

<http://dx.doi.org/10.18222/ee.v30i73.6229>

# TRENDS AND CHALLENGES IN PROGRAM EVALUATION: AN INTERVIEW WITH CHRIS CORYN

ADRIANA BAUER<sup>I</sup>

FABIANA SILVA FERNANDES<sup>II</sup>

## ABSTRACT

*In this interview, Chris Coryn, Director of the Interdisciplinary PhD in Evaluation (IDPE) Program, Professor of Evaluation at Western Michigan University (WMU) and executive editor of the Journal of MultiDisciplinary Evaluation, explains key concepts and characteristics of program evaluation, a dimension of evaluation that is still unexplored in the educational area in Brazil. The researcher comments on some of the current challenges in the field of program evaluation, highlighting methodological tendencies and ethical issues in professional practice. He also discusses knowledge, competences and skills that should be considered in the training of the professional evaluator. Coryn criticizes current evaluative approaches that disconnect program evaluation from values and valuing, which, in his view, are distinctive elements of program evaluation.*

**KEYWORDS** PROGRAM EVALUATION • EVALUATIVE RESEARCH DESIGN • EVALUATOR TRAINING • EVALUATION APPROACHES.

<sup>I</sup> The Carlos Chagas Foundation (FCC) and the University of São Paulo (USP), São Paulo-SP, Brazil; <http://orcid.org/0000-0002-5942-9181>; [adbauer@fcc.org.br](mailto:adbauer@fcc.org.br)

<sup>II</sup> The Carlos Chagas Foundation (FCC), São Paulo-SP, Brazil; <https://orcid.org/0000-0002-3458-9963>; [fsfernandes@fcc.org.br](mailto:fsfernandes@fcc.org.br)

## TENDÊNCIAS E DESAFIOS NA AVALIAÇÃO DE PROGRAMAS EDUCACIONAIS: ENTREVISTA COM CHRIS CORYN

### RESUMO

Nesta entrevista, Chris Coryn, diretor do Programa de Doutorado Interdisciplinar em Avaliação, professor de Avaliação na Western Michigan University e editor executivo do *Journal of MultiDisciplinary Evaluation*, explica conceitos e características essenciais da avaliação de programas, dimensão da avaliação ainda pouco explorada na área educacional no Brasil. O pesquisador comenta sobre alguns dos desafios atuais do campo da avaliação de programas, destacando tendências metodológicas e questões éticas recorrentes na prática avaliativa profissional; discute sobre conhecimentos, habilidades e competências que deveriam ser consideradas na formação do avaliador profissional; e critica abordagens avaliativas atuais que desconectam a avaliação de programas da atribuição de valor e da discussão sobre valores e valoração que, ao seu ver, são elementos distintivos da avaliação de programas.

**PALAVRAS-CHAVE** AVALIAÇÃO DE PROGRAMAS • DESENHOS DE PESQUISA AVALIATIVA • FORMAÇÃO DO AVALIADOR • ABORDAGENS AVALIATIVAS.

## TENDENCIAS Y DESAFÍOS EN LA EVALUACIÓN DE PROGRAMAS EDUCATIVOS: ENTREVISTA CON CHRIS CORYN

### RESUMEN

En esta entrevista Chris Coryn, director del Programa de Doctorado Interdisciplinario en Evaluación, profesor de Evaluación en la Western Michigan University y editor ejecutivo del *Journal of Multidisciplinary Evaluation*, explica conceptos y características esenciales de la evaluación de programas, dimensión de la evaluación todavía poco explorada en el área educacional en Brasil. El investigador comenta algunos de los actuales desafíos del campo de la evaluación de programas, destacando tendencias metodológicas y temas éticos recurrentes en la práctica evaluativa profesional; discute conocimientos, habilidades y competencias que habría que considerar en la formación del evaluador profesional; y critica abordajes evaluativos actuales que desconectan la evaluación de programas de atribución de valor y de la discusión sobre valores y valoración que, en su opinión, son sus elementos distintivos.

**PALABRAS CLAVE** EVALUACIÓN DE PROGRAMAS • DISEÑOS DE INVESTIGACIÓN EVALUATIVA • FORMACIÓN DEL EVALUADOR • ABORDAJES EVALUATIVOS.

