

Contributing Factors to the Continued Blurring of Evaluation and Research: Strategies for Moving Forward

Ralph Renger

Center for Rural Health, School of Medicine and Health Sciences
University of North Dakota

Abstract: *Despite many studies devoted to the different purposes of evaluation and research, purpose-method incongruence persists. Experimental research designs continue to be inappropriately used to evaluate programs for which sufficient research evidence has accumulated. By using a case example the article highlights several contributing factors to purpose-method incongruence, including the control of the federal level evaluation agenda by researchers, confusion in terminology, and the credible evidence debate. Strategies for addressing these challenges are discussed.*

Keywords: *barriers, credibility, discipline, evaluation, research*

Résumé : *Malgré le grand nombre d'études portant sur les divers objectifs de l'évaluation et de la recherche, une incongruité persiste au niveau des objectifs/méthodes. L'utilisation inappropriée de méthodes expérimentales de recherches pour l'évaluation de programmes déjà abondamment documentés perdure. En présentant un cas, l'article met en lumière plusieurs facteurs qui contribuent à cette incongruité au niveau des objectifs/méthodes, incluant le contrôle du programme fédéral d'évaluation par les chercheurs, la confusion terminologique, et le débat des preuves crédibles. Des stratégies pour s'attaquer à ces défis sont proposées.*

Mots clés : *obstacles, crédibilité, discipline, évaluation, recherche*

BACKGROUND

Each year since 2005 the U.S. Department of Education (DOE) has awarded an average of \$13 million to fund programs such as the Arts in Education (AED) program (DOE, 2013). One goal of the AED program is to document and assess the results and benefits of integrating the arts into school curricula (DOE, 2013). At the 2012 AED evaluation conference, federal evaluators overseeing the AED program insisted grantees employ experimental research designs, preferably the randomized control trial (RCT) design, to evaluate the benefits of the program.

Corresponding author: Ralph Renger, Center for Rural Health, School of Medicine and Health Sciences, University of North Dakota, 501 N. Columbia Road Stop 9037, Grand Forks, ND, USA, 58202-9037; ralph.renger@med.und.edu

The experimental research design requirement was distressing because presenters in conference sessions shared convincing research evidence of the effectiveness of arts integration in school curricula. Further, the AED solicitation for proposals states that one absolute priority of the initiative is to support the evaluation of projects “that are based on research and have demonstrated their effectiveness” (Federal Register, 2010, p. 2523). The decision to take the program to a national scale was based on the accumulation of sufficient research evidence to support the effectiveness of arts integration in school curricula. Why then would there be a need to accumulate more research evidence using costly replication studies?

I became involved in the evaluation when one AED program in a California school district terminated the services of their evaluator. I first conducted a review of the original evaluation plan. The original evaluator was compliant with the AED evaluation requirements, using a randomized nested design (students within schools) to assess program impact. The experimental design involved thousands of students, multiple data collection instruments, and over 50 variables. In short, it included all the criteria necessary for it to score well during the grant review process (Rezmovic, Cook, & Dobson, 1981).

However, my review also revealed the implementation of the evaluation plan was highly problematic. The total annual budget for the project was \$275,000, with an average evaluation budget of approximately \$30,000. Although the budgeted value amount fell within the often cited and recommended 10% (Office of Educational Assessment, 2005), it was grossly insufficient to support an experimental design of the planned scope and magnitude. As a consequence, most of the data collection burden fell on the teachers and the lone data management person for the entire school district. School administrators felt the evaluation plan was taking away teachers’ classroom time and negatively affecting students’ opportunity for learning; arguably an ethical issue and a form of distributive injustice (Bickman & Reich, 2009; Newman & Brown, 1996).

Comparison schools not receiving the arts integration intervention were understandably upset by the investments they were making in contributing to a study with no perceived benefit to their students (personal communication, September 30, 2010). Program staff believed researchers were more concerned with scientific rigour and less concerned about the program, its operation, and its impact on the students. Program staff became frustrated (personal communication, September 30, 2012). The net result was the evaluation plan was not completed with fidelity. Evaluation milestones were not met. Very little data was collected. What data were collected were of limited utility in decision-making.

