RESEARCH ON EVALUATION: A NEEDS ASSESSMENT

Michael Szanyi
Tarek Azzam
Matthew Galen
Claremont Graduate University
Claremont, California, USA

Abstract: This survey study attempts to understand the research questions that evaluators were most interested in answering. The findings suggested that there is a great deal of interest in research efforts that (a) explore factors that increase the impact of evaluation, (b) help develop new methodologies, (c) examine the influence of context on evaluations, and (d) help to address ethical dilemmas. Respondents also provided research questions for each topic, revealing a diverse body of concerns and issues. The study also indicated that research on evaluation is viewed as an important endeavour with strong support from the community.

Résumé : Cette étude enquête a tenté de comprendre les questions de recherche auxquelles les évaluateurs ont été les plus intéressés à répondre. Les résultats suggèrent qu'il y a un grand intérêt dans les efforts de recherche qui (a) explorent les facteurs qui augmentent l'impact de l'évaluation, (b) aident à développer de nouvelles méthodologies, (c) examinent l'influence du contexte sur les évaluations, et (d) permettent d'aborder les dilemmes éthiques. Les répondants ont aussi fourni des questions de recherche pour chaque thème, révélant un corps diversifié de préoccupations et d'enjeux. L'étude indique également que la recherche sur l'évaluation est considérée comme un projet important avec un fort soutien de la communauté.

███████ Research on evaluation (ROE) can help connect theory with practice, contribute to the development of alternative approaches to conducting evaluations, further refine work in the field, and allow for a better understanding of the influence of context on evaluations (Christie, 2003). Smith (1993) argues that empirical studies of evaluation practice would aid in the development of evaluation theory, which can eventually guide practitioners in their work. Furthermore, there are many untested ideas that inform evaluation

Corresponding author: Tarek Azzam, Claremont Graduate University, 123 East 8th Street, Claremont, CA, USA 91711; <tarek.azzam@cgu.edu>

practices, and there is a need to examine those ideas and to help our practice become more grounded in empirical work.

In the past, calls for more research on evaluation have gone mostly unheeded (Scriven, 1991; Worthen, 1990). Henry and Mark (2003) also noted that "there is a serious shortage of rigorous, systematic evidence that can guide evaluation or that evaluators can use for self-reflection or for improving their next evaluation" (p. 69). However, after the many calls for more research (e.g., Alkin, 2003; Mark, 2007; Smith, 1993), there are encouraging signs of its development as research on evaluation publications is emerging more frequently in the literature and has addressed many issues such evaluation utilization, ethics, and theory.

The best studied area in ROE is research focusing on the factors that influence evaluation utilization. Cousins and Leithwood (1986) offered one of the first comprehensive syntheses of this area of research and found that the most prevalent factors influencing overall use were evaluation quality and decision characteristics (e.g., type and significance of decision). In an extension of that work, Johnson et al. (2009) used the Cousins and Leithwood (1986) framework to address evaluation use in more recent years and added new categories of stakeholder involvement and evaluator competency to reflect our evolving understanding of evaluation use. These studies, along with others that examine evaluator practice (Fleischer & Christie, 2009; Patton, 2008; Preskill & Caracelli, 1997), have helped our community understand how use is achieved in real-world situations, and how we, as practitioners, act to facilitate its development and success. For example, Preskill and Caracelli (1997) surveyed the Evaluation Use Topical Interest Group of the American Evaluation Association (AEA) and found that 80% of participants agreed that eliciting stakeholder involvement in the evaluation process was part of the responsibility of the evaluator. In a follow-up study, Fleischer and Christie (2009) asked the same question of AEA members in general and found that 98% of participants felt that eliciting stakeholder involvement in the evaluation process was part of the responsibility of the evaluator.

Research has also helped us learn more about the ethical challenges we face as evaluators. A prime example of such research was Morris and Cohn's (1993) survey, which asked a sample of AEA members to identify the ethical problems they encountered most frequently, as well as the most serious ethical challenge they have faced. Using the Standards for Program Evaluation as a guideline, they coded the

responses and found that 65% of participants had faced an ethical problem, that the most frequent ethical issue surrounded the presentation of findings, and that the most serious ethical pressure emerged when stakeholders asked to alter the presentation of findings (Morris & Cohn, 1993). In a similar study of the Canadian Evaluation Society (CES), Buchanan, Babcock, and MacDonald (2011) found that 77% of participants had encountered an ethical dilemma, that it most frequently occurred in the Integrity dimension of the CES Guideline for Ethics, and that the most frequent challenges were conflicting stakeholder expectations and stakeholder pressure to alter presentation of findings. In assessing the extent of wrongdoing, they found that when an issue arose, 87% of participants took the initiative to discuss the problem with those who commissioned the evaluation while only 47% found that action to be helpful (Buchanan et al., 2011). These findings can help the community understand the common challenges that we face in the field and can be used to design training and educational programs as well as policies that can help address these issues.

Research has also informed our understanding of the effects of culture and context on evaluation. Chouinard and Cousins' (2009) synthesis of cross-cultural evaluations suggests that culture permeates every aspect of evaluation, that the construct of evaluation is often viewed as a collaborative activity across cultures, and that even though there is significant variation in cross-cultural evaluation practice, simply adopting a participatory approach does not reduce the need to understand cultural complexity and challenges (e.g., power imbalances and funding requirements). Chouinard and Cousins (2009) used their synthesis to develop a framework that identifies relational, ecological, methodological, organizational, and personal factors that influence the relationship between evaluators and stakeholders. This knowledge can help improve how evaluators interact and address the unique challenges and advantages that come when working across cultures.

However, even with this research and with the continued growth of the evaluation field, there remains much to be understood about the practice of evaluation through the use of systematic inquiry (Christie & Fleischer, 2010). Ramirez, Waldman, and Lasser (1991) have suggested that research topics in the discipline tend to focus on areas that are publishable and where data are readily available, but which often do not accurately address the major issues that face practitioners.

