New Directions for Evaluation

WILEY

ORIGINAL ARTICLE OPEN ACCESS

Redefining Research on Evaluation to Unlock Its Full Empirical and Conceptual Potential

Ghislain Arbour

Faculty of Education, University of Melbourne, Victoria, Australia

Correspondence: Ghislain Arbour (ghislain.arbour@unimelb.edu.au)

Received: 18 July 2025 | **Accepted:** 21 July 2025

ABSTRACT

This article redefines the term Research on Evaluation (RoE) by applying principles from concept analysis and terminology. A central contribution is the distinction between two types of evaluation-related research. RoE refers to the scientific inquiry of evaluation phenomena, including evaluators and their practices, but also other evaluation social objects. In contrast, Research in Evaluation (RiE) denotes the scientific inquiry of the value of social interventions, the subject matter of evaluation as a discipline. RiE is essential to the field, as it provides the substantive theories that underpin the defining activity of evaluation: making warranted judgments of value. The article further clarifies RoE and RiE by providing several conceptual adjustments. It argues that both must be understood within the broader principles of scientific research, which extend beyond notions of systematic or purposeful inquiry. It challenges the exclusion of non-empirical work as inconsistent with the general nature of research and as unnecessarily limiting for both RoE and RiE. Finally, it contends that the domain of RoE is broader than evaluation practice alone and encompasses a wider array of evaluation-related phenomena.

1 | Introduction

This article reconsiders the concept of *Research on Evaluation* (RoE) by critically examining its current definitional assumptions. The analysis is grounded in the field of program evaluation, rather than evaluation in general or other domains such as personnel evaluation. As such, this article considers the concept of *Research on Program Evaluation*, and unless otherwise specified, *evaluation* is used here to mean *program evaluation*.

To situate this redefinition, it is helpful to begin by sketching the prevailing understanding of RoE. In their Editors' Notes, Aston et al. (2025) revisit several influential definitions, ending with the relatively authoritative one by Coryn et al. (2017), in which 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 (161).

Further conceptual grounding can be found in taxonomies proposed by Henry and Mark (2003) and revised by Mark (2008), who identify the main subjects of inquiry in RoE as evaluation contexts, evaluation activities, evaluation consequences, and professional issues. Finally, there is a general agreement among RoE champions, including many of the contributors to this issue: RoE matters because it strengthens evaluation knowledge, which should, in turn, support the practice of evaluation and solidify its legitimacy in society. Taken together, these sources conceptualize RoE around four key characteristics:

1. It exhibits specific research qualities, like systematism;
2. It is empirical;

3. Its domain of investigation is connected to evaluators and their practice; and
4. As a consequence of the previous characteristics, it contributes to the theoretical development of the evaluation discipline.

Before discussing these four key characteristics in more detail, I draw on work from concept analysis to provide the general framework used in this article to establish meaningful conceptual distinctions around RoE.

2 | General Approach to Concept Analysis and Definition-Making

This article challenges each of these four assumptions by drawing on established work in concept analysis and terminology. Specifically, it uses Giovanni Sartori's framework for concepts in the social sciences (Collier and Gerring 2009; Sartori 1984) and the general theory of terminology developed by Eugen Wüster and formalized in *Terminology work—Principles and Methods* by the International Organisation for Standardization (ISO 704 2022).

Sartori, drawing on Ogden and Richards (1946, as cited in Sartori 1984), distinguishes three key elements in concept analysis:

1. Term—the word used to refer to a concept (e.g., evaluation standard);
2. Concept—the abstract unit of meaning or what the word evokes in our minds (e.g., a reference point used to judge the performance of an evaluand); and
3. Referent—the real-world phenomenon the concept represents through abstraction (e.g., all the instances of benchmarks, norms or other references that effectively correspond to the concept of standard).

Moreover, a *definition* attempts to capture the essence of a concept (Goertz 2006, 2020; Sartori 1984). Good definitions are clear and unambiguous, encompassing all possible cases of the referent and only them (Sartori 1970, 1984). For example, it would be undesirable for a definition of RoE to exclude true embodiments of RoE, or to include forms of inquiry that do not truly fit the concept. Wüster adds that definitions should be economical: they must include all necessary characteristics of a concept and only those necessary (ISO 704 2022, 7–8, 33–34). These terminological prescriptions provide a concept an exclusive space of meaning and ensure coherent relationships with other concepts (Gerring and Christenson 2017, 32–34; ISO 704 2022, 8, 33–34).