Analysis of the Case Example

It is reasonable to posit that much of the tension between the California school district and the DOE evaluation plan can be attributed to a disagreement regarding the evaluation purpose. On the one hand, the DOE was clear in its request for experimental designs. The original evaluation plan was successful in meeting

this requirement, as evidenced by the funding decision. However, the California school district had integrated arts education into their curriculum for numerous years, had evidence of its effectiveness, and wanted to focus on tracking whether benefits they valued (e.g., creating a nurturing classroom environment) were being realized. In the view of the California school district, the DOE design was good research but not proper evaluation (Levin-Rozalis, 2003). In their view the evaluation was a hostage of scientific curiosity (Sanders, 1994).

The DOE's experimental design policy suggests the evaluation purpose at the federal level was knowledge development (Mark, Henry, & Julnes, 2000). However, the California school district was interested in evaluating the merit and worth of the program (Mark et al., 2000) and was upset that a research agenda, not stakeholder values, was driving the program evaluation. In the opinion of the school district, sufficient research evidence had already accumulated to show the effectiveness of the AED program on math and science scores (President's Committee on the Arts and Humanities, 2008), and therefore their objectives should take priority.

What is bewildering from an evaluation perspective is why federal evaluators insisted on a research design to evaluate a program for which ample research evidence had accumulated. In my 18 years of conducting local, state, federal, and international program evaluation, this inappropriate application of experimental research methods in program evaluation has been a recurring and troubling problem. It's as though no one is paying attention to the difference between research and evaluation (Beney, 2011; Levin-Rozalis, 2003; Scriven, 1991, 2003; Stufflebeam, 1983; Suchman, 1967).

The purpose of research is to uncover new knowledge (Mark et al., 2000). Evaluation is intended to establish the merit, worth, or value of a program (Julnes & Rog, 2009; Scriven, 1991). The DOE wanted to know whether the AED program was the cause for changes in math and science scores. On the other hand, the California school district wanted to know whether the AED program brought value to the students, teachers, and school district. The values extended far beyond grades and included student personal growth, teacher self-efficacy integrating arts into a curricula, and creating a learning environment valued by students, teachers, and parents.

Evaluation and research are clearly distinct but related (Scriven, 1991). The skills and attributes of the evaluator needed to demonstrate credibility are clearly different in both contexts (Patton, 2008; Scriven, 2003). The pioneers in evaluation such as Suchman, Scriven, Stufflebeam, Mark, and so forth have championed the effort to differentiate evaluation from research. However, it is clear the failure to recognize the distinction between evaluation and research continues to contribute to resource, ethical, and credibility concerns. It is time for the next generation of evaluators to champion this cause, to take responsibility for moving the understanding of the distinction forward, and to ensure the message is no longer ignored. This article postulates a few contributing factors to the problem and offers several strategies for moving the distinction between evaluation and research from paper to practice.

CONTRIBUTING FACTORS TO THE BLURRING BETWEEN EVALUATION AND RESEARCH

I. Federal-Level Evaluation Agenda Driven by Researchers

In the AED case example, the federal evaluators were researchers who assumed the title of evaluator. It stands to reason researchers will prefer methods with which they are trained and familiar. Thus, it is no surprise the federal evaluation of the AED program insisted on experimental designs. This also explains the constant pressure placed on local-level program evaluators to employ research methods (Levin-Rozalis, 2003) and perhaps why the initial AED evaluation plan acquiesced to meeting the federal mandate without considering the values of the school district.

How did researchers obtain these influential positions in the U.S. federal government? In his plenary speech at the Australasian Evaluation Society, Scriven (2013) showed the rise and fall in the status of evaluation over time. He contends that around 1910 the positivist movement toward the value-free doctrine led to a crash in the status of evaluation. In the 1960s, the government's push for accountability sought the advice of researchers and so they [researchers] became part of what Scriven (2013) refers to as the *inner core*. The inner core of researchers is still alive and well today, as evidenced by the federal government reliance on researchers, not evaluators, to set priorities (Donaldson, Christie, & Mark, 2009). Ironically, the problems encountered in the case example are related to a program funded by the DOE, whose 2003 policy to give preference to experimental and quasi-experimental designs triggered the credible evidence debate in our field (Donaldson et al., 2009)

