

---

# Research on Evaluation Outside Journal Publications: An Analysis of Proposals Accepted for the 2019 American Evaluation Association Annual Conference

*Journal of MultiDisciplinary Evaluation*  
Volume 21, Issue 51, 2025

**JMDE**  
Journal of MultiDisciplinary Evaluation

ISSN 1556-8180  
<http://www.jmde.com>

Dana J. Linnell  
*University of Wisconsin-Stout*

Michael Harnar  
*Western Michigan University*

Esther C. Nolton  
*Everstead Strategies*

Seema Mahato  
*Washington University in St. Louis*

Travis R. Moore  
*Tufts University*

**Background:** Research on evaluation (RoE) may be shared outside of academic journal articles.

**Purpose:** Explore the extent to which RoE is being shared at American Evaluation Association (AEA) conferences.

**Setting:** Not applicable.

**Intervention:** Not applicable.

**Research Design:** Systematic review.

**Data Collection and Analysis:** Conference proposals from the 2019 AEA conference were coded for whether they were RoE.

**Findings:** 15% of 2019 AEA conference proposals were coded as RoE.

**Keywords:** *research on evaluation; conference proposals; American Evaluation Association*

---

## Research on Evaluation Outside Journal Publications: An Analysis of Proposals Accepted for the 2019 American Evaluation Association Annual Conference

Research on evaluation (RoE) is “Any purposeful, systematic, empirical inquiry intended to test existing knowledge, contribute to existing knowledge, or generate new knowledge related to some aspect of evaluation processes or products, or evaluation theories, methods, or practices.” (Coryn et al., 2016, p. 161). The purpose of this study is to examine the extent to which RoE is being presented at evaluation conferences by examining proposals accepted for presentation at the 2019 American Evaluation Association (AEA) conference.

### *Research on Evaluation*

In a survey of AEA members (Coryn et al., 2016), of those who responded, 96% stated that RoE is important, and 97% acknowledged its influence on their thinking. Additionally, 41% of respondents stated that RoE improves, informs, and guides the practice of evaluation. This suggests that evaluators find RoE important and valuable for the field.

To examine the prevalence of RoE, Coryn et al. (2017) conducted a systematic review of evaluation journals published between 2005 and 2014. They identified 7.6% of articles as RoE, with variations in the presence of articles across different journals ranging from 1% in *Educational Evaluation and Policy Analysis* to 43% in *African Evaluation Journal*.

Coryn et al. (2017) further coded the RoE articles based on Henry and Mark’s (2003) and Mark’s (2008) frameworks. Henry and Mark (2003) described an agenda for RoE, including six types of research: research on evaluation outcomes, comparative research on evaluation practice, metaevaluation, analog studies, practice component studies, and evaluation of technical assistance and training. Coryn et al. (2017) found that 41% of RoE studies focused on comparative research on evaluation practice, 31% had a focus categorized as “other,” 14% examined RoE outcomes, 12% were practice component studies, and less than 10% were the remaining three categories of RoE (Henry & Mark, 2003).

Later, Mark (2008) elaborated on the earlier Henry and Mark (2003) agenda for RoE by providing a framework for RoE, describing various subjects of inquiry (i.e., what is being studied in the

RoE) and modes of inquiry (i.e., how it is being studied) based on his experience with RoE. He argued that these could be used both to understand the current landscape of RoE and also to identify gaps in RoE, particularly by examining the cross-section of various subjects of inquiry and modes of inquiry (e.g., evaluation activities across evaluation contexts). Coryn et al. (2017) found that evaluation practice was the most frequent subject of inquiry (51%), followed by evaluation consequences (27%), evaluation contexts (23%), professional issues (20%), and domain-specific issues (9%; Mark, 2008). The vast majority of articles used description as a mode of inquiry (83%), and a minority used tool development or validation (7%), causal analysis (6%), classification (5%), or values inquiry (4%; Mark, 2008).

### *Disseminating RoE*

Coryn et al. (2017) concluded their article by highlighting that RoE tends to be predominantly an academic endeavor, despite only a third of AEA members working in academia (AEA, 2018a). This is likely because most academics are expected to conduct and publish research and are incentivized to do so. Practitioners, on the other hand, often do not have research as part of their job descriptions, nor are they typically incentivized to conduct and publish research. Coryn et al. (2017) encouraged evaluators outside academia to contribute to RoE; however, in other research, practitioners mentioned practical challenges such as time constraints, limited resources, and unfamiliarity with academic writing styles as obstacles that hindered their participation (Mata et al., 2010).

Although research tends to be published primarily as papers in academic journals, there are alternative opportunities to share RoE, including through grey literature and conference presentations. Conferences offer valuable professional development experiences and foster networking, skill-building, mentoring, peer review, and collaboration (Mata et al., 2010). They also offer a platform to disseminate research and lessons learned directly to an audience, often with less effort than journal article submissions. Though one limitation of conferences is the time and money associated with attendance, conference presentations may yield greater rewards for less effort than other avenues for sharing research, especially when research is not a primary job responsibility. Consequently, practitioners may primarily engage in and share their RoE through conference participation.

The annual American Evaluation Association (AEA) conference offers evaluators an opportunity to disseminate their work. It is attended each year by more than 2,000 professional and academic evaluators from around the world. The conference changes locations every year. The theme of each year's conference is determined by the AEA president. There are a variety of types of AEA conference proposal submissions,<sup>1</sup> including individual papers (which are organized by the topical interest group (TIG)<sup>2</sup> leaders to be presented in multi-paper sessions with other individual papers by TIG leaders), Birds of a Feather gatherings, demonstrations, skill-building workshops, expert lectures, Ignite presentations, panels, roundtables, think tanks, and multi-paper sessions (intact sessions proposed by presenters). A typical journal article is anywhere from 2,000 to 10,000 words, whereas an AEA conference session proposal only requires a title, a 150-word abstract, and a 500-word relevance statement, along with information about the audience level and which of AEA's guiding principles (2018b) and evaluator competencies (2018c) the proposal most aligns with. Some proposal types require additional information; multi-paper sessions and panels also have titles and 150-word abstracts for each paper or panel presentation within a session.

Submitters are asked to select the TIG that most closely aligns with their proposal. In 2019, AEA had 61 TIGs covering a wide range of topics relevant to evaluation. Once submitted, proposals enter a review process coordinated by the respective TIG leaders. At the time of our research study, this review process varied across TIGs. In the typical case, TIG leaders solicit reviewers from AEA membership to evaluate proposals based on a set of standardized criteria established by AEA. Reviewers typically provide scores for each proposal, resulting in a cumulative average score. Most proposals are reviewed by approximately three reviewers, and proposals with the highest average scores are selected for presentation. However, the review process is not uniform across all TIGs, and additional variations may occur. For instance, some TIGs may use slightly different criteria, prioritize certain types of proposals (e.g., those from emerging evaluators or underrepresented voices), or apply qualitative assessments alongside scoring. Notably, we are not aware of any TIG using RoE as a criterion for reviewing proposals. In some cases, TIG leaders may also directly influence final selections based on

thematic fit, session balance, or broader representation goals. These inconsistencies across TIGs highlight the flexible and decentralized nature of the review and selection process.