To be useful, evaluation research must focus on topics that are valued by evaluation scholars and practitioners; to refine this focus, specific areas and questions of study need to be identified. Adopting a research inquiry approach can offer direction for the evaluation field and provide additional knowledge for the scholarly development of evaluation theory and practice. Mark (2007), for example, argues that more evidence-based practice would highlight the contributions that evaluation actually makes toward social betterment and would ultimately strengthen those contributions. He further suggests that research on evaluation may be able to answer the fundamental question that plagues many evaluators: "How should one choose from among the multitude of options available for evaluation?" (Mark, 2007, p. 113). By knowing how to select the appropriate methods, practices, and techniques for a given context, evaluators can conduct their work more efficiently and effectively.

With these issues in mind, we sought the input of evaluation scholars and practitioners to identify the areas of greatest interest and potential need in the field of evaluation, with the hope of providing guidance for evaluation researchers and increasing the impact of such research. This approach has been used in other disciplines. In the health arena, for example, Selgrade, Cooper, Germolec, and Heindel (1999) utilized information collected from practitioners and researchers to identify relevant research areas for the study of autoimmune disease and to help set an agenda for future research in the discipline. Similarly, O'Fallon, Wolfle, Brown, Dearry, and Olden (2003) conducted 16 meetings that brought together academic experts, public health practitioners, and community members to discuss the most pressing environmental health issues, to generate research questions, and to ultimately guide research in the area of environmental health. Within the financial management discipline, Ramirez et al. (1991) directly surveyed a sample of managers and chief financial officers about potential research questions and needs; their findings identified areas of confusion and concern (e.g., the impact of regulation on finance) that led to further investigations and insights.

As the field of evaluation continues to grow and develop, we need similar explorations that focus on the research needs and interests of the evaluation community. Understanding these needs can help guide future research efforts and increase the relevance of research on evaluation to our community. With these goals in mind, we conducted a study that attempted to answer the question *What are the*

research topics and questions that evaluators most want explored and answered? To answer this question, evaluators were asked to respond to a set of closed and open-ended questions about the relevance of different research on evaluation topics, their willingness to engage in research on evaluation efforts, and to suggest other research on evaluation issues and topics. The following is a description of the methods we utilized when designing and conducting the study, the main findings, and implications for the future of the field.

METHOD

To answer our research question, we utilized a survey instrument designed to capture respondents' views on the importance of different evaluation topics. The survey format allowed respondents to quantitatively rate the level of importance of various research topics from *not at all important* to *highly important* and to qualitatively suggest areas for research in each topic area. The research areas covered in the survey came from a scan of major topics that appear in the evaluation literature, the Topical Interest Group themes of AEA, and feedback received from pilot respondents. Specifically, an initial draft of the survey was constructed and piloted on a sample of 18 evaluation scholars, practitioners, and students at two universities with established evaluation programs. The feedback received was used for survey modification and improvement.

The final 10 research topics included (a) background issues, (b) ethics, (c) conceptual topics, (d) impact, (e) context, (f) professional development/training, (g) methods, (h) policy issues, (i) culture, and (j) technology, with the addition of (k) an "other" category, used to capture research areas that were missed. Respondents were also asked to rate the importance of research on evaluation in general and to answer a set of background characteristic items. These included the number of years the evaluators had conducted evaluations, their methodological training, the roles they played in their last evaluations (internal/external), and the degree to which evaluation was part of their professional identities.

Sample

The survey was distributed to all members of AEA who were listed in the 2009 directory and who had previously agreed to allow survey research to be sent to them. In mid-October of 2010 a total of 5,839 individuals were contacted via e-mail, and a reminder was sent one

week later. In all, 1,683 respondents completed the survey, for a response rate of 28.8%.

Many of the respondents were very experienced evaluators: 37.3% had more than 11 years of evaluation experience and 48.9% held doctorate degrees. Most respondents indicated that evaluation was a primary part of their professional identities (61.7%) and that they were at an intermediate (46.2%) or advanced level (39.4%) of knowledge and experience in evaluation. Many respondents indicated that their primary training was either in mixed methodology (48.3%) or quantitative approaches (31.1%). Most respondents were currently working on evaluations (78.1%); 29.0% had worked as internal evaluators and 51.0% had worked as external evaluators in their last evaluations. These and other descriptive statistics are presented in Table 1.

Background characteristic data of AEA members was used to compare the study sample with the overall AEA population. AEA provided aggregated data collected through a member survey, which had a 49% response rate (AEA, 2008). A descriptive comparison between the member survey and the study sample was conducted, which did not reveal many major differences between the groups (Table 2). In each case, the study sample followed the same ratios and trends as the AEA sample. For example, both samples contained a ratio of 1 male to 2 females, and both samples showed the same pattern of experience where most respondents had 1 to 5 years or more than 15 years of experience. These are important similarities that could increase the potential generalizability of the study. However, the lack of a more definitive comparison group and the low response rate could reduce the representativeness of the sample and the opinions expressed in the responses.

Limitations

This study contains limitations that should be acknowledged at the outset. The first is the relatively low response rate, which does limit the generalizability of the findings. This limitation also implied that those who did not respond to the survey may have differing opinions on the importance of research on evaluation. Second, this survey had a broad definition of the areas in which participants worked (e.g., education, policy, health evaluation) and was unable to conduct any specific analysis about differences in work settings (e.g., university vs. evaluation firm). Research should continue to look at an evaluator's place of work, as well as the areas and types of evaluations

