activities to influence program and organization personnel's evaluative practice directly and purposefully (e.g., Preskill & Boyle, 2008). In this chapter, we will present ECB as more of a by-product of the continued interactions among program stakeholders and evaluators over the course of a project, rather than the result of a planned design. We examine the ECB that occurred among both program personnel and junior evaluators. The ECB occurred during an evaluation of a three-year middle- and high-school inquiry-based science teacher professional development project, entitled *Teaching Science as Inquiry—Aquatic Science* (TSI-A), that was funded by the U. S. Department of Education's Institute of Education Sciences. The personnel who we profile include two marine scientists—Duncan-Seraphin, the project principal investigator, and Philippoff, the project manager, called the *development team* —and two relatively junior evaluators—Harrison, the evaluation manager, and Vallin, the graduate assistant, called the *evaluation team*. The development team's primary responsibilities were to prepare the project's content and professional development activities. The evaluation team's primary responsibilities were to work under the principal investigator in developing, validating, and administering data collection instruments, and in analyzing and reporting results.

The project was largely developmental, and, as a result, the evaluation was constantly adapting to the changing landscape, requiring a high level of interaction between the development team and the evaluation team. As the project progressed, it was determined that some instruments needed to be added to the evaluation design and others needed to be revised. The development team assisted heavily in reviewing revisions to the evaluation plan, selecting instruments, writing and revising instrument items, serving as content-validity experts, and scoring qualitative-item responses. We consider this type and degree of involvement to characterize a *stakeholder-based evaluation* (Cousins & Whitmore, 1998).

The study of ECB as conducted to date used three methods applied sequentially over a period of some weeks: an introductory focus group (facilitated by Lawton) of all the team members, as well as the evaluation principal investigator; a follow-up peer debriefing session (facilitated by Cooper); and the production, sharing, and discussion of each group's written reflections. Peer debriefing is a method of inquiry in which a disinterested peer explores aspects of research or evaluation to help participants understand their disparate points of view and, ultimately, to strengthen the credibility of a study (Lincoln & Guba, 1985). We added the peer debriefing component because we wanted an outsider to help us dig deeper into the results of the first focus group. We gathered the written reflections to allow the participants to ponder and describe their discussion in the first two sessions and to provide the fodder for a rich discussion likely to produce nuanced findings about each individual's ECB.

We began our research by examining the extent to which ECB occurred among the development team. In the focus group discussion, the two team members described how they were trained zoologists who had some prior experience (Duncan-Seraphin) or formal training (Philippoff) in program evaluation but had come to realize in the TSI-A project that evaluation studies are not always implemented as intended (Smith, 2012). This, at times, resulted in frustration and some tension between the development team and the evaluation team. However, as the project progressed and the development team became more involved in the discussion about the evaluation team's plans and in instrument development, data collection, and the scoring of responses, there was a shift in the development team's knowledge, skills, and behavior about evaluation. Thus, the focus group results yielded good evidence that the development team was thinking more evaluatively. The shift in the development teams' knowledge and skills also had a positive effect on the evaluation, creating an environment of accountability and transparency. That is, as the development team's evaluation knowledge and skills increased, they became more inquiring about the evaluation team's decisions, which in turn required the evaluation team to provide thorough descriptions and clear reasons about why certain choices were made.

In the peer-debriefing session group, the peer debriefer helped us examine the evaluation teams' ECB as well. The two evaluation team members were fairly new to conducting evaluations. The discussion began to uncover that, like the development team, the evaluation team was also experiencing some misunderstanding about the process of conducting an evaluation—particularly the number of changes that occurred in the initial evaluation plan—an occurrence not uncommon to evaluations of development projects. This discovery helped provide a more accurate and complete picture of ECB within the context of the TSI-A project.

In the third phase of the study, each member of the two teams prepared written reflections describing (a) his or her experiences in the evaluation, (b) how the interactions among the team members shaped their professional relationships, and (c) advice that they would give themselves or to other novices if they could go back to the beginning of the evaluation. Two excerpts of reflections by a member of the development team and a member of the evaluation team, respectively, are examples highlighting the benefits that the evaluation had on ECB:

> If I had to give advice to myself in the beginning of this project, or to another novice, I would say that having a map doesn't make the adventure less significant. I would tell myself that a map might make the journey more meaningful, but don't beat yourself up if you have to do some extensive exploration before you can create a useful map. Your work with the evaluation team will need to be more extensive than planned in the proposal; next time plan for more interaction!

If I were to go back in time and give my former self advice about this project, I'd say, "It'll work out. You'll still have some outcome data to use for the evaluation. Be aware, though, that this is a development grant and that false starts with instruments or construct definitions will undoubtedly occur." I also would have told my former self to invest in some good ol' fashioned butcher paper and pens to have the development team develop their logic model on their own before we engage in construct-definition discussions. This might have ameliorated some of the stress I felt and made my role more responsive. Still so much to learn. . . .