Scriven (2013) notes researchers in the inner core only accept the work of others who share their point of view. Thus, the positivist, value-free doctrine became self-perpetuating. This inner core has largely ignored the criticisms levied against their methods and "begrudgingly acknowledge the need for methodological diversification" (Henry, 2009, p. 33). Although the status of evaluation is steadily recovering, it is still viewed as an intellectual outcast (Scriven, 1991, p. 140). Solutions are needed to address and improve the status of evaluation so evaluators can become valued partners in the inner core, if not overthrow it completely, as Scriven (2013) quipped.

II. The Failure to Distinguish Between Evaluation and Research

There is a clear difference between evaluation and research. Evaluation considers the values of stakeholders and assists in programmatic decision-making (Mark et al., 2000). Research is focused on knowledge development, is driven by the researchers' agenda, and attempts to be value-free (Mark et al., 2000). Yet there is considerable confusion surrounding the terms *evaluation* and *research* (Levin-Rozalis, 2003). At least two reasons for this confusion relate to overlapping terminology and the credible evidence debate.

(i) Overlapping Terminology

In the AED example the federal evaluators saw no problem representing themselves as evaluators although they were researchers. This is not to suggest any deliberate intention to deceive, but at best they were unmindful of the distinction between evaluation and research. Levin-Rozalis (2003) notes some of the confusion in terminology can be attributed to the changing definition of research and evaluation, especially within the social sciences. However, in many ways evaluators must bear some of the responsibility for this confusion. We also tend to be unmindful and to use the terms interchangeably and simultaneously. For example, Hawe and Potvin (2009) note considerable confusion between the terms *intervention research*, *implementation research*, and *evaluation research*. Suchman (1967) used the term evaluation research to depict the idea that judgements can still be made scientifically. Donaldson (2009) uses the terms *basic research* and *applied research* to differentiate between evaluation and research. The former refers to knowledge development and is value-free; the latter focuses on solving practical problems, inherent in which is the need to include stakeholder values.

Using hybrid terms contributes to the confusion and does neither evaluation nor research any good (Levin-Rozalis, 2003). Add to this that often similar designs are used in research and evaluation, albeit for different purposes, and it is not difficult to see how designs needed to develop and confirm knowledge are inappropriately used to try to answer questions about a program's value (Levin-Rozalis, 2003).

(ii) The Credible Evidence Debate

The credible evidence debate is another contributing factor to the blurring between evaluation and research (Donaldson et al., 2009). The debate was supposed to help bring clarity to the differences between the disciplines. However, after attending seminars and workshops at all the major evaluation conferences worldwide, it is my experience the debate is having the opposite effect; it is contributing to the blurring.

As Mark (2009) observed, there really are two debates: one regarding what constitutes credible evidence in knowledge development (i.e., research) and the other as to what constitutes credible evidence in evaluation. Schwandt (2009) suggests one reason for the confusion is that the credible evidence debate is misguided; many researchers and evaluators are not clear there are really two debates. The debate as to what constitutes credible evidence in research has somehow and unnecessarily crept into the program evaluation debate.

This blurring can be partly attributed to how conference seminars devoted to the topic are often presented together under an umbrella of credible evidence. Thus evaluators and researchers, each with different interests and agendas, are brought together at the same time and place. This only serves as a catalyst for confusion: we talk apples and oranges. Within these settings an inordinate amount of time is devoted to the discussion of RCTs as the true design for generating credible

evidence, supporting Mark's (2009) assertion that researchers tend to dominate these discussions. The dominance of the RCT discussion creates the impression experimental research designs are a *fait accompli* when evaluating merit and worth. Of course nothing can be farther from the truth.

III. Failing to Recognize the Shift from Research to Application

It is a well-established tenet in science that the method must meet the purpose (Henry, 2009; Julnes & Rog, 2009; Mark & Henry, 2004). Hawe and Potvin (2009) note many "great failures" in research occurred because of the failure to match the method to the purpose. The existence of purpose-method incongruence creates the sequelae of effects such as those illustrated in the AED program case example.