Table 1
Background Characteristics of Responding Evaluators

	N	%		N	%
Gender			*Professional identity*		
Male	466	27.7%	Evaluation is a primary part	1038	61.7%
Female	1066	63.3%	Evaluation is a secondary part	606	36.0%
Missing	151	9.0%	Evaluation is not a part	30	1.8%
Total	1683	100.0%	Missing	9	0.5%
			Total	1683	100.0%
Years conducting evaluation					
No evaluation experience	9	0.5%	*Knowledge and experience level*		
Less than 1 year	65	3.9%	Relative beginner	236	14.0%
1 to 5 years	446	26.5%	Intermediate level	777	46.2%
6 to 10 years	389	23.1%	Advanced level	663	39.4%
11 to 15 years	204	12.1%	Missing	7	0.4%
More than 15 years	424	25.2%			
Missing	146	8.7%	*General methodological approach*		
			Primarily quantitative	233	13.8%
Highest obtained degree			Primarily qualitative	143	8.5%
Some college	2	0.1%	Mixed methods	1157	68.7%
Certificate	2	0.1%	Missing	150	8.9%
Associate's	2	0.1%			
Bachelor's	67	4.0%	*Methodological training*		
Master's	640	38.0%	Primarily quantitative	524	31.1%
Doctorate	823	48.9%	Primarily qualitative	108	6.4%
Missing	147	8.7%	Mixed methods	813	48.3%
			No formal methods training	89	5.3%
Evaluator's position in last evaluation			Missing	149	8.9%
Lead evaluator	1020	60.6%			
Full-time team member	328	19.5%	*Last evaluation occurred*		
Part-time team member	212	12.6%	Never conducted one	15	0.9%
Other	112	6.7%	Currently working on one	1314	78.1%
Missing	11	0.7%	Less than a month ago	80	4.8%
			1 to 6 months ago	128	7.6%
Evaluator's role in last evaluation			7 to 12 months ago	47	2.8%
Internal	488	29.0%	More than 1 year ago	95	5.6%
External	858	51.0%	Missing	4	0.2%
A mix of both	293	17.4%			
Don't know	25	1.5%	*Primary evaluation area*		
Missing	19	1.1%	Education	477	28.3%
			Health	259	15.4%
			Policy	103	6.1%
			Social/community	236	14.0%
			Organizational	86	5.1%
			Developmental	77	4.6%
			Environmental	58	3.4%
			Undefined	26	1.5%
			Missing	361	21.4%

they work on (or other important demographic information), to get a more nuanced view of the research areas and questions they find most important to address.

Finally, another important limitation of the study is the American focus of the findings. The original intent of this study was to survey a more global sample of evaluators, and this effort is still ongoing. Unfortunately, we were not able to separate out Canadian evaluator responses in the current study. However, the views of some Canadian evaluators are represented in the data because the membership of AEA includes a meaningful number of Canadian evaluators. Future research should attempt to reach a wider and more diversified audience of evaluators to determine if the same pattern of results emerges.

Table 2
Background Characteristics of Responding Evaluators Compared to AEA Members

	N	ROE study survey	AEA 2008 member survey
Gender			
Male	466	27.7%	33%
Female	1066	63.3%	67%
Missing	151	9.0%	N/A
Total	1683	100.0%	100%
Years conducting evaluations			
No evaluation experience	9	0.5%	N/A
Less than 1 year	65	3.9%	N/A
1 to 5 years	446	26.5%	33%
6 to 10 years	389	23.1%	24%
11 to 15 years	204	12.1%	16%
More than 15 years	424	25.2%	27%
Missing	146	8.7%	N/A
Highest obtained degree			
Some college	2	0.1%	N/A
Certificate	2	0.1%	N/A
Associate's	2	0.1%	N/A
Bachelor's	67	4.0%	6%
Master's	640	38.0%	42%
Doctorate	823	48.9%	52%
Missing	147	8.7%	N/A

Analysis Procedures

Several analyses were conducted on the quantitative data, including basic descriptive statistics, factorial and one-way analyses of variance, and appropriate post hoc tests. These analyses were used to detect differences between groups in their quantitative ratings of the importance of various topics within research on evaluation. The researchers also read the open-ended responses (i.e., the suggested research questions) and coded them using an iterative and inductive procedure. Two coders read the responses and independently developed initial codebooks for the open-ended responses. After comparison, the coders created a single codebook to be used. Then, the coders re-read a random sample of the open-ended responses and engaged in the process of refining, combining, disaggregating, eliminating, or adding codes. Multiple rounds of this process occurred until a consensus emerged around the codes and their meanings. The remaining responses were then coded independently. Coding agreement for this sample of responses was then categorized as a dichotomous variable to determine interrater reliability, which is presented in the results section.

RESULTS

Quantitative Analysis

What were the highest rated research areas?

When respondents were asked to rate the importance of each research on evaluation area using a 5-point rating scale (1 = *not important* to 5 = *highly important*), the four top-rated areas were research on the impact of evaluation ($M = 4.53, SD = 0.66$), research on evaluation methods ($M = 4.31, SD = 0.76$), research on evaluation context ($M = 4.19, SD = 0.81$), and research on ethics in evaluation ($M = 3.97, SD = 0.89$). Table 3 presents descriptive information on all 10 areas included in the survey, arranged from the highest to the lowest rated.

Factorial and one-way analyses of variance were conducted on the top four areas to see if any demographic interactions would emerge. The analyses revealed four significant main effects between the research areas and evaluators' demographic characteristics. The first significant main effect occurred in the rating of research on impact for people who professionally identified with the field of evaluation in different ways ($F (2, 1560) = 5.92, p = 0.003$). Post hoc comparisons using the Tukey HSD indicated that those who stated evaluation was a secondary part of their professional identity ($M = 4.46, SD = 0.70$) rated research on impact significantly lower than did those who

Table 3
Descriptive Statistics on Areas of Research on Evaluation (*n* = 1683)