AEA is just one of many voluntary organizations for professional evaluation (VOPEs) that offer conference presentation opportunities. For instance, the Canadian Evaluation Society, European Evaluation Society, and African Evaluation Association each hold an annual international conference. There are also a plethora of regional and local conferences held, including the Eastern Evaluation Research Society and the Arizona Evaluation Network.

### *Purpose of This Study*

Despite many conference opportunities to share research, no RoE studies have focused only on the prevalence of RoE presented at conferences. Building on Coryn et al.'s (2017) findings, we focused in this study on the extent to which RoE is being presented at the annual AEA conference. Furthermore, we examined proposals that we determined to include RoE against Mark's (2008) framework examining subjects of inquiry and modes of inquiry. Lastly, we added a third dimension—methods of inquiry—to Mark's (2008) framework to examine not only the broad mode of inquiry (e.g., descriptive, causal, categorization) but also the methods used in the mode of inquiry (e.g., survey, interview, document analysis).

### **Methods**

In this study, we built upon the findings of Coryn et al. (2017) to investigate the extent to which RoE is present in accepted proposals for the annual AEA conference. We began this project in 2020, and because we were interested in current representation of RoE, we analyzed data from the most recent annual AEA conference, which was held in 2019 in Minneapolis, MN, and led by that year's president, Tessie Catsambas, with the theme "Paths to the Future of Evaluation: Contribution, Leadership, Renewal." We screened the proposals that were accepted for this conference in order to identify those containing RoE content.

<sup>1</sup> <https://www.evaluationconference.org/Attendees/Presentation-Types>

<sup>2</sup> <https://www.eval.org/Community/Topical-Interest-Groups>

## Data

Data in this study consisted of all proposals accepted for presentation at the 2019 AEA annual conference ( $n = 1,414$  intact sessions and individual papers). Data were provided by the American Evaluation Association for analysis and included session titles, abstracts, relevance statements, session types (e.g., panel, multi-paper), related topical interest groups (TIGs), paper titles and abstracts (if applicable), audience level designations, and author information (i.e., names, organizations). We did not examine audience level or author information in the present study.

*Inclusion Criteria.* Of the 1,414 proposals accepted for presentation at the 2019 AEA conference, we only coded session types that we thought were likely to have RoE content according to the definition of RoE we used for this study. We limited our analysis to individual papers ( $n = 433$ ) in their respective TIG multi-paper sessions ( $n = 125$ ), posters ( $n = 208$ ), Ignite presentations ( $n = 45$ ), and intact multi-paper sessions ( $n = 46$ ). We did not examine panels ( $n = 193$ ), roundtables ( $n = 156$ ), demonstrations ( $n = 114$ ), skill-building workshops ( $n = 64$ ), Birds of a Feather gatherings ( $n = 55$ ), think tanks ( $n = 52$ ), expert lectures ( $n = 44$ ), or keynotes ( $n = 4$ ), because preliminary examination of proposals and previous experience with attending AEA conferences showed that these session types tend to be more interactive discussions or reflective presentations, in which presenters are less likely to be sharing research studies. We also excluded one poster because the abstract was written in another language. Figure 1

illustrates these criteria and summarizes our coding procedures.

## Coding Procedures

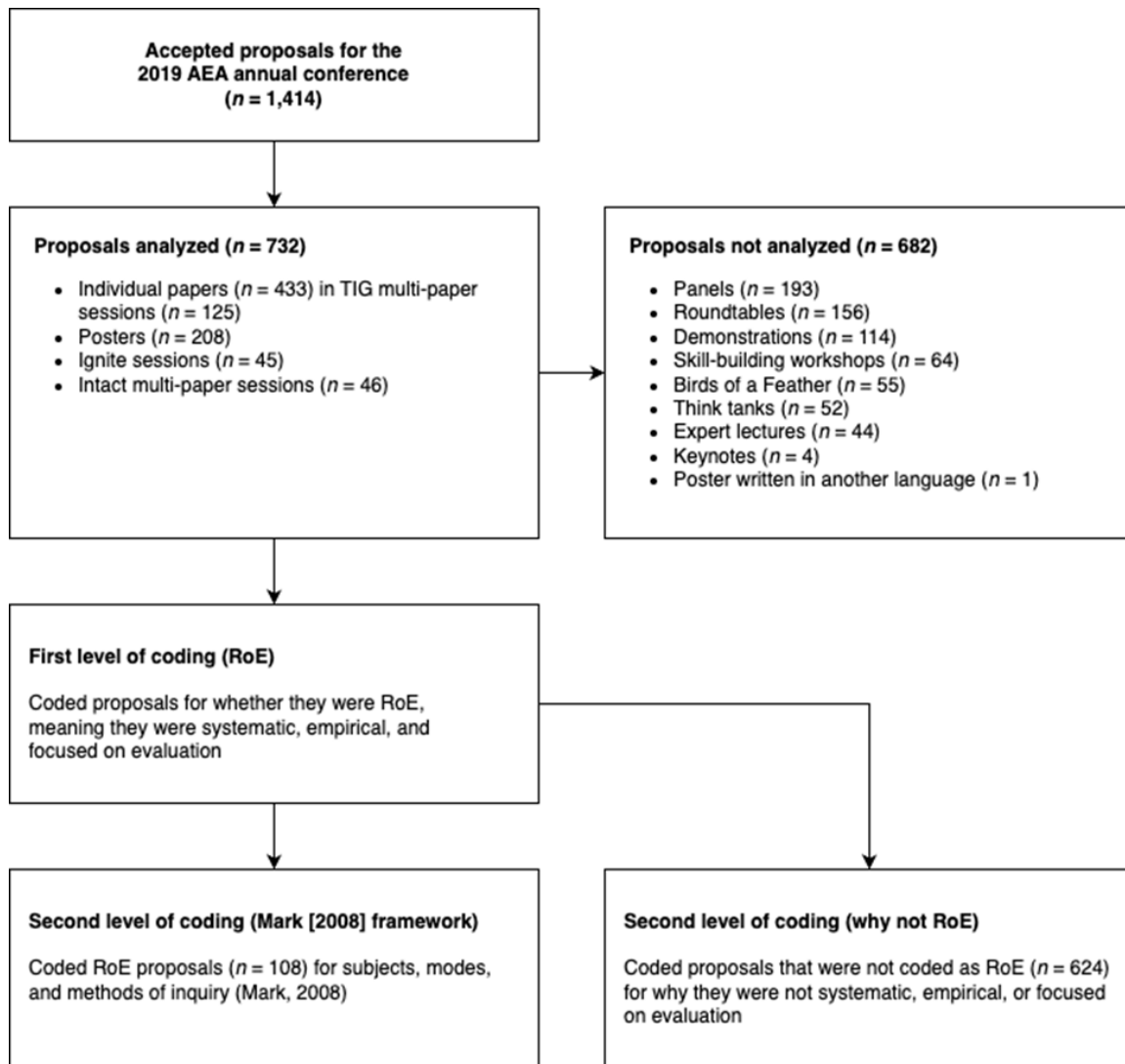
The data were imported into DistillerSR<sup>3</sup> for coding at two levels. The first level of coding was to determine whether or not the proposal was considered RoE. The second level of coding depended on the first level of coding; proposals that were coded as RoE were coded for subjects, modes, and methods of inquiry (e.g., Mark, 2008), whereas proposals that were not coded as RoE were further coded to indicate the reasons why they were not considered to be RoE. The subsequent sections provide details about the coding schemes for both levels of coding.