In the case of evaluation-related research, clear definitions support unambiguous dialogue among researchers and promote a more deliberate recognition of what kinds of research are being conducted—and what might be missing. They may help evaluation to mature as a discipline by equipping its contributors with the conceptual tools needed to identify, organize, and expand their work with additional clarity and intentionality.

3 | An Argument to Untangle Confusion

The core issue this article identifies is a conflation of referents: it is generally assumed that evaluation-related research constitutes a single real-world phenomenon, captured by the current concept of RoE with its four key characteristics. Specifically, two types of research are collapsed into one concept: the empirical study of evaluation practice (aligned with characteristics no. 1, 2, and 3) *and* the research that underpins and advances substantive evaluation theory (the purpose detailed in characteristic no. 4).

However, the forthcoming analysis reveals that these two types of inquiry do not share the same essential characteristics. Their respective intellectual natures fundamentally differ and therefore they cannot be abstracted into a single concept or governed by the same definition. This situation renders necessary the acknowledgment of two distinct and complementary concepts of evaluation-related research:

> Research on Evaluation (RoE): the scientific inquiry of evaluation phenomena.

> Research in Evaluation (RiE): the scientific inquiry of the value of social interventions.

In RoE, evaluation is the object of investigation. In RiE, evaluation is the disciplinary context.

These definitions represent the conceptual destination of this article's argument. The argument unfolds across four main components:

1. It considers the underlying concept of scientific research for RoE and RiE;
2. It rejects the exclusion of non-empirical work for RoE and RiE;
3. It argues for expanding the scope of RoE beyond evaluation practice; and
4. It reveals the misalignment between a discipline of value determination and a research program focused on evaluation practice.

4 | Four Characteristics of RoE

4.1 | RoE Exhibits Specific Research Qualities, Like Systematism

This section reconsiders the first characteristic of the current concept of RoE (i.e., that it displays specific research qualities like systematism). It argues that while systematism is a legitimate scientific principle, it should not be a strict and limiting feature of either RoE or RiE.

Defining RoE and RiE demands considering the shared underlying concept of *research*, or, more precisely, *scientific research*, given the intended connection with the development or testing of evaluation theories. The term research is used in this article in that scientific sense.

Research is characterized by the adoption of scientific principles, discussed at the intersection of research methodology and epistemology (e.g., Blaikie 2007; Brady and Collier 2010; Crotty 1998; Gerring 2012b; Mahner 2007). These norms aim to ensure that the findings generated by research efforts are warranted and suitable for inclusion in the accumulated knowledge of their disciplines.

Systematic inquiry is a principle that frequently appears in definitions of scientific research. Many research textbooks and disciplinary accounts define social science research as the systematic inquiry of social phenomena (e.g., Babbie 2016). Therefore, it is not surprising to find the characteristic of systematism appears in RoE definitions. In its most general understanding, "systematic" is the quality of something done "according to a system, plan, or organized method" (Oxford University Press n.d.). This concept is not extensively developed in the literature about RoE, but reviews conducted by Coryn et al. (2017) and Linnell and Stachowski (2025) describe it with the words "planned; methodical" as part of their operational definition of RoE.

Nevertheless, systematism is one of many legitimate scientific principles. Gerring, for instance, considers that research is, in addition to systematic, "cumulative, evidence-based, falsifiable, generalizing, nonsubjective, replicable, rigorous, skeptical, transparent, and grounded in rational argument" (Gerring 2012b, 2, 11). Limiting the definition of RoE and RiE to one or few principles, like systematic inquiry, is inaccurate, as no single principle is sufficient on its own.

Conversely, an expansive list of principles invites debates and poses risks of rejection across the wide range of perspectives in evaluation. Moreover, the relative weight or even applicability of these principles varies depending on the nature of the investigation at hand, including the possibility of pragmatic tradeoffs between principles (Gerring 2012b; Gerring and Christenson 2017). While all research works should be grounded in rational argument, some may not be cumulative, empirical, or even systematic, and still provide warranted insights into the world.

To address these definitional challenges, the definitions of RoE and RiE proposed in this article start with the phrase "the scientific inquiry." This clearly asserts the scientific character of research with its underlying scientific principles without specifying any exhaustive list of said principles.