Areas of research on evaluation	M*	SD	Example research questions
Research on impact	4.53	0.66	When and how are evaluations successfully used? Is there a relationship between evaluation and program improvement?
Research on methods	4.31	0.76	How do different stakeholders view the credibility of different methodological approaches? How can different methods be applied in novel ways? Which methods are the most cost-efficient?
Research on context	4.19	0.81	What contextual factors alter the evaluation? How do they affect evaluation methods? How do they affect the evaluated programs?
Research on ethics	3.97	0.89	How do evaluators weigh the needs and concerns of different stakeholder groups? How are decisions about a program's impact made when conflicting evidence emerges?
Research on culture	3.93	0.92	How does an evaluator become culturally competent and effectively use that information? What barriers arise in an evaluation setting that may impede the execution of a culturally competent evaluation?
Research on technology	3.78	0.90	What technology is available that evaluators can use? How might certain software and applications increase the range of and relationship with stakeholders? When does technology hurt or harm the collection of data?
Research on professional development/training	3.74	0.90	What are the current training characteristics of practicing evaluators? What role should professional organizations play in developing local, national, and international evaluation leadership?
Research on policy issues	3.73	0.92	How do evaluation policies vary across organizations? How can evaluators help shape policies of evaluation?
Conceptual research	3.70	0.93	How have past evaluation practices shaped evaluation today? How do ethics inform the choice of evaluation methods? What are the predominant evaluation theories driving practices today?
Background research	3.62	0.97	What sectors do evaluators work in? Which areas do evaluators come from? What type of training do evaluators receive?

*1 = *not important* to 5 = *highly important*

stated evaluation was a primary part of their professional identity (M = 4.57, SD = 0.62).

The second significant main effect occurred in the rating of research on methods for people with different methodological training (F (3, 1513) = 4.85, p = 0.002). Post hoc comparisons using the Tukey HSD indicated that those with no formal methods training (M = 4.03, SD = 0.94) rated research on methods significantly lower than did those with primarily quantitative training (M = 4.36, SD = 0.71), qualitative training (M = 4.38, SD = 0.74), or mixed methods training (M = 4.32, SD = 0.77).

The third significant main effect occurred in the rating of research on context for people with different methodological training (F (3, 1510) = 3.75, p = 0.011). Post hoc comparisons using the Tukey HSD indicated that those with primarily quantitative training (M = 4.10, SD = 0.83) rated research on context significantly lower than did those with mixed method training (M = 4.22, SD = 0.80).

The fourth significant main effect occurred in the rating of research on ethics for people with different methodological training (F (3, 1512) = 3.83, p = 0.018). Post hoc comparisons using the Tukey HSD indicated that those with no formal methods training (M = 3.75, SD = 0.99) rated research on ethics significantly lower than did those with primarily qualitative training (M = 4.09, SD = 0.88) or mixed methods training (M = 4.01, SD = 0.88).

How important is research on evaluation?

When respondents were asked to rate the importance of research on evaluation using a 5-point rating scale (1 = *not important* to 5 = *highly important*), they indicated that research on evaluation is very important (M = 4.18, SD = 0.81), personally conducting research on evaluation is somewhat important (M = 3.35, SD = 1.24), and accessing current research on evaluation is very important (M = 4.19, SD = 0.85) (Table 4).

Table 4
Descriptive Statistics of Stances on Research on Evaluation

Stances on research on evaluation	M*	SD	Most frequent response (n, %)
Research on evaluation	4.18	0.81	Very important (672, 39.9%)
Personally conducting research on evaluation	3.35	1.24	Somewhat important (477, 28.3%)
Accessing current research on evaluation	4.19	0.85	Very important (697, 41.4%)

*1 = *not important* to 5 = *highly important*

These ratings remained consistent across varying demographic characteristics, except in the two cases. First, there was a significant difference in the ratings of the importance of personally conducting research on evaluation for people with different methodological training (F (3, 1517) = 3.38, p = 0.018). Post hoc comparisons using the Tukey HSD indicated that those with no formal methods training (M = 2.93, SD = 1.27) rated personally conducting research on evaluation significantly lower than did those with primarily quantitative training (M = 3.34, SD = 1.23) or mixed methods training (M = 3.37, SD = 1.23). Second, there was a significant difference in respondents' ratings of the importance of accessing current research on evaluation for people who professionally identified with the field of evaluation in different ways (F (2, 1625) = 3.41, p = 0.033). Post hoc comparisons using the Tukey HSD revealed that those who stated evaluation was not a part of their professional identities (M = 3.93, SD = 0.98) rated accessing current research on evaluation lower than did those who stated evaluation was a primary part of their professional identities (M = 4.23, SD = 0.85).

Qualitative Analysis

Our qualitative analysis only focused on the four highest rated research areas. Because the survey contained 10 research areas plus an "other" category, it was impractical to go into sufficient qualitative depth in all of them. Moreover, a focus on the four highest rated areas was in keeping with the overall goal of the study and made reporting the findings more feasible within the space limitations of scholarly journal articles. Our analysis centred on the types of research questions that respondents offered for the evaluation research topics they deemed most important for future inquiry.

Research on impact

Research on the impact of evaluation was rated as the most important area for research. The prompt for this topic described it as research that explores how the evaluation process and findings can impact/influence individuals, groups, and organizations. Within this area of inquiry, respondents were asked to offer research questions that they considered important. Their open-ended responses were coded using 11 categories that emerged with an interrater reliability of 0.84. Table 5 describes each code label and how frequently it was used to code the open-ended responses.