All authors or reviewers (used interchangeably going forward) of this paper were involved with coding. Each proposal was coded by two independent reviewers. Discrepancies in coding were resolved through the involvement of a third reviewer, through discussion between the original two reviewers, and/or through group discussion. Our focus was on resolving disagreements at the first level of coding for all proposals, and at the second level of coding for proposals that were coded as RoE. We did not attempt to resolve disagreements in the second-level coding for proposals that were not coded as RoE. We did not calculate interrater reliability for our coding. This was because we used a deliberation process to come to an agreement, which was more appropriate for this study since consensus was difficult to achieve (see the discussion section for more details).

---

<sup>3</sup> <https://www.distillersr.com/>

Figure 1. PRISMA Diagram for Inclusion Criteria and Coding Levels



*First Level of Coding: Whether Proposals Are RoE or Not.* We adopted the Coryn et al. (2016) conceptual definition of RoE: “Any purposeful, systematic, empirical inquiry intended to test existing knowledge, contribute to existing knowledge, or generate new knowledge related to some aspect of evaluation processes or products, or evaluation theories, methods, or practices” (p. 161). However, our operational definition had slight variations. Specifically, we screened proposals as RoE only if they met the following criteria: (a) They were empirical (based on observation or data); (b) they were systematic (planned and methodical); and (c) they focused on evaluation (related to evaluation processes, products, theories, methods, or practices). We chose not to include two of Coryn et al.’s (2017) criteria: (d) purposeful (done with

research intent) and (e) intended to test or contribute to knowledge. We reasoned that if a proposal was systematic and empirical, then it was purposeful. Furthermore, we assumed all conference proposals would be intended to test or contribute to knowledge in that they were shared directly back to the evaluation community

*Second Level of Coding: Coding RoE Proposals for Subjects, Modes, and Methods of Inquiry.* For proposals that were classified as RoE, we applied Mark’s (2008) framework of RoE. This framework encompasses two key components: subjects of inquiry and modes of inquiry. The subjects of inquiry focus on what is being studied, while the modes of inquiry delve into how the study is conducted.

Mark (2008) identified five *subjects of inquiry* within the RoE framework. The first is the evaluation *context*, which includes the circumstances in which the evaluation occurs, such as societal, organizational, evaluation-specific, and evaluand-related factors. The evaluation *activities* subject encompasses the procedures, methods, and practices involved in planning, executing, and disseminating evaluations, including different approaches, components within an evaluation model, and various practices. Evaluation *consequences* are the changes that occur or do not occur as a result of the evaluation; inquiries into this subject include investigating the use and influence of the evaluation; changes in participants or organizations (such as empowerment, evaluation capacity, or evaluative thinking); consequences on other categories; and outcomes specific to the evaluation itself. *Professional issues* constitute concerns related to the structure, norms, and development of the evaluation field; this includes aspects like training, standards, ethics, competencies, meta-RoE, metaevaluation, evaluation policy, and mainstreaming, as well as associations and networks within the field. The *domain-specific* subject of inquiry addresses issues specific to particular domains or sectors. It explores factors such as the sector under evaluation, subject areas within the evaluation, or specific topic areas being studied. Lastly, the *other* category, added by Coryn et al. (2017), encompasses any factors or considerations that are not covered by the aforementioned categories.

In addition to the subjects of inquiry, Mark's (2008) framework includes five *modes of inquiry* for classifying studies. The first mode is *description*, which involves measuring and quantifying the characteristics of clients, services, or outcomes under evaluation. Another mode is *classification*, which identifies and categorizes different groups or entities, such as clients, services, or settings. *Causal analysis* focuses on estimating causal effects and identifying their underlying mechanisms. *Values inquiry* delves into identifying the value positions held by stakeholders and the public, using evaluation methods to explore the values embedded in and associated with a program. *Tool/model development or validation* pertains to developing or validating evaluation-specific instruments, models, frameworks, or methods, including approaches like psychometrics, factor analyses, and grounded theory. Lastly, an *other* category added by Coryn et al. (2017) includes any modes not covered by the aforementioned categories.

Beyond the broad design (*mode of inquiry*) of each RoE proposal, we were interested in the specific types of *methods of inquiry* that were used. Mark's framework did not include these categories, so we added a third set of questions on design and methods (e.g., quasi-experiment, survey, interviews, case study, most significant change). We listed potential methods of inquiry at the start of our coding process, then added more methods as we coded. The subjects, modes, and methods of inquiry are summarized in Table 1.

Table 1. Codes Applied to RoE Proposals

Subjects of inquiry (Mark, 2008)	Modes of inquiry (Mark, 2008)	Methods of inquiry
Context	Description	Quasi-experiment
Activities	Classification	Survey
Consequences	Causal analysis	Interviews
Professional issues	Values inquiry	Case study
Domain specific	Tool/model development or validation	Most significant change
<i>Other</i>	<i>Other</i>	<i>Others identified during coding</i>

We treated codes within all three areas—subjects, modes, and methods of inquiry—as check-all-that-apply. Proposals could have more than one subject of inquiry, mode of inquiry, or method of inquiry; however, they were required to have at least one of each (otherwise they would not have been considered RoE). For proposals that described both an evaluation and a RoE study related to the evaluation, only the portion of the proposal that

described the RoE study was coded, so as not to include subjects, modes, or methods of inquiry that were related to an evaluation itself rather than the RoE aspect of the proposal.