One final point concerns the adjective purposeful, suggested in the definition by Coryn et al. (2017) to distinguish RoE from accidental or incidental inquiry. This label is ultimately superfluous. The term systematic in the current definition already implies intentionality, and the term scientific inquiry, as proposed in this article, even more so: rationality, rigor, or the search for evidence in scientific research all presume a purposeful process. To be clear, scientific findings may emerge through serendipity. Yet for such discoveries to count as scientific contributions, they must be retrospectively appraised against relevant scientific principles and therefore fall within the proposed definitions of RoE and RiE.

4.2 | RoE Is Empirical

This section addresses the second characteristic of the current concept of RoE, which views it as a form of empirical research. This section argues that non-empirical work should not be excluded from definitions of RoE and RiE.

Research involves a complex relationship between theoretical reasoning and empirical data. Not all knowledge contributions emerge from empirical work. The kind of work conducted in this very article, concept analysis, falls within non-empirical processes such as abstraction and logic. Concept analysis is itself foundational for empirical research phases: researchers rely on it to express meaningful categories of things to discuss, measure, or compare, and, without it, there are no empirical accounts such as descriptions or causal claims (Gerring 2012a; Sartori 1984).

Moreover, many concepts and theories are not meant to be empirically verified in the traditional sense, but instead prove their value in conceptual illumination. This is true for evaluation, which needs to rely, among other things, on conceptual systems about human values in society and evaluation reasoning, discussing topics like the nature of value (e.g., Putnam 2002), needs (e.g., Scriven and Roth 1978), justice (e.g., Rawls 1971), utility and capability (e.g., Adler 2015; Sen 1999), or the structure of evaluation arguments (e.g., Fournier 1995; Toulmin 2003).

Even in the development of empirically grounded theories, shifts are not always driven by new empirical observations. Instead, as Kuhn argued in *The Structure of Scientific Revolutions*, new theoretical frameworks arise because they prove better at explaining or making sense of existing observations (Kuhn 2012).

It is worth mentioning that the requirement of an empirical character is problematic even within meta-RoE using RoE typologies. Mark's classification of RoE based on modes of inquiry (Mark 2008) includes a category covering the development and validation of evaluation models and tools. Yet this category cannot be meaningfully understood without allowing for a significant theoretical component, that is, the model itself. This creates a paradox: RoE is defined as *strictly empirical*, but one of its categories has a *necessarily theoretical* component. In comparing the RoE reviews of Coryn et al. (2017) and Linnell and Stachowski (2025), Linnell and Stachowski note that differences in interpretations for that category might have led to an increased prevalence in their review.

Given the role of theoretical work in research, it is unnecessarily restrictive to define RoE (and RiE) as strictly empirical. The definitions proposed in this article avoid specifying an empirical or theoretical character, by simply referring to the more inclusive phrase "scientific inquiry" instead. This approach accommodates the full range of legitimate research activities, including conceptual work essential to the development of evaluation theory.

4.3 | RoE's Domain of Investigation Is Connected to Evaluators and Their Practice

This section challenges the third characteristic of the current concept of RoE, which stipulates that its domain of investigation relates to social phenomena close to evaluators and their practice. The argument here is twofold: first, RoE should not be confined to studying evaluation practice; and second, RoE should not be restricted to the theoretical and methodological traditions of the evaluation discipline.

The phrase "research on" typically signals a research program unified by a common object of study, regardless of disciplinary affiliation. Journals like *Research on Aging*, *Research on Social Work Practice*, or *Research on Language and Social Interaction*, offer collections of contributions organized around the study of a phenomenon, not around the consolidation of a specific academic field. RoE borrows this same naming convention, a logical choice considering it has been traditionally built around a shared object, evaluation practice.

However, while evaluation practice is a legitimate focus, it is only a subset of the broader category of evaluation phenomena to do "research on." This includes, for instance, evaluation language and terminology, evaluation policies and institutions, the supply and demand of evaluation services, the public access to evaluations and its legal or ethical implications, the functions of evaluation in different forms of government, the ethos and advocacy purposes of evaluation associations, or, again, the emergence of evaluation mechanisms in self-regulated industries. All of these are evaluation social objects. A concept called RoE should include research activities relating to them as well as to other evaluation phenomena.