Table 5
Frequency of Codes for Research on the Impact of Evaluation

Code	Description	Frequency	Example question
Activities/ approaches related to evaluation impact	Applied to responses that were interested in the relationship between the impact of evaluations and various evaluator activities or theoretical approaches.	37	What is the relation between stakeholder in-volvement at every stage of evaluation and the impact of evaluation?
Impact of evaluation on programs/ policies	Applied to responses that expressed an interest in understanding the impact that evaluations have at the program and policy level.	20	Does evaluation impact how funding is allocated for new and existing pro-grams?
Contextual variables related to evaluation impact	Applied to responses that referred to the relationship between the impact of evalu-ations and various physical, temporal, political, and stakeholder variables.	15	Under what circumstances is a participatory approach more or less important in order to have the evalua-tion findings used?
Operationalization of evaluation impact	Applied to responses that were interested in how impact can be defined and mea-sured.	14	What is the difference between short-, mid-, and long-term impact?
Activities/ approaches related to evaluation use	Applied to responses that were interested in the relationship between the use of evaluations and various evaluator activities or theoretical approaches.	10	What is the difference between different reporting styles and evaluation use?
Products of evaluation use	Applied to responses that referred to how evaluations are used, or what occurs as a result of this use.	9	What do organizations ac-tually do with evaluations?
Impact of evaluation (no unit specified)	Applied to responses that expressed an interest in understanding the impact that evaluations have, but did not explicitly state a unit of analysis for this impact.	8	Does evaluation improve democratic reasoning or decision-making?
Impact of evaluation on stakeholders	Applied to responses that expressed an interest in understanding the impact that evaluations have at the stakeholder level.	8	How does evaluation help staff, funders, and partici-pants think about program priorities and strategies?
Impact of evaluation on society-at-large	Applied to responses that expressed an interest in understanding the impact that evaluations have at the societal level.	7	What impact do different evaluation models have on social conditions in differ-ent sectors?
Contextual variables related to evaluation use	Applied to responses that referred to the relationship between the use of evalu-ations and various physical, temporal, political, and stakeholder characteristics.	6	What are the facilitators of and barriers to using evaluation findings?
Other	Applied to responses that were either too vague for the coders to decipher or were unrelated to research on impact.	28	
Total		162	

Research on methods

Research on evaluation methods was rated as the second most important area of inquiry. The prompt indicated that research on evaluation methods could investigate different methodological approaches and their connections to evaluation findings, perceived credibility, stakeholder involvement, and evaluation design and budget. Respondents were asked to offer research questions that they considered important within this area. These open-ended responses were coded using nine categories that emerged with an interrater reliability of 0.87. Table 6 describes each code label and how frequently it appeared in the open-ended responses.

Table 6
Frequency of Codes for Research on Evaluation Methods

Code	Description	Frequency	Example question
Perceptions/impact of methodological choices	Applied to responses that were interested in knowing the perceptions or actual impact of the methodological choices they made on the evaluation.	21	How does the use of different evaluation methods influence evaluation outcomes, and which methods are associated with bringing about change?
Choosing appropriate methods	Applied to responses that were not as specific to the context but that were more generally interested in choosing the best methodology in any given situation.	18	What methods are most appropriate for which settings and types of programs?
Choosing appropriate methods for political/cultural contexts	Applied to responses that were particularly interested in the intersection of methodological choice and politics or culture.	15	Are there data collections methods that are non-invasive and culturally sensitive, and are they used in a way that makes data generalizable?
Importance of good methodology	Applied to responses that were simply enthusiastic about research on evaluation methods in general.	13	How important is good methodology?
Using mixed methods	Applied to responses that referenced how to better use and integrate mixed methodology.	9	How do you meaningfully combine and integrate quantitative and qualitative evaluation results?
Balancing methodological rigor and feasibility constraints	Applied to responses that were concerned about the practical issues that can affect the quality of the methodology used.	8	How can evaluators balance the need for rigorous methods and the reality of less than ideal data?
Using novel/cutting-edge methods	Applied to responses that were interested in knowing more about new and innovative methodologies.	7	What new methods are emerging, and how is their credibility perceived?

Descriptions of methods used in different contexts	Applied to responses that wanted to better understand the current state of the field and know which methodologies are being used where and by whom.	6	What methods get used on which problems and why?
Other	Applied to responses that were either too vague for the coders to decipher or were unrelated to research on methods.	17	
Total		114	

Research on context

Research on context was rated as the third most important area for research on evaluation. The prompt indicated that inquiry of this type could explore how contextual factors such as political climate, budget, community characteristics, and other social and environmental factors might affect the evaluation process and its outcomes. Respondents were asked to offer research on evaluation context questions that they considered important. These open-ended responses were coded using eight categories that emerged with an interrater reliability of 0.85. Table 7 describes each code label and how frequently it appeared in the open-ended responses.

Table 7
Frequency of Codes for Research on Evaluation Context

Code	Description	Frequency	Example question
The impact of contextual features on evaluation	Applied to responses that mentioned a diverse set of contextual features that could have some impact on evaluation and were generally concerned with the intersection of impact and context.	19	To what degree does the imperative to raise money influence the way in which evaluations are conducted?
Operationalization of evaluation context	Applied to responses that referred to defining and constructing meaning around context.	15	Is there academic versus practical context and how would we deal with potential differences?
Practices related to adapting to contextual features	Applied to responses interested in specific activities one should engage in to address contextual factors in an evaluation.	14	What specific evaluation approaches and techniques should be used if the context is complex and adaptive?
The impact of discipline-level context on evaluation	Applied to responses that considered the larger disciplinary field of evaluation as the context and wanted to know what effect the field would have on conducting evaluations.	13	How do evaluation publication practices affect evaluation methods?
Importance of considering context	Applied to responses that were interested in research on evaluation context in general	12	Is there an evaluation that is not influenced by context?

Code	Description	Frequency	Example question
The impact of political context on evaluation	Applied to responses that considered political features of a situation as the context and wanted to know what effect those features would have on conducting evaluations.	11	How do changes in administration affect evaluation planning, measures, and outcomes?
Impacts/practices related to cultural contexts	Applied to responses that were particularly interested in how the cultural context is addressed and affects evaluation practice.	10	How and in what ways does culture influence, or should influence, methodological choices, data collection, and evaluation communication?
Other	Applied to responses that were either too vague for the coders to decipher or were unrelated to research on context.	16	
Total		110	

Research on ethics

The issue of ethics was rated as the fourth most important area for research on evaluation. The prompt indicated that research on ethics in evaluation could investigate evaluator responses to ethical dilemmas. Respondents were asked to suggest questions in this area that were of relevance to them. These open-ended responses were coded using eight categories that emerged with an interrater reliability of 0.88. Table 8 describes each code and how frequently it appeared in the open-ended responses.