It is possible one reason for the purpose-method incongruence observed in the AED program case example is a failure to recognize when the purpose has shifted from research to evaluation. One of the important goals of research is establishing cause and effect (Bickman & Reich, 2009). Once there is initial evidence for cause and effect, the focus shifts to accumulating evidence, through replication and generalizability studies (Burman, Reed, & Alm, 2010; Julnes & Rog, 2009). With confidence in the research, the knowledge can be used to design evidence-based programs, policies, and services to meet the goal of social betterment (Feder, 2003; Henry & Mark, 2003; SAMHSA, 2003). This is what Donaldson (2009) would refer to as shifting from basic research to applied research.

The shift from research to evaluation should result in a shift in the methodological approach. But often it doesn't. This is especially evident in professions where both research and service are equally important and professionals are engaged in research and service activities (e.g., public health, psychology, sociology, anthropology). Researchers may fail to realize they have moved along the knowledge-application continuum and continue to apply experimental methods, design, and measurement to the evaluation of programs. Because researchers often advise funders, the shift goes unrecognized at multiple levels (Donaldson et al., 2009). As a result the funding agency, like the DOE, continues to insist on using the same methods, design, and measures to evaluate the program as those used to establish the cause and effect of the program.

STRATEGIES TO ADDRESS THE CONTRIBUTING FACTORS

I. Elevating the Status of Evaluation

Two approaches for elevating the status of evaluation are forwarded. The first is within the evaluation discipline's direct control and takes the needed steps to move our discipline toward a profession. The second is an indirect approach, which acknowledges the contributions of research and frames arguments using research principles for creating movement in the research community toward understanding, accepting, and valuing the contributions of evaluation.

(i) Direct Approach: Moving from a Discipline to a Profession

When the competence of one evaluator is questioned, it affects our profession (Mayer, 2008). More countries, led by their respective evaluation associations, need to endorse Canada's lead in developing credentialing boards like the one managed by the Canadian Evaluation Society (CES, 2013). Further, standards of acceptability within the certification-credentialing process should include the ability to (a) articulate the difference between evaluation and research, (b) understand the consequences when there is purpose-method incongruence, and (c) apply this knowledge in practice.

Having credentialed evaluators will be a major step forward in changing the image of evaluation from an intellectual outcast to a respected profession. It will also pave the way for placing evaluators in positions of influence within government, so realistic, meaningful, and appropriate solicitations for proposals will be developed. Credentialing and certification are necessary first steps for realizing Scriven's (2013) vision of becoming an alpha discipline (i.e., in charge of quality control as a subject), replacing latent positivists (i.e., those who only accept and do value-free studies), becoming the exemplar discipline (i.e., the central core of applied social science), and eventually dealing with the alpha value of ethics (i.e., understanding and respecting value systems).

The Canadian credentialing process has been slow and not without its problems (Brousselle, 2012). One problem is researchers are in control of defining the evaluator skill set (Altschuld, 2005). The evaluation discipline needs to move forward and be comfortable knowing our membership may be without researchers who are unwilling to adapt their approaches to meet evaluation purposes.

More evaluation-specific programs are needed to develop credentialed evaluators. Programs such as those at Claremont Graduate University (Claremont Graduate University, 2013) are still in short supply. We need to compete with research programs for rising stars and make students aware of the viability of evaluation as a career option.

(ii) Indirect Approach: Educating the Research Community

Given the glacial pace at which our discipline is moving toward certification and credentialing, other strategies are needed to elevate the status of evaluation. If knowledge is power, then perhaps educating researchers about evaluation is a way forward. However, because researchers are in the inner core and self-perpetuate their values, they may not be open to accepting arguments that evaluation is a valued discipline. In my experience a more indirect approach is needed when trying to educate researchers.

Framing arguments using the purpose-method tenet has proven very useful. The importance of purpose-method congruence is well documented in the research literature (Henry, 2009; Julnes & Rog, 2009). Undergraduate and graduate research courses repeatedly discuss the pitfall of not having a toolbox of methods as captured by the adage "It is tempting, if the only tool you have is a hammer, to treat everything as if it were a nail" (Maslow, 1966). The key to opening the

discussion is to first remind researchers of this tenet. Once this is accepted, then a meaningful discussion about how evaluation and research purposes differ can occur. With that understanding, the discussion about appropriate methods can then occur without any perceived threat.