The narrow association of RoE with evaluation practice is likely due to its perceived role as the research counterpart of a theory of practice, a matter addressed in the next section. Nevertheless, there is no semantic justification to equate RoE with research on evaluation practice alone. For that reason, the proposed definition avoids enumerating specific evaluation objects as in Coryn et al. (2017), referring instead to "evaluation phenomena" as the domain of inquiry. In other words, *what* RoE studies should not be limited to a specific class of evaluation social objects.

An adjacent argument is that *how* RoE studies evaluation phenomena should not be limited to the theoretical and methodological traditions of the evaluation discipline. While the idea is not rejected in evaluation literature per se, it is worth stressing it as part of appraising the potential range of RoE. Regardless of a predictably higher interest of evaluation scholars for evaluation phenomena, the latter belong to any discipline that can shed theoretical light on them, raise original questions, or offer valid methodologies for their investigation. Other disciplines could, in principle, pursue their own research programs on evaluation phenomena, such as the sociology of evaluation or evaluation economics, just like the sociology of education and education economics investigate education phenomena. As such, RoE should be seen as the realm of the social sciences and the humanities in general, not the dominion of the evaluation discipline.

To illustrate the point above, one may consider the category "evaluation context" from Mark's typology of RoE subjects (Mark 2008), discussed by Li and Westine in their synthesis (Li and Westine 2025). This topic can be studied using theoretical and methodological tools from the evaluation discipline, but also from well-established traditions that offer alternative lenses. For instance, New Institutional Economics theorizes about how institutional contexts shape behavior (e.g., Crawford and Ostrom 2005; North 1990; Ostrom 1990), potentially illuminating how governance structures influence evaluation use. Another example is symbolic anthropology (Geertz 1973), which conceptualizes context in terms of systems of meanings, offering thick description as a method to study local evaluation cultures. Such extra-disciplinary openness already exists in RoE—Neubauer et al. (2025), for example, draws on contemporary education theory to organize their research program on evaluator education and training.

These two observations—that the objects of RoE should include all evaluation phenomena and that their study should not limit itself to the perspectives of the evaluation discipline—has important consequences on how to categorize RoE studies. At the intersection of many possible evaluation objects and an equally large number of research perspectives lie unlimited types of RoE. They may be categorized according to the actors they focus on, the epistemologies they adopt, their theoretical and disciplinary influences, their localized or generalizable concerns, or their research aim in terms of description, exploration, or explanation. Rather than striving for a definitive classification system, which appears impractical and conceptually incoherent—all these potential types would not fit into a single conceptual system—, the field may opt for multiple coexisting typologies, each tailored to different analytical needs. In this light, progress in meta-RoE should not be tied to the revision or expansion of any single classification framework.

4.4 | RoE Contributes to the Theoretical Development of the Evaluation Discipline

This section considers the fourth characteristic of the current concept of RoE, which assumes that RoE is the research program that contributes to the theoretical development of the evaluation discipline. As stated in the introduction, the RoE intellectual community wishes for evaluation to connect its research, theory, and practice into an effective rapport for disciplinary growth and the credibility of its practice. The general assumption is that RoE (understood primarily as research on evaluation practice) generates evaluation theories (understood primarily as evaluation practice theories) which, in turn, informs evaluation practice (see Figure 1).

This section works on the basis of a different set of assumptions that revises the functional relationships between the evaluation discipline and its connection to research, theory, and practice. It argues that RoE cannot fulfill the role of being the generator of evaluation theory in the disciplinary sense of the term and that a complementary research concept, RiE, is necessary. This is, admittedly, an abstract operation. Consulting the visual representation of the revision in Figure 2 can facilitate the reading.

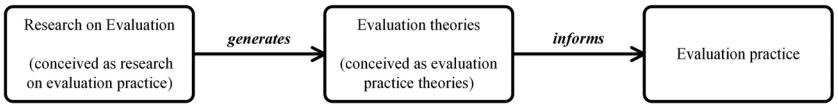

FIGURE 1 | Current understanding of functional relationships between the concepts of research, theory, and practice.