Table 8
Frequency of Codes for Research on Ethics in Evaluation

Code	Description	Frequency	Example question
Current/best practices related to evaluation ethics	Applied to responses that were interested in specific activities one does or should engage in to deal with broad ethical issues.	17	What can evaluators do to ensure they are implementing ethical evaluation?
Current/best practices related to political/cultural issues	Applied to responses that were interested in specific activities one does or should engage in to deal with political and/or cultural issues related to ethics.	13	Can evaluation ethics differ from one region or culture to another and are evaluation ethics absolute or relative?
Importance of ethics in evaluation	Applied to responses that were interested in research on evaluation ethics in general	11	Without ethics, can you even call it evaluation?

Professional development and training	Applied to responses that mentioned the continued learning and professional development of evaluators.	9	What ethical values should evaluators be trained to uphold, what are some of the most critical dilemmas faced, and how should they be resolved?
Ethical frameworks/definitions	Applied to responses that referred to exploring what ethics means and using/developing ethical frameworks in evaluation.	9	What classical and non-mainstream ethical frameworks can we use with confidence?
Issues related to board review	Applied to responses that were concerned with the evaluation review process, including confidentiality and informed consent.	6	How have internal review boards harmed or helped evaluation and evaluation clients?
Issues related to the ethics of inclusion	Applied to responses that were interested in the ethical challenges and opportunities of including diverse people in the practice of evaluation.	6	How do members of the community participate in discussions on ethical issues?
Other	Applied to responses that were either too vague for the coders to decipher or were unrelated to research on ethics.	8	
Total		79	

Table 9 provides a summary of the main findings from the results and analysis sections.

Table 9
Summary of Main Findings

What were the highest rated research areas? (Top four, starting with highest)

Research topic	Description of findings
Research on impact	Respondents were interested in studies that examine the potential impacts of various stakeholder engagement approaches, evaluator roles, methods, or evaluation theories.
Research on methods	Respondents were concerned about what constitutes credible evidence, for whom, and under what circumstances, and the assumptions that diverse evaluation methods make about reality, human nature, and social organizations.
Research on context	Respondents were interested in knowing more about how accommodation to contextual factors—such as different stakeholder groups, accessibility to data, evaluation timing and funding—might affect the use of evaluation results.
Research on ethics	Respondents were interested in knowing how evaluators currently deal with a variety of ethical issues related to evaluation use, reporting, communication, internal and external roles, and how evaluator actions compare to suggested ethical practices.

How important is research on evaluation?

	Results indicated that research on evaluation was considered very important to the evaluation community, and that accessing current research on evaluation was also very important. However, there were mixed levels of interest in personally conducting research on evaluation studies.

DISCUSSION

This study set out to shed light on some of the most pressing questions in research on evaluation in the hopes of providing guidance for evaluation researchers and increasing the potential impact of research on the field. The findings indicate that research on evaluation is viewed as an important endeavour and has strong support. Individuals' interest levels in this type of research and in specific topics within the discipline are influenced by evaluators' levels of methodological training (i.e., no training vs. training) rather than the type of methodological training (i.e., qualitative, quantitative, or mixed), or individuals' identification with the evaluation field (i.e., primary vs. secondary). The findings also suggest that while there is a great deal of interest in the products of research on evaluation, there is mixed interest among respondents in actually conducting research on evaluation. Certain individuals are very interested in conducting research on evaluation and believe it is critically important, while others believe in its importance but are not necessarily interested in designing and implementing the studies themselves. This contrast highlights the pattern that emerged when calls for more research on evaluation were made and produced research in some areas but not in others.

Overall, evaluators rated the impact of evaluation, evaluation methods, evaluation context, and ethics in evaluation as the four most important topics to be explored in research on evaluation. Research on the impact of evaluation represented the highest rated research topic and was particularly important for those who professionally identified with the field of evaluation. Research on impact is closely tied to Mark's (2007) "evaluation consequences" research dimension, which he defines as the study of evaluation consequences and the systematic examination of "changes that do (or do not) occur as a result of evaluation" (p. 117). Questions from the open-ended responses further suggest that there is a great deal of interest in knowing about the impact of evaluation.

Evaluators were especially interested in learning more about the impact of using different evaluation activities or approaches. Evaluators wanted to know which evaluation theory leads to larger impacts, how stakeholder involvement helps or hinders evaluation impact, and what evaluation practices are most influential in making change. Beyond these practice questions, evaluators also wanted to know how impact can be measured or captured in programs and policies, its re-

lation to cost, and what other contextual variables might ultimately affect an evaluation's impact. These questions point to an existing need to further understand what effects evaluation actually has. Although this interest is not necessarily new (Weiss, 1977; Welch & Sternhagen, 1991), there have been good strides toward answering many of these impact questions in the studies examining evaluation utilization (Cousins & Leithwood, 1986; Johnson et al., 2009). Nonetheless, it does suggest that, as the field continues to grow, the importance of demonstrating the impact evaluations can have on programs, policy, and society at large is also growing.

Research on evaluation methods represented the second highest rated research topic. This area of research was viewed as especially important by evaluators who had more evaluation experience and methodological training. Research on evaluation methodology could offer novel ways to examine complex situations or offer better insights into how different methodologies are implemented, viewed, and used.

In their open-ended responses, evaluators indicated that they wanted to know more about the consequences of different methodological choices and how to select appropriate methods for varying political and cultural contexts. These concerns were reflected in the proposed research questions where respondents expressed a desire to know more about the impact of mixing qualitative and quantitative methods, stakeholder perceptions of the credibility of different methodological approaches, and the effectiveness of non-invasive techniques for collecting evidence. These questions can be viewed as critical methodological questions facing the evaluation field because they go beyond the technical merits of different methodological approaches and focus more on the credibility, adaptability, and relevance of methods to different contexts. Finding answers to these questions could, in the long term, help develop new approaches that could be adapted to changing contexts.