## PRESENTATION

Chris L. S. Coryn is Director of the Interdisciplinary PhD in Evaluation (IDPE) Program and Professor of Evaluation at Western Michigan University (WMU). He received a B.A. in Psychology and a Master's degree in Social and Experimental Psychology, both from Indiana University. His Ph.D. in Philosophy of Assessment from Western Michigan University was guided by Michael Scriven, a leading theorist whose contributions to the development of program evaluation are numerous. With extensive experience in the field of evaluation, Coryn's research interests include history of evaluation, research on evaluation theories, quantitative, qualitative and mixed research methods, measurement theories and psychometrics, and meta-analysis. His academic career is marked by numerous studies and evaluations in various fields, including education, science and technology, health and medicine, community and international development, and social and human services. He has authored or co-authored more than one hundred research reports. He lectures, speaks and gives workshops in the United States and Canada, as well as in other countries in Europe, Asia and Oceania. His academic production is extensive, with more than ninety contributions among scientific

articles, entries for specialized encyclopedias, chapters and books focusing on conceptual, theoretical and methodological discussions on program evaluation. For his contribution to the fields of research and evaluation he has already been awarded four prizes, including the American Educational Research Association (AERA) Research on Evaluation Distinguished Scholar Award. Along with Daniel Stufflebeam (*in memoriam*), he wrote *Evaluation Theory, Models and Applications*, published by Jossey-Bass in 2014.

**EAE:** Considering the Brazilian context, in the educational field it seems more common to research public policy using a framework of policy analysis than using program evaluation theories and models. What are the main differences and the common principles, if any, between these research fields? Additionally, could you provide us with a definition of evaluation (in the program evaluation context) that you consider good/adequate?

**CC:** *Policy analysis and evaluation are often used synonymously, and the distinctions between the two are not, necessarily, methodological. Both are concerned with similar objects of investigation. However, policy analysis is typically concerned only with questions of ‘what’s so?’ (i.e., descriptive, fact-based conclusions) and does not usually address questions of ‘so what?’ (i.e., evaluative, value-based conclusions). Value, values, and valuing are the central features that distinguish evaluation from related disciplines, including policy analysis and applied research, among many others (e.g., assessment). By definition – and, therefore, given that the root of the term evaluation is ‘valu’ – evaluation is the process of determining the merit, worth, or significance of something, or the product of that process. This definition is neither too broad nor too narrow and is sufficient for encompassing most types of evaluation and objects of evaluation (e.g., programs, policies, personnel).*

**EAE:** In Brazil, policies and programs are marked by ruptures and significant changes in their designs due to power changes in government. In your point of view, how to conduct program evaluations considering the challenge imposed by this political and administrative context of federative entities?

**CC:** *Politics always have been, and likely always will be, a major inhibitor to evaluation. Without delving into the specifics of politics and the implications of politics on programs, the simple answer is that evaluation (and evaluators) should be flexible and adapt to changing contexts and political forces. Rigid, fixed evaluation designs almost always fail.*

**EAE:** Evaluation rises in a scenario of public accountability. In Brazil, the discussion about the relations between evaluation and accountability are very controversial. How do you understand those relations?

**CC:** *Accountability is one function – or purpose – of evaluation. Accountability is simply to “be held accountable or responsible for;” which could include results (or lack thereof), expenditures, activities, and ultimately, transparent disclosure (including evaluations). So, in some respects, accountability is an important aspect of evaluation. But, not the only one.*

**EAE:** How would you summarize, in theoretical terms, the main contemporary trends and movements in program evaluation?

**CC:** *At this point in the brief history of professional evaluation, the field is rife with a variety of ‘fads’ and ‘trends’ that are not necessarily beneficial to the development of the discipline or profession. These include, for example, the various social justice-oriented approaches, systems thinking and systems theory, theory-driven, and many, many others. Fundamentally, none of these concepts are unique to evaluation, and are pervasive in a variety of similar fields and disciplines. They do, however, share one common feature: none are explicitly concerned with determining merit, worth, or significance; or, again, value, values, and valuing. Some of the popular approaches currently in vogue are misguided, in my humble opinion, given biases disguised as advocacy or approaches that are overly concerned with explaining phenomenon, for instance. These concerns are legitimate and worthy, but as previously mentioned, all miss the point. Evaluation is, first and foremost, concerned with quality (merit) and value (worth). Consider the underlying implications of the various systems and theory-oriented approaches. Put simply, it is of exceptional difficulty to merely determine whether something works (i.e., simple causal description), or produces beneficial (or harmful) effects, let alone attempt to explain how, for whom, and under what conditions (i.e., causal explanation)! We are not in agreement about how to go about the latter, let alone the former.*