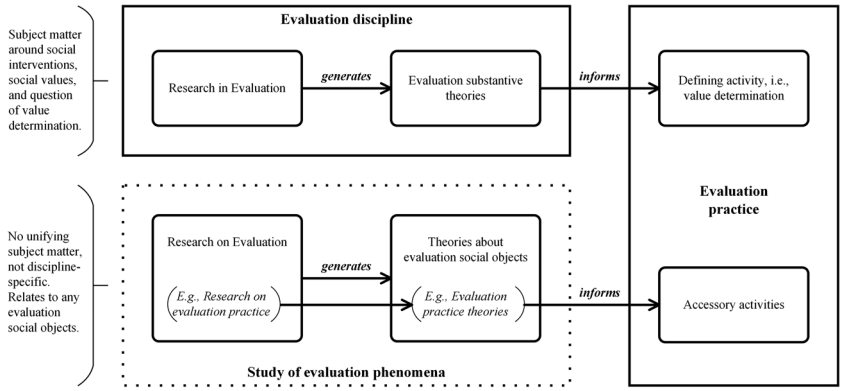

FIGURE 2 | Revised functional relationships between the concepts of discipline, research, theory, and practice.

A scientific discipline has a dual nature. It encompasses, on one hand, the accumulated and systematized knowledge in a domain (i.e., its theories, specific methodologies, and important findings, which is shortened by a reference to "evaluation substantive theories" in Figure 2). On the other hand, it includes the research efforts that feed and regenerate that knowledge (Lakatos 1978; Merton 1973). What anchors this knowledge and research is the *subject matter* of the discipline (i.e., the objects, topics, or ways to investigate them that draw together the intellectual boundaries of the discipline; see the large rectangle labelled "evaluation discipline" in Figure 2).

Theories that match the subject matter of a given discipline are called *substantive theories* for the sake of this demonstration. For example, the subject matter of chemistry relates to the substances, their structure, their properties, and the reactions that change them (see Pauling 1947). Evaluation takes as its distinctive subject matter the value of social interventions and how it is determined (Arbour 2024; Scriven 1991). This inescapably encompasses objects like social interventions and social values, and how they can be integrated into warranted judgments of value. Substantive evaluation theories, logically, are about *that*, not evaluation practice. Research that attends to these substantive matters is therefore the scientific inquiry of the value of social interventions.

By contrast, practice is the "actual application or use of an idea, belief, or method, as opposed to theories relating to it" (Oxford University Press 2019). A practice can be distinguished from others by its defining activity. This defining activity is informed by specific substantive theories. The defining activity of evaluation practice (i.e., the value determination of social interventions) has to be supported by theories aligned with that subject matter (i.e., evaluation substantive theories; Arbour 2024).

Practice itself is a social phenomenon that can be studied, and, incidentally, be the object of its own kind of theory, a theory of practice. The latter covers concepts, explanations, and prescriptions regarding how to engage in specific practical activities (Bourdieu 1977). Compared to a substantive theory, a theory of practice plays an accessory role: it informs activities that support the practice, but do not define it. Such theory informs important activities in evaluation practice (e.g., project management, stakeholder engagement, communication of results, utilization facilitation, or the political functioning of evaluation), but does not inform the defining activity of value determination (Arbour 2020).

The difference between substantive theory and theory of practice should be striking, but has been historically eclipsed in evaluation literature, where evaluation theory is largely assimilated to a prescriptive theory of practice (e.g., Christie 2003; Leeuw and Donaldson 2015; Shadish et al. 1991). Agreeing with this fundamental distinction implies that different research programs are needed for each.

As stated in the previous section, evaluation practice is one of many evaluation phenomena that can be studied under RoE. Research on evaluation practice is therefore a subset of RoE research, the same way that the practice theories it generates coexist with other theories about other evaluation objects (see lower section of Figure 2). These research programs and the theories they generate are not unified by the subject matter of the evaluation discipline but rather share a general object in common (i.e., evaluation in society).

This leaves the research dealing with the subject matter of evaluation—the scientific inquiry of the value of social interventions—without a term. A few terminological options may be considered to fill the gap. First, the evocation of the name of a discipline may suffice to refer to the research activities that correspond to its subject matter. Journals titles like *American Political Science Review* and *Sociology* are examples of this approach. Additionally, the suffix *research* can be attached to distinguish the research component from the discipline as a whole, as in "economic research" or "political science research."