*Second Level of Coding: Coding Non-RoE Proposals for Why They Are Not RoE.* For proposals that were not classified as RoE, we further coded the reasons why they were not

classified as RoE. Coders had the option to indicate that the proposal was one or more of the following: (a) not empirical, (b) not systematic, or (c) not focused on evaluation. Additionally, we assigned codes to indicate why a given proposal was not empirical (e.g., not empirical because it was only a reflection, not empirical because it did not specify data sources, not empirical because it was a conceptual paper, not empirical because the proposed study had not yet been conducted); not systematic (e.g., not systematic because it was only a reflection, not systematic because methods or analyses were not clearly described, not systematic because it was not an RoE-focused question, not systematic because it only described a process, tool, or framework); and/or not focused on evaluation (e.g., not focused on evaluation because it addressed a topic unrelated to evaluation, not focused on evaluation because it only described an evaluation itself, not focused on evaluation because it was not an RoE-focused question, not focused on evaluation because it was more focused on advancing research than evaluation). These codes were included to provide a better contextual understanding of what qualifies as RoE and what does not. Codes related to why proposals were not

empirical, not systematic, and/or not focused on evaluation were provided as check-all-that-apply options, so coders could indicate multiple reasons for not meeting each of the three aspects of RoE. Furthermore, coders had the option to add more reasons as new responses beyond our original set.

## Results

The AEA proposals were analyzed with descriptive statistics, including frequencies and percentages of proposals at both the first and second levels of coding. Examples are also provided throughout to exemplify some of the coding categories.

### *First Level of Coding: Whether Proposals Were RoE or Not*

Of the 732 proposals analyzed, a total of 108 (15%) were coded as RoE (see Table 2). Within each session type, multi-paper sessions had the highest percentage of RoE proposals, followed by individual papers, Ignite sessions, and poster sessions.

Table 2. Proposals by Session Type

Session type	Not RoE ( <i>n</i> )	RoE ( <i>n</i> )	RoE (%)
Multi-paper	34	12	26%
Individual paper	359	74	17%
Ignite	39	6	13%
Poster	192	16	8%
Total	624	108	15%

The TIG with the highest percentage of RoE proposals was the RoE TIG, followed by Teaching of Evaluation, Social Impact Measurement, Design and Analysis of Experiments, Extension Education Evaluation, Use and Influence of Evaluation, Indigenous Peoples in Evaluation, Quantitative Methods: Theory and Design, and Theories of Evaluation; see Appendix A for a full list of the number and percentage of RoE proposals by TIG. A total of 18 TIGs did not have a single RoE proposal.

*Second Level of Coding: Coding RoE Proposals for Subjects, Modes, and Methods of Inquiry*

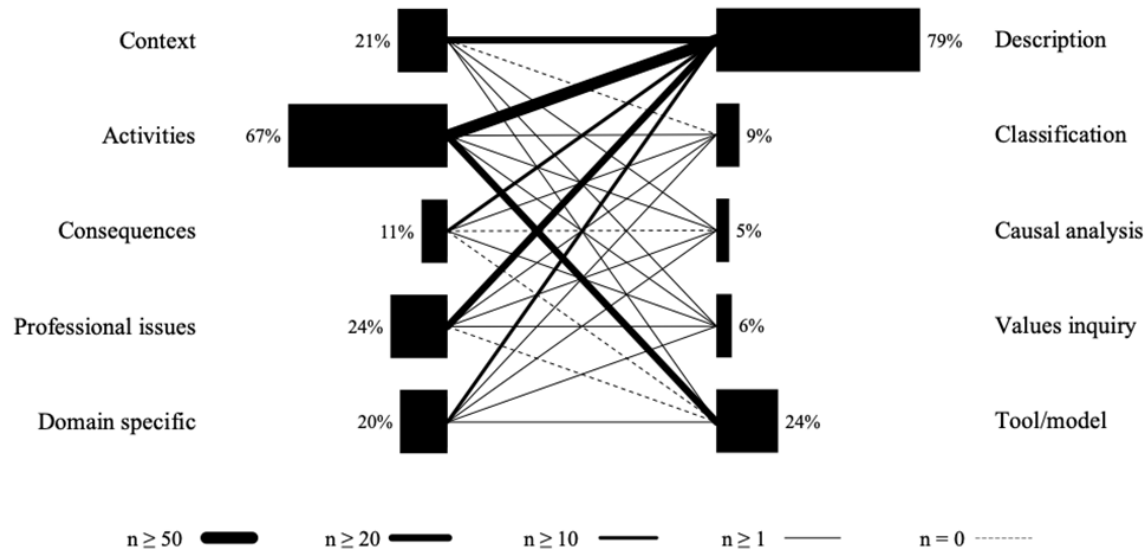
We further coded the 108 RoE proposals for their subjects of inquiry, modes of inquiry, and methods of inquiry; proposals could have multiple subjects of inquiry, modes of inquiry, and methods of

inquiry, so percentages can exceed 100%. Figure 2 visualizes the frequencies of subjects of inquiry and modes of inquiry.

For subjects of inquiry, most proposals examined evaluation activities ( $n = 72$ , 67%). Fewer proposals examined professional issues ( $n = 26$ , 24%), evaluation context ( $n = 23$ , 21%), specific domains ( $n = 21$ , 20%), or evaluation consequences ( $n = 12$ , 11%). The domains of inquiry were international cooperative extension, behavioral or mental health, NGOs, human services, translational research, clinical trials, English language learners, disaster response, youth education, public health, community-based evaluation, Indigenous peoples evaluation, government agencies, mental health, developing countries, STEM education, agriculture, and cancer research. Some proposals had at least two subjects

of inquiry; the most frequent pairings included activities and domain-specific ( $n = 15$ , 14%), context and activities ( $n = 13$ , 12%), and context and domain-specific ( $n = 8$ , 7%).

Figure 2. Combinations of Subjects of Inquiry and Modes of Inquiry



*Note.* The bars show the percentage of subjects of inquiry and modes of inquiry, which could be multiple, so percentages do not add up to 100%. The weight of the lines indicates how frequently each subject was paired with each mode of inquiry; a thicker line indicates a higher frequency, whereas a dotted line indicates zero frequency.

For modes of inquiry, most proposals used description ( $n = 84$ , 79%). Fewer proposals used tool/model development or validation ( $n = 26$ , 24%), classification ( $n = 10$ , 9%), values inquiry ( $n = 6$ , 6%), or causal analysis ( $n = 5$ , 5%). Some proposals had two or more methods of inquiry. The most frequent pairings included description with tool/model development or with validation ( $n = 8$ , 7%). Furthermore, all proposals that used values inquiry also used description ( $n = 6$ , 6%).

Figure 2 also visualizes the relationships between subjects of inquiry and modes of inquiry through the weighted lines. Most studies examined evaluation activities using either description ( $n = 52$ , 49%) or tool/model development or validation ( $n = 23$ , 21%). Many studies also used description in examining professional issues ( $n = 24$ , 22%), context ( $n = 21$ , 20%), specific domains ( $n = 15$ , 14%), or evaluation consequences ( $n = 12$ , 11%). No proposals featured the following combinations: (1) classification studies examining evaluation context, (2) causal analyses examining

evaluation consequences, (3) tool/model development or validation studies examining evaluation consequences, or (4) tool/model development or validation studies examining professional issues. See Appendix B for the tabular details of the intersections within Mark's (2008) framework.