The final step of the third stage of our research, where the team members and peer debriefer meet to discuss their reflections, has yet to be completed. In the meeting, the participants will focus on how the ECB that occurred within each team affected the ECB of the other team. Together, the findings of the three stages of research will result in a full description of the changes in the team members' attitudes, behaviors, and knowledge that occurred as the project progressed, as well as how the relationships between them evolved over the project. The chapter will strengthen the view about how stakeholder involvement can enhance the overall quality of an evaluation of a project in development and will add to the literature by contributing a discussion of ECB among junior evaluators.

References

Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. In E. Whitmore (Ed.), Understanding and practicing participatory evaluation. *New Directions for Evaluation, 80*, 5–23.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.

Preskill, H. & Boyle, S. (2008). A conceptual model of evaluation capacity building: A multidisciplinary perspective. *American Journal of Evaluation*, 29, 443–459.

Smith, N. L. (Chairperson). (2012, October). *Adaptive study designs and strategies used in evaluating large-scale complex interventions.* A symposium held at the meeting of the American Evaluation Association, Minneapolis.

Taylor-Powell, E. and Boyd, H. H. (2008). Evaluation capacity building in complex organizations. *New Directions for Evaluation*, *120*, 55–69.

Chapter 7

a. Title: Evaluation Education: The State and Role of Evaluation Certificate Programs

b. Author

John M. LaVelle
School of Social Science, Policy & Evaluation
Claremont Graduate University
McManus 240
170 E. 10th Street
Claremont CA 91711
(909) 607 - 9016
john.lavelle@cgu.edu

c. Description

Evaluation practice has grown in leaps and bounds in recent years (Donaldson & Christie, 2006). As this nascent profession has moved towards a more mature profession, the pre-service education and socialization of evaluators has been a recurring topic of interest. Periodic research has been conducted on the universities that provide education and socialization for would-be evaluators with the goal of describing and understanding what is really going on in evaluation education (e.g. Altschuld, Engle, Cullen, Kim & Macce, 1994; Engle, Altschuld, & Kim, 2006; LaVelle & Donaldson, 2010; May, Fleischer, Scheirer, & Cox, 1986). These descriptive studies have added richness to the ongoing dialogue about *what* is taught, *where* it is taught, *how* it is taught, and ultimately *which* academic credentials can be earned by participating in these programs. The aforementioned studies might be critiqued for their methodological limitations, most significantly their overwhelming reliance on self-report surveys (for a full description, see LaVelle & Donaldson, 2010). The studies also have been conceptually limited in that the researchers did not disentangle the course curricula from the degrees sought by the participants. For example, the courses that lead to a master's degree have not been reported separately from the courses that lead to a doctoral degree; while some of the courses and degree requirements might overlap, they are conceptually and practically different experiences. It is important to note as well that master's and doctoral degrees are all important contributions to the patchwork quilt of evaluation practitioners and scholars, but they are not the only educational experiences available.

An understudied—some might argue *neglected*—educational experience is the certificate program. This paper will provide a conceptual overview of the role certificate programs can play in building a profession, as well as explore what criteria might be used to judge if a certificate program should really be called an "evaluation certificate program" and worthy of inclusion in the directories of evaluation education programs. Patterning itself after previous research conducted by LaVelle & Donaldson (2010), this study will apply an online search and analysis procedure and provide a snapshot of the dramatic rise in the number of evaluation certificate programs since the 1980s. More specifically, the researcher utilized the Boolean capabilities of the Google Search Engine to search for the terms "MA or MS or Certificate or PhD: 'evaluation' .edu*" which told the search program to look for the terms "MA" or "MS" or "PhD" or "Certificate" on the same webpage as the term "evaluation" that ended in .edu*. This lent the researcher the ability to look for university-based webpages both in the domestic United States and abroad. When located, the researcher triangulated the curricular data on the webpages with the publicly available information available on the university Registrar pages.

A preliminary analysis of the data suggests the presence of at least 37 evaluation certificate programs operating in the United States during the 2011-12 academic year. These programs were housed in a range of academic disciplines, including education, educational psychology, psychology, and public policy. In addition, several were officially housed in the Graduate School rather than in a specific academic department. When analyzed in their entirety, the programs ranged from 1.4 to 60 graduate credits, but not all would have met the official inclusion criteria of "2 or more courses specifically in evaluation" (Altschuld et al., 1994; Engle et al., 2006; LaVelle & Donaldson, 2010). However, examining the concentration or evaluation-specific courses with the entirety of the set curriculum revealed an interesting result: on a per-credit basis, certificate programs offer a much more concentrated dose of evaluation content (~33%) than either master's degrees (~18%) or doctoral degrees (~13%) (LaVelle, dissertation study, forthcoming). Implications for further study and analysis will be discussed.