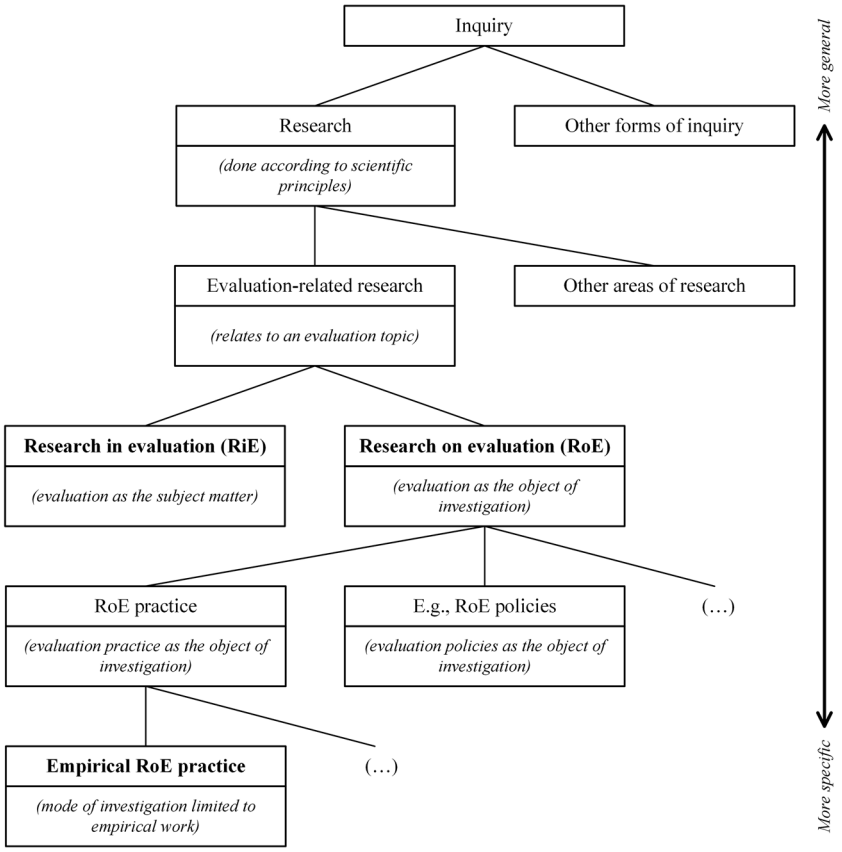

FIGURE 3 | Subdivision of evaluation-related research concepts.

Evaluation could, in principle, adopt this naming convention, with "evaluation research" representing the scientific inquiry of the value of social interventions. Unfortunately, established usage in research methodology has associated it with evaluation projects themselves (Babbie 2017, 364; Clark et al. 2021, 49). The pursuit of overall clarity dictates that the same term is not used for both concepts.

The use of the phrase "research in," resulting in the proposed term RiE, is another viable possibility. This form often suggests (though, admittedly, not systematically) a research program focused on the subject matter of a discipline or, at least, a more unified, coherent research program. This approach gives, for example, a term like *Research in Economics* and the journal of the same name. Another advantage of employing RiE is that its form mirrors that of RoE, thus facilitating comparisons and distinctions between the two concepts.

Ultimately, the distinction between substantive theory and theory of practice lies in their focus: the former centers on the subject matter of the discipline, while the latter attends to the practical user of that knowledge. In this light, the evaluator and their practice are not the subject of evaluation any more than the chemist is the subject of chemistry. If, as Shadish (1998) claimed, "evaluation theory is who we are," it must not be because evaluation theories are about evaluators, but because they concern the defining activity that gives evaluation its identity: the act of evaluating social interventions. The comprehensive development of the field, therefore, necessitates not only the study of the practice through RoE but also the study of the subject matter of evaluation through RiE.

5 | Discussion and Conclusion

This article has systematically distinguished between key concepts around evaluation-related research. Figure 3 summarizes these distinctions using a model of hierarchical generic relations (ISO 704 2022). Each concept is presented with its delimiting characteristic, clarifying how it is distinct from related terms. For ease of comparison, RiE, RoE, and "empirical research on evaluation practice" appear in bold.

Notably, the latter concept aligns with the four key characteristics of RoE outlined at the outset of this article, and thus represents the current understanding of RoE. It is, simultaneously, the most specific concept identified in this analysis and the one most frequently treated as stand-in for the entirety of evaluation-related research. Figure 3 illustrates the conceptual distance between this narrow focus and the full extent of evaluation-related research embodied by the complementary RiE and RoE concepts proposed here.