However, there will still be those researchers who insist designs needed to isolate cause and effect are appropriate when conducting a program evaluation: they genuinely believe they have purpose-method congruence. These researchers must be challenged to provide a reasonable defence of why there is a need to isolate variables and create contrived closed systems through methods such as randomization when there is already ample research evidence to support the expected benefits of the program. Most problems encountered in a program evaluation environment are in fact not solvable by research and statistical approaches (Rezmovic et al., 1981; Victora, Habicht, & Bryce, 2004).

In my experience researchers are often unaware of the consequences of the purpose-method incongruence as it relates to evaluating programs. This is because, by its very nature, knowledge inquiry is guided by the researcher and aims to be value-free (Scriven, 2013). As there is no need to engage stakeholders, the consequences of not doing so are not realized; it is an error of omission. Focusing on the consequences of purpose-method incongruence with researchers is often helpful in (a) making clear the importance of distinguishing between evaluation and research, and (b) elevating the value of evaluation methods. Some of these consequences are now presented.

Possible Consequence 1: Low Project Morale

In the AED case example, program staff assumed the evaluation responsibilities and added responsibilities created by the demand for a RCT design that fell outside their job description. The staff had no desire or time to engage in these added responsibilities (Rezmovic et al., 1981). When evaluation responsibilities take away from primary responsibilities, staff become overwhelmed and lose morale.

Another factor affecting morale relates to the nature of research-based protocols. These protocols can be daunting and strict (Rezmovic et al., 1981). From a research perspective strict adherence to an implementation protocol is essential to isolating cause and effect: everything other than the independent variable needs to be controlled. Of course this is counterintuitive to program evaluation, where ongoing corrective actions are welcomed to ensure the highest quality delivery of the program given the changing circumstances (Rezmovic et al., 1981; Victora et al., 2004). Not being able to make necessary changes to ensure participants receive the highest quality services disheartens program staff (Mayer, 2008). It also affects participants, who are more likely to drop out if the program does not adjust to meet their needs.

Possible Consequence 2: Fiscal Irresponsibility

Evaluators have an obligation to be stewards of taxpayer dollars. Insisting on experimental and quasi-experimental designs to reaffirm a well-established

knowledge base is irresponsible. Such designs and level of rigour are unnecessary when the focus is on evaluating what stakeholders value and whether expected outcomes, based on research, have been met.

Sample size requirements and the cost associated with trying to maintain the integrity of a research design in the field affect staff morale and produce unusable reports (Rezmovic et al., 1981). Service programs often target a specific population demographic, serve a small number of participants, or are of insufficient duration to be able recruit the needed number of participants over time to satisfy sample size requirements. For example, the Seven Generations Center of Excellence (SGCoE, 2013) in the Center for Rural Health at the University of North Dakota has numerous activities designed to address the shortage of American Indian behavioural health professionals. For the SGCoE, placing even five American Indian behavioural health professionals on reservations would be a monumental success. Yet many programs are burdened by sample size requirements when there are other methods, such as single-subject designs, to produce reliable information upon which to base decisions (Shadish & Rindskopf, 2007). Programs serving a relatively small and/or targeted population are set up to fail if required to use experimental research methods to evaluate their effectiveness.

Possible Consequence 3: Ethical Concerns

Many programs have small, finite budgets (Rezmovic et al., 1981). Programs want to serve the maximum number of people with their budget. Shifting budget dollars away from services to support the accumulation of additional, unnecessary research evidence becomes an ethical concern.

Another ethical concern pertains to the withholding of benefit and cost-burden to those assigned to control conditions (Bickman & Reich, 2009). For example, in the AED program, there were no plans and no resources to provide the benefits of the arts integration to the students participating in the control schools. The control group was exploited for the purpose of replicating past research findings.