Research on the evaluation context focuses on exploring and examining the situations in which evaluation happens (and, usually, its effects on evaluation). This area of research was the third highest rated research topic. Many of the questions that emerged were concerned with the effects that context has on the design and conduct of evaluations. Respondents were particularly interested in knowing more about the influence of the political climate on the design and use of evaluations, the effects that stakeholder groups have on

what actually gets measured, and the relationship between culture and the ability of evaluators to conduct their work. These questions reflect a heightened level of interest in the ways that context may affect evaluation practice and outcomes.

Importantly, open-ended responses related to evaluation context revealed confusion surrounding the term "context." In many responses, evaluators wanted to know how to operationalize the term and suggested that any research should begin with a clear articulation of what contextual elements would be examined. For example, will a researcher focus on contextual factors such as organizational size, cultural diversity, public or private programs, resource availability, political situations, or a host of other factors? These questions point to a need for more conversations and investigations around the definition of context and its potential categories. Classification efforts can help researchers and practitioners focus their scholarship and practice when addressing this issue.

Research on ethics in evaluation was the fourth highest rated research topic. This area of research was viewed as especially important by evaluators who had more evaluation experience and methodological training. Ethics is a foundational issue in evaluation that has historically generated a great deal of interest and debate within the field (Morris, 2011), and this was reflected in the research questions that emerged from the survey. Many of the questions broadly focused on three issues: What are the common ethical dilemmas facing evaluators in the field? How do they actually deal with these ethical dilemmas? Are their responses considered ethical?

Such studies have been conducted in the past (Morris, 2007; Penuel, Sussex, Korbak, & Hoadley, 2006; Walker, Hoggart, & Hamilton, 2008) and have produced valuable lessons for the field. In addition, books and book chapters containing insightful case studies and suggestions for best practices also add to the resources available to evaluators (Morris, 2008; Newman & Brown, 1996; Simons, 2006). But, as Morris (2011) observed, relatively few studies have examined these questions. Moreover, one of the missing pieces is consistent research on this topic. The regular tracking of common ethical dilemmas facing evaluators can act as a bellwether for emerging challenges in our field. The consistent production of research on ethics would help to generate healthy debates around existing and new ethical issues, provide continued presence in the literature, and ultimately help improve evaluator practice.

These four research topics can be viewed as the areas where the field of evaluation could use additional development. The main goal of research on impact is to gain a better understanding of the value we offer stakeholders and society; research on methods shows the concerns we have about the reliability, validity, and credibility of the designs we implement; research on context highlights the challenges we face when conducting evaluation in real-world settings; and research on ethics exposes the dilemmas we encounter as we make evaluation choices. The questions emerging from these areas, among others, point to our desire for a more nuanced view of the world and how we interact with it; the answers to these questions are often missing in methodological and evaluation textbooks. This is why additional research on evaluation would be useful to move the field forward.

CONCLUSION

Calls for more research on evaluation will most likely continue. In a series of presentations and publications on "Advances in Evaluating Evaluation Theory," Smith (2010), Miller (2010), and Kirkhart (2010) all discuss the state of current research on evaluation in relation to evaluation theory. They point out that even though the field has advanced in its thinking, there is still much work to be done to empirically test the ideas and claims that emerge from different evaluation theories, and to increase the connection between evaluation practice and research on evaluation. This study is intended to complement and further inform those calls by offering a better understanding of the more general areas and topics that are of concern to evaluators.

The multifaceted research agenda put forth in this article has been compiled through input from a diverse group of evaluators. It informs the field of the issues that evaluation scholars and practitioners find to be the most in need of research. Hopefully, evaluation experts can use this information to identify questions and topics that they can integrate into their work as they move the field of evaluation forward. Knowing what evaluators value, and subsequently acting on that knowledge, will ideally increase support for research on evaluation and enthusiasm for actually doing the research, and ultimately improve the quality of our evaluation practice.

To continue this trend, incentives should become available to increase the motivation to conduct this kind of work outside of a university setting. Although it would seem that practitioners are not studying

their own practice, there is a growing pool of grey literature, which is often empirical, where evaluators are reflecting on their work. Much of this can be found in both the AEA and CES conference sessions and websites. One such example of this is the new section in the *Canadian Journal of Program Evaluation* titled "Addressing Challenges in Evaluation Practice," which is devoted to presenting real-life cases by evaluation practitioners. By offering incentives, professional societies and evaluation journals may encourage these practitioners to contribute to the more traditional knowledge base.

As CES continues this work and continues to operate so closely with AEA, it is important to share the views and concerns of their memberships. Specifically, as evaluation efforts continue to cross borders, and as evaluation becomes more of a global effort, it is important that the field as a whole move forward with a cohesive set of fundamental evaluation principles. Through the combined research efforts of evaluators in Canada, the United States, and abroad, the field will hopefully be able to uncover empirically based evaluation knowledge that all evaluators can use in their practice. The authors hope that this initial article will open additional opportunities to access a larger evaluation community so that this effort can continue in the future.

REFERENCES

Alkin, M. C. (2003). Evaluation theory and practice: Insights and new directions. In C. Christie (Ed.), *The practice-theory relationship in evaluation* (pp. 81–89). San Francisco, CA: Jossey-Bass.

American Evaluation Association. (2008). *American Evaluation Association internal scan report to the membership*. Retrieved from http://www.eval.org

Buchanan, H., Babcock, K., & MacDonald, W. (2011). *Ethical challenges in evaluation: 2010 Canadian survey*. Ottawa, ON: Canadian Evaluation Society.

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

Christie, T. A., & Fleischer, D. N. (2010). Insight into evaluation practice: A content analysis of methods used in evaluation studies published in North American evaluation focused journals. *American Journal of Evaluation, 31,* 326–346.