The RoE community, rightly so, cares about quality research that can support consequential theory for its professional practice and its impact on society. Yet, the conceptual re-examination offered in this article suggests that areas of evaluation-related research could be overlooked and the role of others misunderstood; the

conceptual tools proposed in this article can help to recognize and understand them.

Words matter. They allow practitioners and scholars to carry, organize, and communicate complex ideas. In the end, though, the adoption of terms depends on the acceptance of their semantic community. Some evaluation scholars may choose differently than the terms and definitions suggested in this article. Still, behind words are concepts themselves, and this is where the attempted contribution of this article lies. Greater conceptual clarity about the nature of evaluation-related research may help the field to take proper account of what is already known and more clearly imagine what will be learned next.

Acknowledgments

Open access publishing facilitated by The University of Melbourne, as part of the Wiley - The University of Melbourne agreement via the Council of Australian University Librarians.

References

Adler, M. D. 2015. "Value and Cost-Benefit Analysis." In *The Oxford Handbook of Value Theory*, edited by I. Hirose and J. Olson. Oxford University Press.

Arbour, G. 2020. "Teaching Programme Evaluation: A Problem of Knowledge." *Evaluation and Program Planning* 83: 101872. https://doi.org/10.1016/j.evalprogplan.2020.101872.

Arbour, G. 2024. "Lack of Discipline: Answering Scriven's Call for Value in (program) Evaluation." In *Core Concepts in Evaluation: Classic Writings and Contemporary Commentary*, edited by L. A. Wingate, A. S. Boyce, L. Wilson Becho, and K. N. Robertson, 55–65. Sage Publications.

Aston, R., D. L. Linnell, and C. Westine. 2025. "Editors″ Notes." *New Directions for Evaluation*.

Babbie, E. 2016. *The Practice of Social Research*. Cengage.

Babbie, E. 2017. *The Basics of Social Research*. 7th ed. Cengage.

Blaikie, N. W. H. 2007. *Approaches to Social Enquiry: Advancing Knowledge*. 2nd ed. Polity Press.

Brady, H. E., and D. Collier. 2010. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Rowman & Littlefield Publishers.

Bourdieu, P. 1977. *Outline of a Theory of Practice* (R. Nice, Trans.). Cambridge University Press. https://doi.org/10.1017/CBO9780511812507.

Christie, C. A. 2003. "What Guides Evaluation? A Study of How Evaluation Practice Maps Onto Evaluation Theory." *New Directions for Evaluation, Spring* 2003, no. 97: 7–35.

Clark, T., L. Foster, L. Sloan, and A. Bryman, Social research, m. 2021. *Bryman's Social Research Methods*. 6th ed. Oxford University Press. https://research.ebsco.com/linkprocessor/plink?id=1dfd179a-b128-3c87-9cf1-c738c2de4a75.

Collier, D., and J. Gerring. 2009. *Concepts and Method in Social Science: The Tradition of Giovanni Sartori*. Routledge.

Coryn, C. L. S., L. N. Wilson, C. D. Westine, et al. 2017. "A Decade of Research on Evaluation: A Systematic Review of Research on Evaluation Published Between 2005 and 2014." *American Journal of Evaluation* 38, no. 3: 329–347. https://doi.org/10.1177/1098214016688556.

Crawford, S., and E. Ostrom. 2005. "A Grammar of Institutions." In *Understanding Institutional Diversity*, edited by E. Ostrom. Princeton University Press.

Crotty, M. 1998. *The Foundations of Social Research: Meaning and Perspective in the Research Process*. Sage.

Fournier, D. M. 1995. " Establishing Evaluative Conclusions: A Distinction Between General and Working Logic." *New Directions for Evaluation* 1995, no. 68: 15–32.

Geertz, C. 1973. *The Interpretation of Cultures: Selected Essays*. Basic Books. https://research.ebsco.com/linkprocessor/plink?id=81e0a4a5-5cc5-3522-99f0-9dc8154002d9.

Gerring, J. 2012a. "Mere Description." *British Journal of Political Science* 42, no. 4: 721. https://doi.org/10.1017/S0007123412000130.

Gerring, J. 2012b. *Social Science Methodology: A Unified Framework*. Cambridge University Press.