Possible Consequence 4: The Failure to Produce Usable Evaluation Reports

In the AED example the school district did not see the relevance of the federal program evaluation requirements. For example, stakeholders struggled with what a $p < .05$ meant for decision-making. As Levin-Rozalis (2003) put it,

[w]hat happens is that the research apparatus gives highly generalizable, abstract answers appropriate to research: valid answers gathered with the aid of highly replicable research tools. But the quality of the evaluation is damaged because the operators of the project do not get answers relevant to their own work—answers that are directly related to the different activities, audiences, and questions of the project. (p. 7)

The AED evaluation plan failed to provide the stakeholders with usable data upon which to make decisions (Sanders, 1994). As a result data were not collected and the evaluation reports were either not completed or shelved (Patton, 2008).

II. Addressing Overlapping/Confusing Terminology

The terms evaluation and research must be kept distinct. Using hybrid terms only adds to the confusion (Levin-Rozalis, 2003). There is consensus that the purpose of research is knowledge development (Mark et al., 2000) and the purpose of evaluation is to provide information to assist stakeholders in making decisions of value to them (Renger, Bartel, & Foltysova, 2013)—to improve, not prove (Stufflebeam, 1983). The Integrated Theory of Evaluation nicely differentiates and recognizes the importance of each of these purposes and is a framework I have successfully used over the last decade to differentiate evaluation purposes and ensure purpose-method congruence (Julnes & Rog, 2009; Mark et al., 2000).

With respect to the contribution of the credible-evidence debate to the confusion between evaluation and research, Mark (2009) noted these are really two separate debates. Therefore, one strategy is to hold two separate discussions: one discussion for what constitutes credible evidence in evaluation, the other discussion for what constitutes credible evidence in research. Creating two forums will send a clear message that there is a distinction. It also creates a safe place for evaluators to discuss credible evidence for evaluating merit and worth, while not being dominated or pressured by researchers (Levin-Rozalis, 2003; Mark, 2009).

III. Providing a Conceptualization that Recognizes the Continuum of Evaluation to Research

One way I have sharpened the distinction between evaluation and research is to discuss them in terms of a research-application continuum, an idea consistent with SAMHSA's (2003) "from science to service" motto. Doing so is less threatening to researchers; they can clearly see their place and importance on the continuum. Research on the one end is focused on knowledge development. Once this knowledge is replicated and validated it is used to develop programs (Henry, 2009; Julnes & Rog, 2009). Programs are a way of bringing this knowledge to the public for social betterment (Mark et al., 2000).

SUMMARY

The distinction between evaluation and research is well articulated in the evaluation literature by many of our discipline's leaders. However, the misapplication of experimental research methods and designs in the evaluation context continues. The unnecessary accumulation of knowledge using replication studies must be stopped. It is fiscally irresponsible to taxpayers, unethical to program participants, and demoralizing to staff, and it does not produce usable results for program administrators.

One contributing factor to the problem at hand is the evaluation agenda is set by researchers. Scriven's observation decades ago that evaluation is an intellectual outcast still holds true today. Becoming credentialed and certified is an important step in elevating the status of our discipline. However, this is a time-consuming top-down approach that is impeded by some researchers.

An ongoing educational effort is needed. For this educational effort to be successful, researchers' perceived threat of evaluation must be calmed. This can be accomplished by (a) focusing on the purpose-method tenet of science, (b) reinforcing the importance of research as the foundation for evidence-based programs, and (c) highlighting the harmful consequences of purpose-method incongruence.

The next generation of evaluators needs to champion the fight in elevating the status of our discipline. We need to take every opportunity to educate government, researchers, funding agencies, and, last but not least, evaluators on the importance of distinguishing between research and evaluation. We must speak with a common language, using agreed-upon definitions of evaluation and research, and avoid using hybrid terms that contribute to the confusion. Finally, ongoing discussions as to what constitutes credible evidence at evaluation conferences must remain focused on evaluation, not research.