References

Altschuld, J. W., Engle, M., Cullen, C., Kim, I, & Macce, B. R. (1994). The 1994 directory of evaluation training programs. In J. W. Altschuld & M. Engle (Eds.) The Preparation of professional evaluators: Issues, perspectives, and programs. *New Directions for Program Evaluation*, *62*, 71–94.

Donaldson, S. I., & Christie, C. A. (2006). Emerging career opportunities in the transdiscipline of evaluation science. In S. I. Donaldson, D. E. Berger, & K. Pezdek (Eds.), *Applied psychology: New frontiers and rewarding careers*. Mahwah, NJ: Erlbaum.

Engle, M., Altschuld, J. W., & Kim, Y. (2006). 2002 Survey of evaluation preparation programs in universities: An update of the 1992 American Evaluation Association-sponsored study. *American Journal of Evaluation*, 27, 353–359.

LaVelle, J. M., & Donaldson, S. I. (2010). University-based evaluation training programs 1980–2008: An empirical review. *American Journal of Evaluation*, 31, 9–23.

May, R. M., Fleischer, M., Scheirer, C. J., & Cox, G. B. (1986). Directory of evaluation training programs. In B. Gross-Davis (Ed.), Teaching evaluation across the discipline, *New Directions for Program Evaluation*, *29*, 71–98.

Chapter 8.

a. Title: The Future of Research on Evaluation

b. Author

Tarek Azzam, PhD
Assistant Professor
Associate Director of the Claremont Evaluation Center
Claremont Graduate University
School of Behavioral & Organizational Sciences
123 East 8th Street
Claremont, CA 91711
(909) 374-5355
tarek.azzam@cgu.edu

c**. Description**

"If we knew what it was we were doing, it would not be called research, would it?"
- Albert Einstein

Attempting to predict the future trends of a field is a challenging task given shifting priorities, technological developments, and the emergence of unexpected ideas that can shift the foundation of what we think we know. One way of dealing with this challenge is to present trends in the field and discuss their future implications, while embedding beliefs and hopes for the future. For example, one of the current trends in the field is a relatively high interest in accessing research on evaluation (RoE) topics (Szanyi, Azzam, & Galen 2012). However, the same study showed there was less interest in actually doing RoE by the evaluation community at large. Given this high level of interest in RoE products (demand) but relatively low interest in producing RoE (supply), it could be hoped that this would lead to increased future funding for RoE, increased community recognition of RoE efforts, and hopefully this would also increase the quality and quantity of RoE-based knowledge and practices within the evaluation community. The direction of these future RoE efforts can span a wide variety of topics, and the following paragraphs will present likely domains and developments that can influence the future supply and demand for RoE. Within each domain, I will highlight current trends, their implications, and discuss the advantages that can emerge from these future pursuits.

RoE and the Theory-to-Practice gap

Understanding how RoE integrates into daily evaluation practice is an area ripe for continued development and one that would offer the largest potential impact on the community. Christie (2003) observed that "the scarcity of empirical research on evaluation can be partially explained, perhaps, by evaluation's inherently pragmatic, rather than empirical aim" (p. 8). Although this statement is true, other pragmatic professional organizations, such as medicine, devote ample resources and effort into empirically testing new ideas, disseminating findings to working professionals, and studying the implications of adoption across the profession and identifying issues and correcting them (Torpy, Lynm & Glass, 2009). This model creates a constant feedback loop to help inform practice and develop theory. At this stage in our professional development, there is a gap between theory and practice, and the loop connecting theory to practice is barely present (Christie, 2003). Efforts to connect theory to practice have been recently conducted, with Miller and Campbell's (2006) study of the implementation of empowerment evaluation in practice being the most notable example. The findings of this suggested that empowerment evaluation was misinterpreted in practice. This has significant implications for the continued development of empowerment evaluation and should lead to more clarity and concrete operationalization of the concepts used in empowerment evaluation. However, this has yet to occur in any meaningful way, as we do not know how the community of empowerment evaluation practitioners has adjusted their practice in response to this study. This broken feedback loop requires urgent RoE attention if we are to increase the impact of theory on practice and the impact of practice on society.