**EAE:** You serve as on the Editorial Advisory Boards of two important journals on evaluation field: *American Journal of Evaluation* and *Evaluation and Program Planning*. You always contribute to the field with books reviews on evaluation and research designs (among other topics). You are executive editor of a specialized journal on evaluation. This experience allows you to know very well the technical and scientific production on program evaluation. In your constant review of the literature on program evaluation, do you see difficulties and failures of authors in conducting evaluative research? If so, which ones would you highlight?

**CC:** *Evaluation reports are typically not published in peer reviewed journals and are not considered scholarly inquiry. Such reports are considered an essential element of practice, but not a means for increasing or contributing to knowledge regarding evaluation. Research ON evaluation theories, methods, and practices, however, has increased dramatically in the past few decades. This type of inquiry has been defined as “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).<sup>1</sup> Much of the current literature published in the field’s leading journals, however, consists of ‘how to’ (e.g., how to use propensity score analysis) or ‘applications of’ (e.g., an application of transformative evaluation in XX context) types of articles rather than empirical research on evaluation theories, methods, or practices that could potentially move the discipline forward. Although ‘how to’ and applications of’ articles are valuable, they do not empirically examine the claims put forth in theoretical approaches to confirm or disconfirm their use in practice (e.g., does empowerment evaluation actually result in empowerment?). As one prominent evaluation scholar once stated: “Shouldn’t evaluation itself be open to systematic inquiry, just as the policies, programs, and practices that we evaluate are?”*

**EAE:** Coined by Scriven, the concepts of formative and summative evaluation are already well-known in the evaluation field. Often such concepts are seen as incompatible purposes in the practice of program evaluation: an evaluation should have a purpose more related to program improvement or accountability. What are the distinctive features between summative and formative evaluation in the program evaluation practice? How formative and summative purposes can be imbricated in the cycle of public policies?

**CC:** *At their essence, the concepts of formative and summative evaluation are simply the functions or purposes associated with the two concepts. The former is explicitly improvement-oriented, whereas the latter is essentially oriented toward accountability and decision making. Both, however, involve decisions and it is the use of evaluative information in that process that determines the function or purpose of an evaluation (i.e., how an evaluation is actually used). A formative evaluation conducted with the intent of improving a program or policy, for example, can result in a decision to terminate*

**1** CORYN, C. L. S.; WILSON, L. N.; WESTINE, C. D.; HOBSON, K. A.; OZEKI, S.; FIEKOWSKY, E. L.; GREENMAN II, G. D.; SCHRÖTER, D. C. A decade of research on evaluation: A systematic review of research on evaluation published between 2005 and 2014. *American Journal of Evaluation*, v. 38, n. 3, p. 329-347, 2017.

*the evaluated program or policy (i.e., a summative decision). Thus, the clear functional purposes of formative and summative are sometimes simple, though both clearly depend on how an evaluation is actually used, rather than an intended use. Even so, the formative and summative distinctions are sufficient for classifying the majority of evaluation foci, despite some claims that evaluation is more than simply formative or summative (e.g., developmental, ascriptive, learning).*

**EAE:** Other main issue in program evaluation seems to be the definition of judgment criteria. If to evaluate is to attribute value and involve judging an evaluand, is there some better criteria to define how it should be judge? In your views, what are the current main theoretical perspectives on the valuing problem?