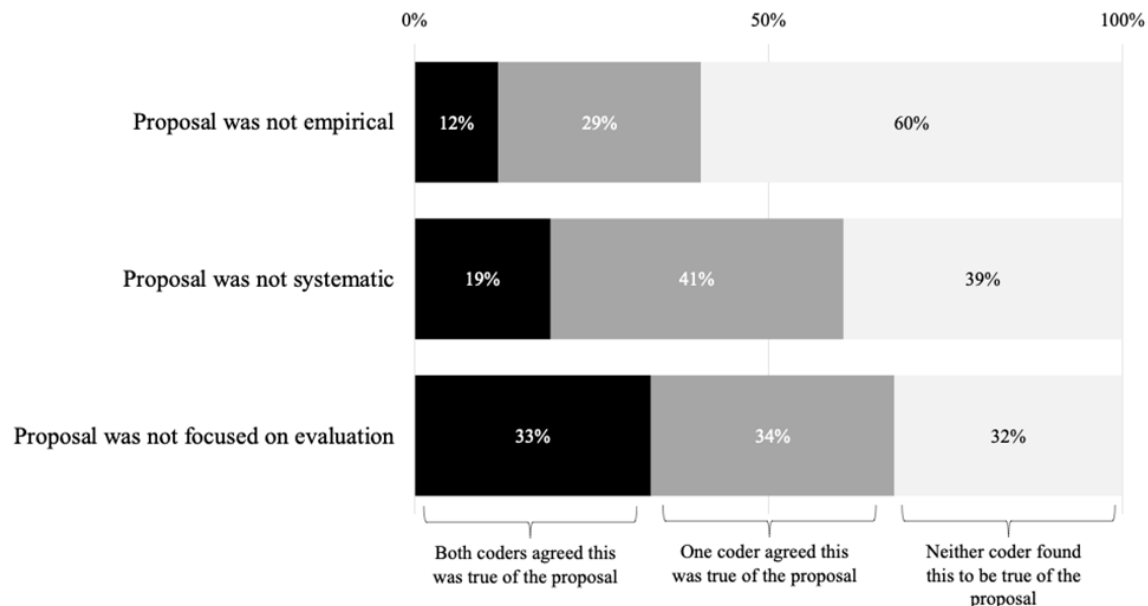
The most frequent methods of inquiry included interviews ( $n = 30$ , 28%), surveys ( $n = 24$ , 22%), case studies ( $n = 21$ , 20%), and literature review ( $n = 15$ , 14%). Less frequently used methods included secondary data analysis ( $n = 7$ , 7%), content analysis ( $n = 6$ , 6%), focus groups ( $n = 6$ , 6%), systematic reviews ( $n = 6$ , 6%), qualitative studies ( $n = 4$ , 4%), factor analysis ( $n = 3$ , 3%), and Delphi technique ( $n = 3$ , 3%). A variety of other methods were used in only one or two proposals: cluster analysis, group deliberation, quasi-experiments, concept mapping, meta-analysis, simulations, document review, metaevaluation, latent class analysis, observations, Q-methodology, social network analysis, logit regression analysis,

force field analysis, SWOT analysis, activities such as games or drawing, latent growth curve analysis, dynamic allocation method, discrete choice model, comparative analysis, multilevel modeling, mixed methods, scoping review, machine learning, benchmark analysis, item response theory, latent profile analysis, structural equation modeling, narrative analysis, cognitive interviews, score analyses, eigenvalue centrality, trend analysis, and environmental scans.

### Second Level of Coding: Coding Non-RoE Proposals for Why They Are Not RoE

We did not deliberate and come to agreement on the second level of coding for non-RoE proposals,

Figure 3. Proposals Coded as Not RoE, by Reason



Most non-RoE proposals were coded as such because they were not focused on evaluation; for example, 33% were coded by *both* reviewers as not focused on evaluation, and a further 34% were coded by *one* reviewer as not focused on evaluation. Of non-RoE proposals coded as not focused on evaluation, most were presenting an evaluation itself rather than conducting a research study on the evaluation; for example, many such proposals presented the methods and findings of an evaluation study. Others were coded as not focused on evaluation because they did not have a question focused on RoE; for example, many such studies presented theories, issues, tools, methods, or

since this was not the primary purpose of our paper. Therefore, given that each proposal was coded by two of the authors, we present three percentages in Figure 3: the percentage of proposals that (a) both reviewers coded as not empirical, systematic, or focused on evaluation [black category on the left], (b) only one reviewer coded as such [gray category in the middle of each bar], and (c) neither reviewer coded as such [light gray on the right]. When describing the specific reasons proposals were found to be not empirical, not systematic, or not focused on evaluation, we do not provide percentages. The information we include on this topic is more general and should be interpreted with care.

approaches that could be used or applied in many different areas and did not focus explicitly on evaluation. A few were presenting research on another topic; for example, one such proposal sought to provide information that could help professionals in another field but not evaluators.

Many non-RoE proposals were coded as such because they were not systematic. Most proposals coded as not systematic were reflections on an evaluation project or process (e.g., presented as retrospective notes observed about the evaluation) and did not systematically examine the evaluation project retrospectively. Others did not fully describe the study, methods, or analyses, so even

though they may have described RoE, they lacked the information we needed to fully discern whether the methods were systematic and to code for modes and methods of inquiry.

Lastly, some proposals were coded as not RoE because they were not empirical. Most non-empirical proposals were reflections (e.g., “lessons learned” following the reporting of an evaluation) and did not indicate data supporting those reflections (which might turn the proposal into more of a reflective case study). Many proposals coded as not empirical were conceptual papers that described theoretical or methodological concepts related to evaluation without providing data. Some proposals were coded as not empirical because their data sources were unclear, and so while they may have described RoE, the lack of information meant we could not fully discern whether they were empirical.

## Discussion

In this paper, we sought to examine the extent to which evaluators are doing RoE and presenting it at AEA conferences. To do so, we coded proposals accepted for presentation at the 2019 AEA conference. We focused our attention on session types that we thought were most likely to have RoE content: papers, multi-paper sessions, posters, and Ignite sessions.

Overall, we coded 15% of accepted proposals as RoE. Although we initially interpreted this as a low percentage, Mark—the discussant for our 2022 conference presentation on this paper (Nolton et al., 2022)—encouraged us to think of this number as an increase from earlier AEA conferences and noted that the absolute number of 107 proposals is an impressive number, especially since many were not in the RoE TIG. Nevertheless, future AEA conferences might consider adding an RoE checkbox to the conference proposal form or using RoE as a criterion for reviewing proposals.