Multidisciplinarity and the Impact of Evaluation on Society

Evaluation has evolved from multiple disciplinary perspectives, including education, psychology, sociology, and philosophy (Shadish, Cook, and Leviton, 1990). The influence of these fields is noticeable in the myriad of approaches and theories that have developed over the years. Further integration of these various disciplinary approaches is needed to push the field forward. More specifically, I anticipate that the only way to understand the impact of evaluation on society (Mark, Henry & Julnes, 2000) is through the systematic use of multidisciplinary methods and approaches and perspectives. Economic approaches could be used to develop models to predict the efficiency associated with policy evaluations, psychologists could examine the mental benefits that are associated with the evaluative process, or sociologists could examine a host of indicators of well-being in evaluation oriented communities. Although some of these examples may appear farfetched, at their core they are meant to push us to think about the value and impact of evaluation from multiple viewpoints. The "impact of evaluation" was identified by the American Evaluation Association members as the most important RoE question for the community (Szanyi et. al., 2012), and I believe that the only way to begin answering this question is through collaborative efforts across disciplines. This can potentially be achieved through the formation and funding of collaborative efforts and teams to work on a particular evaluation issues.

RoE and Technological Developments

Technological developments can also help shape the future or RoE for years to come. Recent developments in new data collection tools such as on-line surveys, social network analysis, qualitative data visualization, and a host of other tools have transformed evaluator practice. However, there are new technological developments that can contribute directly to the creation of new evaluation theories and methods. I think that the most promising development is the use of crowdsourcing for the recruitment of paid participants to complete specific tasks (Azzam & Jacobson, in press). This tool can help the field expand RoE efforts and develop a research agenda that is evaluation-centric. Imagine having access to thousands of people who are nearly representative of the general US population (or perhaps an international sample) and being able to include them throughout the evaluation process. From a methodological development perspective, this access can be used to verify the reliability of different survey measures, have them rate the quality of graphical displays in evaluation reports, translate texts, and even include them as part of the evaluation design. The later example was recently demonstrated in a study that tested the viability of crowdsourcing as a source for creating matched comparison groups (Azzam & Jacobson, in press). This study showed how crowdsourcing could be used to contribute to the design of evaluations when there is a lack of an adequate comparison group. The strengths and limitations of this tool have yet to be determined, and future studies should be conducted to see if this resource can be used in the development of other methodological approaches such as pre-post designs, regression discontinuity, interrupted time series, and so forth. We also have yet to know if this tool can be used for the analysis of qualitative data and information. What would happen if 1,000 people were asked to code an interview transcript? How reliable and valid would their aggregated responses be? Under what conditions would their codes be irrelevant or relevant? These are some of the preliminary questions that need to be investigated if we are to develop new methodological approaches for the evaluation field.

Crowdsourcing can also be used for theory development by helping us test assumptions about the public's values, reactions, and perspectives to the evaluation process. If we are consistently told that our profession is ultimately accountable to the public (Scriven, 1991), then this resource allows us direct access to that public. For example, studies could be conducted to determine the factors that increase/decrease an evaluation's perceived credibility in the public eye. Studies can also be conducted to see the value placed on different outcomes by the general population (e.g., is a 10% increase in a particular social outcome enough?). Such a study can help us determine the significance of a finding by setting a bar that is derived from public perceptions. In essence, we would be able to use the talents of many to conduct a "fleet of studies" (Cronbach et al., 1980) on our own practices and assumptions. I hope that through this approach we can begin to develop an evaluation theory that is more descriptive in nature and that can be applied across multiple contexts.

RoE and Future Generations

Finally, I would like to discuss the role of evaluator training and RoE. I believe that there is a renewed focus on RoE in many of the academic institutions that train evaluators. If this personal observation is true and this trend continues, I hope that this will signify a more optimistic future for the quality and quantity of RoE in our field and its impact on evaluation practice. I am very excited about the prospect and future of RoE, and I think that we can continue to build a wealth of knowledge and help evaluators make decisions that are more influenced by empirical evidence than by trial and error.

References

Azzam, T., & Jacobson, M. (in press). Finding a comparison group: Is online crowdsourcing a viable option? *American Journal of evaluation*.

Christie, C. A. (2003). What guides evaluation? A study of how evaluation practice maps onto evaluation theory. *New Directions for Evaluation*, 97, 7–36.

Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornik, R. C., Phillips, D. C. . , & Weiner, S. S. (1980). *Toward reform of program evaluation*. San Francisco: Jossey-Bass.

Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs*. San Francisco: Jossey-Bass.

Miller, R. L., & Campbell, R. (2006). Taking stock of empowerment evaluation: An empirical review. *American Journal of Evaluation*, *27*, 296–319.

Scriven, M. (1991). Prose and cons about goal-free evaluation. *American Journal of Evaluation*, *12*, 55–62.

Shadish, W. R., Cook, T. D., & Leviton, L. C. (1990). *Foundations of program evaluation: Theories of practice*. Thousand Oaks, CA: Sage.

Szanyi, M., Azzam T. & Galen, M. (2012). Research on evaluation: A needs assessment. *Canadian Journal of Program Evaluation*, *27*, 39–64.

Torpy, J. M., Lynm, C., & Glass, R. M. (2009). Evidence-based medicine. *The Journal of the American Medical Association*, *301*(8), 900.