Cousins, J. B., & Leithwood, K. A. (1986). Current empirical research on evaluation utilization. *Review of Educational Research, 56*, 331–364.

Fleischer, D. N., & Christie, C. A. (2009). Evaluation use: Results from a survey of U.S. American Evaluation Association members. *American Journal of Evaluation, 30*, 158–175.

Henry, G. T., & Mark, M. M. (2003). Toward an agenda for research on evaluation. *New Directions for Program Evaluation, 97,* 69–80.

Johnson, K., Greenseid, L. O., Toal, S. A., King, J. A., Lawrenz, F., & Volkov, B. (2009). Research on evaluation use: A review of the empirical literature from 1986 to 2005. *American Journal of Evaluation, 30*, 377–410.

Kirkhart, K. E. (2010). Eyes on the prize: Multicultural validity and evaluation theory. *American Journal of Evaluation, 31,* 400–413.

Mark, M. M. (2007). Building a better evidence base for evaluation theory. In N. L. Smith & P. R. Brandon (Eds.), *Fundamental issues in evaluation* (pp. 111–134). New York, NY: Guilford Press.

Miller, R. L. (2010). Developing standards for empirical examinations of evaluation theory. *American Journal of Evaluation, 31,* 390–399.

Morris, M. (2007). Foundation officers, evaluation, and ethical problems: A pilot investigation. *Evaluation and Program Planning*, *30*, 410–441.

Morris, M. (2008). *Evaluation ethics for best practice: Cases and commentaries*. New York, NY: Guilford Press.

Morris, M. (2011). The good, the bad, and the evaluator: 25 years of AJE ethics. *American Journal of Evaluation 32*, 134–151.

Morris, M., & Cohn, R. (1993). Program evaluators and ethical challenges: A national survey. *Evaluation Review, 17,* 621–642.

Newman, D. L., & Brown, R. D. (1996). *Applied ethics for program evaluation*. Thousand Oaks, CA: Sage.

O'Fallon, L. R., Wolfle, G. M., Brown, D., Dearry, A., & Olden, K. (2003). Strategies for setting a national research agenda that is responsive to community needs. *Environmental Health Perspectives, 111,* 1855–1860.

Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: Sage.

Penuel, W. R., Sussex, W., Korbak, C., & Hoadley, C. (2006). Investigating the potential of network analysis in educational evaluation. *American Journal of Evaluation, 27*, 437–451.

Preskill, H., & Caracelli, V. (1997). Current and developing conceptions of use: Evaluation Use TIG survey results. *Evaluation Practice, 18,* 209–225.

Ramirez, G. G., Waldman, D. A., & Lasser, D. J. (1991). Research needs in corporate finance: Perspectives from financial managers. *Financial Management, 20,* 17–29.

Scriven, M. S. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: Sage.

Selgrade, M. K., Cooper, G. S., Germolec, D. R., & Heindel, J. J. (1999). Linking environmental agents and autoimmune disease: An agenda for future research. *Environmental Health Perspectives, 107,* 811–813.

Simons, H. (2006). Ethics in evaluation. In I. F. Shaw, J. C. Greene, & M. M. Mark (Eds.), *The SAGE handbook of evaluation* (pp. 243–265). Thousand Oaks, CA: Sage.

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

Smith, N. L. (2010). Characterizing the evaluand in evaluating theory. *American Journal of Evaluation, 31*, 383–389.

Walker, R., Hoggart, L., & Hamilton, G. (2008). Random assignment and informed consent: A case study of multiple perspectives. *American Journal of Evaluation, 29*, 156–174.

Weiss, C. H. (Ed.). (1977). *Using social research in public policy making*. Lexington, MA: Lexington Books.

Welch, W. W., & Sternhagen, F. (1991). Unintended effects of program evaluation. *Evaluation Practice, 12,* 121–129.

Worthen, B. R. (1990). Program evaluation. In H. J. Walberg & G. D. Haertel (Eds.), *The international encyclopedia of education evaluation* (pp. 42–47). Oxford, UK: Pergamon Press.

Michael Szanyi received his B.A. in psychology from Pitzer College and his M.A. in applied social psychology and evaluation from Claremont Graduate University (CGU). He is currently pursuing a doctoral degree in the Evaluation and Applied Research Methods program at CGU. Michael has worked on evaluations ranging from youth social programs to college access initiatives. Recently, he has engaged with after-school dance education programs to examine program logic, implementation, and student outcomes particular to those art-based communities. Michael is interested in research on advocacy and ethics in evaluation, art-based evaluation techniques, and the role that research on evaluation has in advancing theory and practice.

Tarek Azzam, Ph.D., is Assistant Professor at the School of Behavioral and Organizational Sciences, Claremont Graduate University. Dr. Azzam's research focuses on developing new methods suited for real-world evaluations. These methods attempt to address some of the logistical, political, and technical challenges that evaluators commonly face in practice. His work aims to improve the rigour and credibility of evaluations and increase its potential impact on programs and policies. Dr. Azzam has also been involved in multiple projects including the evaluation of student retention programs at the K-12 and university level, Science Technology Engineering Math education programs, pregnancy prevention programs, children's health programs, and international development efforts for the Rockefeller and Packard Foundations.

Matt Galen received his B.A. in psychology from Drew University and his M.A. in positive organizational psychology and evaluation from Claremont Graduate University (CGU). He is currently pursuing a doctoral degree in the Evaluation and Applied Research

Methods program at CGU. Matt works to assist organizations in improving their evaluation effectiveness and efficiency, with the goals of reducing costs and improving evaluation quality. Recently, Matt has worked to improve the David and Lucile Packard Foundation's ability to analyze relevant grantee outcomes and has served as a data efficiency consultant for the Lucile Packard Foundation for Children's Health. Matt currently works for the Claremont Evaluation Center (CEC), where he serves as a project manager for a diverse portfolio of evaluations and collaborates with partners from the Rockefeller Foundation and UNICEF to coordinate international evaluation e-learning programs.