It may be considered surprising that only 71% of RoE TIG proposals were coded as RoE. Four of the fourteen proposals were coded as not RoE, for a variety of reasons. The first proposal, a paper, described a framework but did not describe the systematic process for developing or testing the framework. The second proposal, a paper, described professionalization but did not explore the topic systematically or empirically. The third proposal, a paper, provided examples of collecting RoE data in evaluation projects but was a reflection on past experiences rather than a systematic study. The fourth proposal, a multi-paper session, described multiple evaluations as they related to a

specific topic but did not systematically review the evaluations.

Similar to Coryn et al. (2017), we found the most common subject of inquiry to be evaluation activities and the most common mode of inquiry to be description. However, compared to Coryn et al., we saw significantly more domain-specific proposals (20% versus 9%) and tool/model development or validation (24% versus 7%) and significantly fewer proposals focused on evaluation consequences (11% versus 27%). Additionally, as Mark (2008) has discussed, interesting data lies in the combinations between subjects of inquiry and modes of inquiry, as well as within subjects and within modes of inquiry (Nolton et al., 2022). Nearly half of the proposals were descriptive studies examining evaluation activities. Interestingly, all values inquiry proposals were also descriptive. We also found “empty” combinations, as Coryn et al. (2017) did, including classification inquiries into context, causal analysis of consequences, and tool/model development or validation to inquire into consequences or professional issues.

We echo Coryn et al. (2017) in stating that these findings “reveal numerous opportunities and need for further RoE” (p. 339). In the sections that follow, we discuss these opportunities and future directions for defining RoE, for evaluation practitioners and researchers, for the AEA conference, and for current and potential funders of RoE.

### *Issues in Defining RoE*

RoE has been broadly defined as a systematic inquiry into evaluation practice, methods, theories, and the evaluation profession itself (Coryn et al., 2016). However, there is no single, universally accepted definition, and perspectives on what constitutes RoE can vary across the field. For instance, Henry and Mark (2003) conceptualize RoE as research that is comparative, descriptive, or causal in nature, often examining evaluation practice or outcomes. Mark (2008) further categorized RoE according to its subject of inquiry (e.g., evaluation practice or systems), modes of inquiry (e.g., descriptive or causal studies), and methods of inquiry. This lack of consensus highlights both the complexity and the evolving nature of RoE, as its scope can encompass studies of evaluation theories, practices, and tools, as well as metaevaluations and explorations of the evaluation profession.

Throughout the screening and review of RoE in conference proposals, we discovered that the

project team members had different interpretations of the definition of RoE. As such, we engaged in regular discussions regarding what types of inquiry should be considered RoE and chose not to aim for interrater reliability but rather to build consensus regarding all proposals in which there was disagreement when coding. There were debates surrounding the level of systematicity and empiricism required for a proposal to be classified as RoE. For example, proposals that reflected on an evaluation project were challenging for our group: At what point does a reflection become a systematic inquiry? We also deliberated on whether proposals that presented preliminary findings or conceptual frameworks without reports of empirical evidence should be categorized as RoE. Another aspect we discussed was whether a proposal needed an RoE-focused research question followed by systematic data collection and analysis to be considered RoE. Lastly, we had concerns about the extent to which current definitions of RoE represent diverse forms of knowledge and methodologies in the field of evaluation.

The difficulty in consensus building could have resulted from the fact that all five researchers had different social locations (Lincoln & Guba, 2004) and evaluation experiences, and therefore brought unique perspectives to the coding process. Further, in addition to the inconsistent TIG review process, AEA conference proposals have a space limitation such that presenters can only use 150 words for abstracts and 500 words for relevance statements; perhaps if presenters had more space, they could better articulate the research aspects of their conference proposals. The space limitation may have also encouraged us as coders to “read into” proposals and make inferences from the limited information presented. Another difficulty is that some evaluators may not be aware of or care about sharing the extent to which their conference proposals are RoE. There are also many definitions and categorizations of RoE (Henry & Mark, 2003; Mark, 2008), which can make identifying RoE challenging. Furthermore, the definitional challenges of distinguishing evaluation from research can make the discussion of research on evaluation challenging (Linnell, 2023).

These limitations suggest more work is needed to define RoE and determine the categorization of the subjects of inquiry and modes of inquiry in RoE. There has been preliminary work in defining RoE, including definitions within past RoE studies (Brandon, 2015; Brandon & Fukunaga, 2014; Coryn et al., 2016). Furthermore, the RoE TIG<sup>4</sup> has

explored creating a definition of RoE, but this has not been formally adopted by the TIG or by the field writ large (Linnell, 2020). Adopting a formal definition of RoE may be useful in creating a future research agenda on RoE; however, we recognize the complexity of this challenge and how limiting a single definition might be.

### *Implications for Evaluation Practitioners*

Our analysis of accepted conference submissions also has implications for evaluation practitioners. It provides practitioners with valuable insights into the latest methodologies, tools, and best practices, enabling them to enhance the quality and rigor of their evaluation work. As noted by Mark (2008), RoE can build a better evidence base for evaluation theory and provide practitioners with guidance on best practices.

Second, RoE also opens doors for collaboration between evaluation researchers and practitioners. We argue that a primary goal of RoE is to support evaluation practitioners with the practical issues they face in the field; therefore, close collaboration can bridge the gap between academic research and real-world application (Tseng, 2012). Researchers can benefit from practitioners’ insights and questions, which can help tailor research to address practical needs. This presents an exciting opportunity to explore these inquiries collectively, advancing the field and practice of evaluation through a collaborative effort between researchers and practitioners, ultimately improving the quality and impact of evaluation in diverse settings. However, practitioners could also conduct RoE themselves instead of collaborating with evaluation researchers to conduct RoE; indeed, our analysis of accepted conference submissions suggests many practitioners are doing RoE even if they may not call what they are doing RoE.

To build RoE into their practice, practitioners can attend relevant conference sessions about how to conduct RoE, engage in post-conference discussions and communities of practice around RoE, and actively seek out RoE that aligns with their evaluation projects. However, practitioners remain faced with many challenges in engaging in RoE. One significant challenge is the time and resource constraints often associated with their daily work. Engaging in RoE may require additional time and effort that practitioners might find hard to allocate. To overcome this challenge, practitioners can explore opportunities for partnership and collaboration with researchers, academic

<sup>4</sup> <https://comm.eval.org/researchonevaluation/home>

institutions, or other organizations. Collaborative efforts can help distribute the workload and access necessary resources for RoE. This approach would allow researchers to leverage practitioners' expertise and insights on evaluation practice and assist practitioners in incorporating RoE into their work. Furthermore, certain types of RoE (e.g., certain subjects of inquiry and modes of inquiry) may be easier entry points for practitioners than others; for example, a retrospective descriptive analysis of a past evaluation focusing on its evaluation context or activities may be easier to conduct for practitioners than a causal study on professional practices.