REFERENCES

- Altschuld, J. W. (2005). Certification, credentialing, licensure, competencies, and the like: Issues confronting the field of evaluation. *Canadian Journal of Program Evaluation*, 20(2), 157–168.
- Arts in Education. (2012). *Arts in Education Model Development and Dissemination Program evaluation workshop*. Arlington, Virginia.
- Beney, T. (2011). Distinguishing evaluation from research. Retrieved from <http://www.uniteforsight.org/evaluation-course/module10>
- Bickman, L., & Reich, S. M. (2009). Randomized controlled trials: A gold standard with feet of clay? In S. I. Donaldson, C. A. Christie, & M. M. Mark (Eds.), *What counts as credible evidence in applied research and evaluation practice?* (pp. 51–77). Los Angeles, CA: Sage. <http://dx.doi.org/10.4135/9781412995634.d10>
- Brousselle, A. (2012, May). *Why I will not become a credentialed evaluator*. Presentation at the annual conference of the Canadian Evaluation Society, Halifax, Nova Scotia.
- Burman, L. E., Reed, W. R., & Alm, J. (2010). A call for replication studies. *Public Finance Review*, 38(6), 787–793. <http://dx.doi.org/10.1177/1091142110385210>
- Canadian Evaluation Society. (2013). Canadian Evaluation Society professional designations program. Retrieved from <http://www.evaluationcanada.ca/site.cgi?en:5:6>
- Claremont Graduate University. (2013). The Claremont Evaluation Center. Retrieved from <https://sites.google.com/site/sbosjobs/>
- Department of Education. (2013). Arts in education—Model development and dissemination grants program. Retrieved from <http://www2.ed.gov/programs/artsedmodel/index.html>
- Donaldson, S. I. (2009). In search of the blueprint for an evidence-based global society. In S. I. Donaldson, C. A. Christie, & M. M. Mark (Eds.), *What counts as credible evidence in applied research and evaluation practice?* (pp. 2–18). Los Angeles, CA: Sage. <http://dx.doi.org/10.4135/9781412995634.d6>

- Donaldson, S. I., Christie, C. A., & Mark, M. M. (2009). *What counts as credible evidence in applied research and evaluation practice?* Los Angeles, CA: Sage.
- Feder, J. (2003). Why truth matters: Research versus propaganda in the policy debate. *Health Services Research, 38*(3), 783–787. <http://dx.doi.org/10.1111/1475-6773.00146>; Medline:12822912
- Federal Register. (2010). Department of Education, Office of Innovation and Improvement; Overview Information; Arts in Education Model development and Dissemination Program; Notice inviting New Applications for New Awards for Fiscal year (FY) 2010. *Federal Register, 75*(10), 2523.
- Hawe, P., & Potvin, L. (2009). What is population health intervention research? *Canadian Journal of Public Health, 100*(1), I8–I14. Medline:19263977
- Henry, G. T. (2009). When getting it right matters: The case for high quality policy and program impact evaluations. In S. I. Donaldson, C. A. Christie, & M. M. Mark (Eds.), *What counts as credible evidence in applied research and evaluation practice?* (pp. 32–50). Los Angeles, CA: Sage. <http://dx.doi.org/10.4135/9781412995634.d9>
- Henry, G. T., & Mark, M. M. (2003). Beyond use: Understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation, 24*(3), 293–314.
- Julnes, G., & Rog, D. (2009). Evaluation methods for producing actionable evidence: Contextual influences on adequacy and appropriateness of method choice. In S. I. Donaldson, C. A. Christie, & M. M. Mark (Eds.), *What counts as credible evidence in applied research and evaluation practice?* (pp. 96–132). Los Angeles, CA: Sage. <http://dx.doi.org/10.4135/9781412995634.d12>
- Levin-Rozalis, M. (2003). Evaluation and research: Differences and similarities. *Canadian Journal of Program Evaluation, 18*(2), 1–31.
- Mark, M. (2009). Credible evidence: Changing the terms of the debate. In S. I. Donaldson, C. A. Christie, & M. M. Mark (Eds.), *What counts as credible evidence in applied research and evaluation practice?* (pp. 214–238). Los Angeles, CA: Sage. <http://dx.doi.org/10.4135/9781412995634.d20>
- Mark, M. M., & Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation, 10*(1), 35–57. <http://dx.doi.org/10.1177/1356389004042326>
- Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs*. San Francisco, CA: Jossey-Bass.
- Maslow, A. H. (1966). Acquiring knowledge of a person as a task for the scientist. *The psychology of science: A reconnaissance*. Chapel Hill, NC: Maurice Bassett.
- Mayer, A. M. (2008). *Trust in the evaluator-client relationship* (Unpublished dissertation). University of Minnesota.
- Newman, D. L., & Brown, R. D. (1996). *Applied ethics for program evaluation*. Thousand Oaks, CA: Sage.
- Office of Educational Assessment. (2005). *Frequently asked questions*. Seattle, WA: University of Washington. Retrieved from http://www.washington.edu/oea/services/research/program_eval/faq.html
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: Sage.