Gerring, J., and D. Christenson. 2017. *Applied Social Science Methodology: An Introductory Guide*. Cambridge University Press.

Goertz, G. 2006. *Social Science Concepts: a User's Guide*. Princeton University Press.

Goertz, G. 2020. *Social Science Concepts and Measurement: New and Completely Revised Edition*. Princeton University Press.

Henry, G. T., and M. M. Mark. 2003. "Toward an Agenda for Research on Evaluation." *New Directions for Evaluation* 2003, no. 97: 69–80. https://doi.org/10.1002/ev.77.

ISO 704. 2022. *Terminology Work—Principles and Methods*. 4 ed. ISO.

Kuhn, T. S. 2012. *The Structure of Scientific Revolutions*. 4th, 50th anniversary ed. University of Chicago Press.

Lakatos, I. 1978. *The Methodology of Scientific Research Programmes*. Cambridge University Press.

Leeuw, F. L., and S. I. Donaldson. 2015. "Theory in Evaluation: Reducing Confusion and Encouraging Debate." *Evaluation* 21, no. 4: 467–480. https://doi.org/10.1177/1356389015607712.

Li, Z., and C. Westine. 2025. "The Role of Context: A Synthesis of Empirical Research on Evaluation Context." *New Directions for Evaluation* 2025, no. 187: 21–28.

Linnell, D. J., and A. Stachowski. 2025. "Examining Published Research on Evaluation: A Follow-Up of Coryn et al. (2017) From 2015 to 2021." *New Directions for Evaluation*.

Mahner, M. 2007. "Demarcating Science From Non-Science." In *Handbook of the Philosophy of Science: General Philosophy of Science—Focal Issues*, edited by T. Kuipers, 515–575. Elsevier.

Mark, M. M. 2008. "Building a Better Evidence Base for Evaluation Theory." In *Fundamental Issues in Evaluation*, 111–134. Guilford Press.

Merton, R. K. 1973. *The Sociology of Science: Theoretical and Empirical Investigations*. University of Chicago Press.

Neubauer, L., J. LaVelle, A. Boyce, and T. Archibald. 2025. "Considerations for Research on Evaluator Education & Training." *New Directions for Evaluation*.

North, D. C. 1990. *Institutions, Institutional Change and Economic Performance*. Cambridge University Press. https://doi.org/10.1017/CBO9780511808678.

Ogden, C. K., and J. A. Richards. 1946. *The Meaning of Meaning*. Harcourt Brace Jovanovitch.

Ostrom, E. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press. https://doi.org/10.1017/CBO9780511807763.

Oxford University Press. 2019. "Practice." In *Oxford Dictionaries*. https://en.oxforddictionaries.com/definition/practice.

Oxford University Press. (n.d.). "Systematic, adj. & n." In *Oxford English Dictionary*. Retrieved May 22, 2025, from https://doi.org/10.1093/OED/8372227323.

Pauling, L. 1947. *General chemistry; an introduction to descriptive chemistry and modern chemical theory*. W. H. Freeman.

Putnam, H. 2002. *The Collapse of the Fact/Value Dichotomy*. Harvard University Press.

Rawls, J. 1971. *A Theory of Justice*. Belknap Press of Harvard University Press.

Sartori, G. 1970. "Concept Misformation in Comparative Politics." *American Political Science Review* 64, no. 4: 1033–1053. https://doi.org/10.2307/1958356.

Sartori, G. 1984. *Social Science Concepts: A Systematic Analysis*. Sage Publications.

Scriven, M. 1991. *Evaluation Thesaurus*. 4th ed. Sage.

Scriven, M., and J. Roth. 1978. "Needs Assessment: Concept and Practice." *New Directions for Program Evaluation* 1978, no. 1: 1–11. https://doi.org/10.1002/ev.1196.

Sen, A. 1999. *Development as Freedom*. Random House.

Shadish, W. R. 1998. "Evaluation Theory Is Who We Are." *American Journal of Evaluation* 19, no. 1: 1–19. https://doi.org/10.1177/109821409801900102.

Shadish, W. R., T. D. Cook, and L. C. Leviton. 1991. *Foundations of Program Evaluation: Theories of Practices*. Sage.

Toulmin, S. E. 2003. *The Uses of Argument*. 2nd ed. Cambridge University Press. https://doi.org/10.1017/CBO9780511840005.