Another challenge for practitioners may be the need for specific research skills and knowledge, to which not all practitioners may have access. To address this, practitioners can invest in professional development and training to build their research capacity. Additionally, staying informed about the latest RoE and leveraging existing networks within the evaluation community can facilitate knowledge sharing and support in integrating RoE into their work. Lastly, practitioners can prioritize the alignment of RoE efforts with their evaluation goals and objectives, ensuring that research activities directly contribute to enhancing the quality and effectiveness of their evaluation practice.

### *Implications for the AEA Conference*

This research also suggests implications for the AEA conference itself. Proposals are currently submitted to TIGs, and TIGs then manage the review process. However, currently TIGs are limited to only accepting a certain percentage of submissions they receive. Recently, AEA underwent a scan to reimagine the TIG structure and process, including how to reduce the number of TIGs to make the groups more manageable for AEA. The TIG scan and the Conference Advisory Working Group have been reimagining the TIG structure and conference review process as a result. For example, the past two conference evaluation surveys have asked attendees to reflect on submitting proposals via "tracks" in addition to TIGs. Similarly, conference proposals now indicate which evaluator competencies (AEA, 2018c) and guiding principles (AEA, 2018b) the proposal aligns with. We suggest considering an RoE "track/stream" or allowing applicants to indicate via checkbox that the submission is an RoE proposal.

The conference remains a viable and valuable platform for showcasing RoE and sharing best

practices. Encouraging presentations and sessions specifically dedicated to RoE would foster a stronger focus on the application and relevance of evaluation research within the AEA community. In terms of proposal submissions, professional development sessions on how to better articulate the RoE component in conference proposals could be another useful strategy to promote RoE via the AEA conference.

### *Implications for Funders*

Lastly, this study has implications for funders of RoE (e.g., commissioners of evaluation, funders of research, funders of capacity building). Currently, there are limited funding opportunities for RoE. Funders might consider including RoE in their call for proposals, particularly to help advance the evaluation outcomes of interest their organization is seeking. For example, a funder interested in building evaluation capacity might offer a call for proposals on RoE focused on that topic; findings from these RoE studies could then be used to inform the programs and evaluations they fund moving forward to support their evaluation capacity. This could involve explicitly requesting evaluation projects that have a RoE component, encouraging collaboration between researchers and practitioners, or dedicating funding specifically for RoE initiatives. Adding RoE as an additional or optional component in the RFP could help catalyze thinking about RoE right from the initial phases of designing evaluations. Incorporating RoE components when funding evaluations would help create more opportunities and resources for conducting and disseminating RoE.

### *Future Directions for RoE*

Researchers interested in pursuing RoE could benefit from understanding the breadth and depth of conference abstract submissions. Our coding and deliberation process highlighted differences in opinion on what is considered RoE, and likely these perspectives permeate the field. Future research should examine perceptions of what RoE is to help elucidate a field-accepted definition of RoE. In doing so, researchers could further RoE by creating more inclusive and representative categories of what is considered RoE to help broaden RoE as a discipline within the evaluation field. Future research might also take a more inductive approach to categorizing RoE, using existing RoE to determine the framework itself, as opposed to using pre-existing frameworks on RoE (e.g., Henry & Mark, 2003; Mark, 2008).

Future research could also analyze the other types of conference proposals we did not analyze (e.g., roundtables), additional conference years, how conference theme or location might affect the prevalence of RoE, and the prevalence of RoE at other evaluation conferences. Additional comparative analysis would provide valuable insights into the effectiveness of different evaluation approaches, the outcomes they produce, and the factors that influence evaluation practice. Additionally, researchers should seek out collaborations with practitioners in order to tap into practitioners' expertise and insights on evaluation practice and incorporate RoE into new/ongoing evaluation projects. Finally, it would be valuable to examine how AEA's submission and review processes have evolved over time, particularly as these processes continue to shift. A longitudinal analysis of RoE submissions and reviews could provide insights into trends, emerging priorities, and changes in practice, offering a deeper understanding of how RoE is supported and advanced within the field.

This research also suggests many areas ripe for further study within RoE. The modes of inquiry of causal analysis, values inquiry, classification, and consequences all had prevalence below 11% and could benefit from more focus on these areas in RoE. Additionally, as Dr. Melvin Mark noted during his discussion of our paper at the 2022 AEA conference, researchers should focus more on the intersections between subjects of inquiry and modes of inquiry. All of this suggests many areas of opportunity in the conduct of RoE. However, future research should conduct a more formal gap analysis or needs assessment within the field to identify which areas of RoE require future study.

### *Strengths and Limitations*

A strength of our study lies in its collaborative and inclusive process of coding RoE proposals. Some limitations are that we focused on the 2019 conference, we focused only on accepted proposals rather than on all submitted proposals, and we made assumptions about session types and did not code all proposals presented at the 2019 conference. Furthermore, we acknowledge that not everyone is able to attend or present at conferences. Attendance is affected by many factors, including location, time, and cost; presentation is affected by other factors, including how many proposals are able to be accepted and, at the time, differences in how proposals were reviewed by TIGs.

### *Conclusions*

This research suggests that there is RoE occurring across the AEA conference beyond just the RoE TIG. RoE can help contribute to better evaluation theory and practice, examine untested ideas that inform the field of evaluation, promote the field of evaluation, and document the current and past contributions of evaluation (Christie, 2003; Mark, 2008; Szanyi et al., 2013). RoE can also help evaluators work on their evaluator competencies (e.g., supporting the domain of professional practice or better focusing on the methodology or context domains; AEA, 2018c); ensure ethical evaluations through the AEA (2018b) guiding principles; and promote high-quality evaluations that support the JCSEE program evaluation standards (Yarbrough et al., 2010). Our analysis of conference submissions highlights an evolution and increase of RoE at the AEA conference that should be fostered and explored further. By engaging practitioners, refining a universal definition of RoE, properly coding conference proposals, and securing adequate funding, we can advance the field of evaluation and ensure that research is effectively applied to address practical challenges.

### *Acknowledgements*

All authors contributed equally to this article. Western Michigan University provided the DistillerSR contract and paid for copyediting.