- President's Committee on the Arts and Humanities. (2008). Re-investing in arts education: Winning America's future through creative schools. Retrieved from <http://www.pcah.gov/sites/default/files/photos/PCAH%20Report%20Summary%20and%20Recommendations.pdf>
- Renger, R., Bartel, G., & Foltysova, J. (2013). The reciprocal relationship between implementation theory and program theory in assisting decision-making. *Canadian Journal of Program Evaluation, 28*(1), 27–41.
- Rezmovic, E. L., Cook, T. J., & Dobson, L. D. (1981). Beyond random assignment: Factors affecting evaluation integrity. *Evaluation Review, 5*(1), 51–67. <http://dx.doi.org/10.1177/0193841X8100500103>
- SAMHSA. (2003). From science to service: Making a model program. *SAMSHA News, 9*(1). Retrieved from http://www.samhsa.gov/samhsa_news/volumexi_1/article6.htm
- Sanders, J. R. (1994). *The program evaluation standards: How to assess evaluations of educational programs. The Joint Committee on Standards for Educational Evaluation*. Thousand Oaks, CA: Sage.
- Schwandt, T. A. (2009). Toward a practical theory of evidence for evaluation. In S. I. Donaldson, C. A. Christie, & M. M. Mark (Eds.), *What counts as credible evidence in applied research and evaluation practice?* (pp. 197–212). Los Angeles, CA: Sage. <http://dx.doi.org/10.4135/9781412995634.d18>
- Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: Sage.
- Scriven, M. (2003). Reflecting on the past and future of evaluation. *The Evaluation Exchange, 9*(4). Retrieved from <http://www.hfrp.org/evaluation/the-evaluation-exchange/issue-archive/reflecting-on-the-past-and-future-of-evaluation/michael-scriven-on-the-differences-between-evaluation-and-social-science-research>
- Scriven, M. (2013, September). *The past, present, and future of evaluation*. Paper presented at the international conference of the Australasian Evaluation Society, Brisbane, Australia.
- Seven Generations Center of Excellence. (2013). The Seven Generations Center of Excellence in Native Behavioral Health, The Center for Rural Health, School of Medicine and Health Sciences, The University of North Dakota. Retrieved from <http://ruralhealth.und.edu/projects/coe-in-native-behavioral-health>
- Shadish, W. R., & Rindskopf, D. M. (2007). Methods for evidence-based practice: Quantitative synthesis of single-subject designs. In G. Julnes & D. J. Rog (Eds.), *Informing federal policies on evaluation methodology: Building the evidence base for method choice in government sponsored evaluation* (pp. 95–109). San Francisco, CA: Jossey-Bass. <http://dx.doi.org/10.1002/ev.217>
- Stufflebeam, D. L. (1983). The CIPP model for program evaluation. In G. F. Madaus, M. Scriven, & D. L. Stufflebeam (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation* (pp. 117–141). Boston, MA: Kluwer Nijhof.
- Suchman, E. A. (1967). *Evaluative research: Principles and practice in public service and social action programs*. New York, NY: Russell Sage Foundation.
- Victora, C. G., Habicht, J. P., & Bryce, J. (2004). Evidence-based public health: Moving beyond randomized trials. *American Journal of Public Health, 94*(3), 400–405. <http://dx.doi.org/10.2105/AJPH.94.3.400>; Medline:14998803

AUTHOR INFORMATION

Ralph Renger, PhD, is a professor in the School of Medicine and Health Sciences at the University of North Dakota where he evaluates projects within the Center for Rural Health (CRH).