### *References*

- American Evaluation Association. (2018a). *American Evaluation Association 2018 member survey report*. [https://www.eval.org/Portals/0/Docs/2018%20AEA%20Survey%20Report%20Final%20\(02\).pdf](https://www.eval.org/Portals/0/Docs/2018%20AEA%20Survey%20Report%20Final%20(02).pdf)
- American Evaluation Association. (2018b). *Guiding principles for evaluators*. <https://www.eval.org/About/Guiding-Principles>
- American Evaluation Association. (2018c). *AEA evaluator competencies*. <https://www.eval.org/About/Competencies-Standards/AEA-Evaluator-Competencies>
- Brandon, P. R. (2015), Editor's notes. *New Directions for Evaluation*, 148, 3–6. <https://doi.org/10.1002/ev.20152>
- Brandon, P. R., & Fukunaga, L. L. (2014). The state of the empirical research literature on

- stakeholder involvement in program evaluation. *American Journal of Evaluation*, 35(1), 26–44. <https://doi.org/10.1177/1098214013503699>
- Christie, C. A. (2003). What guides evaluation? A study of how evaluation practice maps onto evaluation theory. *New Directions for Evaluation*, 97, 7–36. <https://doi.org/10/fwg4xq>
- Coryn, C. L. S., Ozeki, S., Wilson, L. N., Greenman, G. D., Schröter, D. C., Hobson, K. A., Azzam, T., & Vo, A. T. (2016). Does research on evaluation matter? Findings from a survey of American Evaluation Association members and prominent evaluation theorists and scholars. *American Journal of Evaluation*, 37(2), 159–173. <https://doi.org/10/f8qqvk>
- Coryn, C. L. S., Wilson, L. N., Westine, C. D., Hobson, K. A., Ozeki, S., Fiekowsky, E. L., Greenman, G. D., & Schröter, D. C. (2017). A decade of research on evaluation: A systematic review of research on evaluation published between 2005 and 2014. *American Journal of Evaluation*, 38(3), 329–347. <https://doi.org/10/ggf5td>
- Henry, G. T., & Mark, M. M. (2003). Toward an agenda for research on evaluation. *New Directions for Evaluation*, 97, 69–80. <https://doi.org/10/b5d6xg>
- Lincoln, Y. S., & Guba, E. G. (2004). The roots of fourth generation evaluation. In M. C. Alkin (Ed.), *Evaluation roots: Tracing theorists' views and influences* (pp. 225–241). Sage.
- Linnell, D. (2020, April 23). What is research on evaluation (RoE)? [Blog post]. <https://www.danalinnell.com/blog/whatisroe>
- Linnell, D. (2023, July 30). Research, evaluation, research ON evaluation?! *AEA365*. <https://aea365.org/blog/research-evaluation-research-on-evaluation-by-dana-linnell/>
- 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). Guilford.
- Mata, H., Latham, T. P., & Ransome, Y. (2010). Benefits of professional organization membership and participation in national conferences: Considerations for students and new professionals. *Health Promotion Practice*, 11(4), 450–453. <https://doi.org/10.1177/1524839910370427>
- Nolton, E. C., Harnar, M., Moore, T. R., Mahato, S., & Wanzer, D. L. (2022, November 11). Exploring RoE across TIGs: RoE within accepted 2019 AEA conference proposals. In M. Harnar & M. M. Mark (Chairs), *(Re)shaping our understanding of research on evaluation (RoE)* [Multipaper presentation]. American Evaluation Association 2022 Annual Conference, New Orleans, LA, United States.
- Szanyi, M., Azzam, T., & Galen, M. (2013). Research on evaluation: A needs assessment. *The Canadian Journal of Program Evaluation*, 27(1), 39–64.
- Tseng, V. (2012). The uses of research in policy and practice. *Social Policy Report*, 26(2). <https://doi.org/10/ggf6bc>
- Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2010). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Sage.

## Appendix A

### RoE Proposals by TIG ( $n = 732$ )

TIG	Not RoE ( $n$ )	RoE ( $n$ )	% RoE
Advocacy and Policy Change	9	0	0%
Arts, Culture, & Museums	8	0	0%
Assessment in Higher Education	14	2	13%
Behavioral Health	13	4	24%
Cluster, Multi-site & Multi-level Evaluation	17	1	6%
Collaborative, Participatory & Empowerment Evaluation	28	3	10%
College Access Programs	3	0	0%
Community Development	8	1	11%
Community Psychology	4	0	0%
Costs, Effectiveness, Benefits, & Economics	6	0	0%
Crime and Justice	7	0	0%
Data Visualization & Reporting	7	0	0%
Democracy & Governance	6	1	14%
Design & Analysis of Experiments	5	3	38%
Disabilities & Underrepresented Populations	9	0	0%
Disaster & Emergency Management Evaluation	8	1	11%
Distance Education & Other Educational Technologies	10	0	0%
Environmental Program Evaluation	12	1	8%
Evaluation Managers & Supervisors	2	0	0%
Evaluation Policy	3	0	0%
Extension Education Evaluation	9	5	36%
Feminist Issues in Evaluation	7	1	13%
Government Evaluation	18	1	5%
Graduate Students and New Evaluators	5	1	17%
Health Evaluation	37	1	3%
Health Professions Education, Evaluation & Research	16	1	6%
Human Services Evaluation	10	2	17%
Independent Consulting	4	0	0%
Indigenous Peoples in Evaluation	4	2	33%
Integrating Technology into Evaluation	11	0	0%
Internal Evaluation	11	2	15%
International & Cross Cultural Evaluation	24	7	23%
Latino Issues in Evaluation	5	1	17%
Leadership & Organizational Performance	9	0	0%
Military & Veterans	5	0	0%
Mixed Methods Evaluation	12	1	8%
Multiethnic Issues in Evaluation	3	1	25%
Needs Assessment	7	1	13%
NonProfit & Foundations	9	1	10%
Organizational Learning & Evaluation Capacity Building	15	2	12%

TIG	Not RoE (n)	RoE (n)	% RoE
PreK-12 Educational Evaluation	36	4	10%
Presidential Strand	6	0	0%
Program Design	6	2	19%
Program Theory & Theory-Driven Evaluation	6	1	14%
Qualitative Methods	17	5	23%
Quantitative Methods: Theory & Design	16	8	33%
Research on Evaluation	4	10	71%
Research, Technology & Development Evaluation	18	6	25%
Social Impact Measurement	3	2	40%
Social Network Analysis	8	1	11%
Social Work	4	0	0%
STEM Education & Training	21	2	9%
Systems in Evaluation	12	0	0%
Teaching of Evaluation	9	6	40%
Theories of Evaluation	7	3	30%
Translational Research Evaluation	15	4	21%
Use & Influence of Evaluation	11	6	35%
Youth Focused Evaluation	25	1	4%
Total	624	108	15%

## Appendix B

### RoE Proposals by the Intersections of Mark's (2008) Framework

		1	2	3	4	5	6	7	8	9	10
Subjects of Inquiry	1. Context	23									
	2. Activities	13	72								
	3. Consequences	4	3	12							
	4. Professional Issues	4	5	2	26						
	5. Domain Specific	8	15	3	2	21					
Modes of Inquiry	6. Description	21	52	12	24	15	84				
	7. Classification		4	1	6	2	5	10			
	8. Causal Analysis	1	4		1	1	2	1	5		
	9. Values Inquiry	4	3	1	1	1	6			6	
	10. Tool/Model	3	23			5	8	2	1	1	26

*Note.* Numbers on the diagonal are proposals coded as each of the subjects of inquiry and modes of inquiry. Below the diagonal are proposals coded as two subjects or modes of inquiry.