**CC:** *Valuing is, perhaps, the great dilemma in program evaluation and there is no ‘correct’ or ‘simple’ answer to how values should be incorporated into an evaluation. Even so, Scriven’s logic of evaluation provides a solid foundation for addressing this problem. This logic and pattern of reasoning consists of (1) identify criteria (i.e., these are the ‘values’ against which an object is to be evaluated [though somewhat of an oversimplification]), (2) determine standards of performance for each criterion, (3) measure performance and compare to standards, and (4) synthesize into an evaluative conclusion of merit and worth. Sources of values (i.e., criteria) are a constant challenge and these values can be derived in many, many varying ways (see Scriven’s Key Evaluation Checklist as a reference for sources of values/criteria). Values – or criteria – are not necessarily program goals or objectives, though these should be seriously considered. Recipient needs are a central source of values. As regards needs as a source of criteria or values for judging programs, instrumental needs should be distinguished from needs ‘treatment’ needs (as well as ‘wants’).*

**EAE:** Nowadays, some approaches to program evaluation are concerned with enhance utilization to improve the program or, yet, to support developmental decision making. Those approaches may challenge the role usually attributed to evaluators since sometimes the evaluator becomes involved in the process of decision-making and program management itself. How do you understand the role of the evaluator in today’s evaluation practice, considering your views about the best approaches for 21<sup>st</sup> evaluations?

**CC:** *The recent trend of evaluator as a program stakeholder or decision maker – largely arising from Patton’s notion of developmental evaluation – is troubling. After all,*

*numerous evaluation scholars (Patton included) have long asserted that such activities are likely to create various forms of bias and severely inhibit one's ability to provide unbiased, impartial evaluations. Arguably, evaluators have important skills and knowledge that are relevant to the design and administration of programs. Even so, such persons should not evaluate programs in which they have a vested interest or clear conflict of interest.*

**EAE:** The nature of evaluation can be very controversial. Some might argue that evaluation is an applied research, based on knowledge and procedures of other disciplinary areas. Others claim that evaluation is a disciplinary field itself, with its own theoretical reasoning. Michel Scriven understands evaluation as a transdiscipline. One of your research interests is research on evaluation theories. So, how would you define the nature of evaluation and, more specifically, the nature of program evaluation? What are evaluation theories? Does research on evaluation theory matter?

**CC:** In some respects, the debates about the disciplinary nature of evaluation are irrelevant to the actual practice of evaluation and really only matters to evaluation scholars. Whether evaluation is applied social science research is, like definitions of evaluation, not agreed upon; though I strongly disagree. Applied research, on one hand, is research intended to provide a solution to a perceived problem. Evaluation, on the other hand, is concerned with whether the solution is meritorious and worthwhile.

Evaluation theories are simply prescriptions regarding how evaluation ought to be practiced. They are not explanatory or predictive theories as understood by the colloquial meaning and usage of the term. Given that such theories are prescriptive, it is extremely important to empirically investigate them in order to ascertain the extent to which theoretical claims are either supported or refuted as well as adhered to in evaluation practiced (see Miller and Campbell's investigation of empowerment evaluation and Coryn et al.'s investigation of theory-driven evaluation).

**EAE:** There is a controversial debate not only in Brazil but also internationally about the use of Randomized Control Trials (RCT) as the golden standard for impact evaluation. Could you please explain your views about the positive and negative aspects around RCTs and other evaluation designs which rely on evidence-based practice?



**CC:** *The arguments in support of and against the use of RCTs is well-documented and the answer is nearly always “it depends”. Although many resist the use of RCTs for evaluation purposes, there are instances where such designs are the appropriate choice. Pharmaceutical trials are clear instance of appropriate application of RCTs. In the case of social and other types of interventions into human affairs, the answer is not so clear. That being said, RCTs answer only one type of question, namely ‘causal.’ In my experience, both those who advocate either for or against RCTs often have no true understanding of them and thus have no basis for claiming that such designs should or should not be used. The logic and reasoning underlying RCTs is complex and requires an understanding of probability, equating, counterfactual inference, and many, many related matters. To simply dismiss RCTs as ‘too expensive’ or “too time consuming” is simply an error in reasoning. The real question is simply whether an RCT is the appropriate methodological choice. In essence, RCTs should be considered only when the primary questions are causal, there are sufficient resources (including monetary, time, and expertise), there is a high likelihood that a program or policy will remain in operation over a lengthy period of time, random allocation of a ‘treatment’ is viable (and ethical), and there are not a large number of RCTs already existing on the program or policy (e.g., is it really necessary to conduct another Head Start RTC?), among other considerations (e.g., political). Not all evaluations are concerned with causal inferences, yet – by their very nature – all RCTs are.*

**EAE:** Usually, evaluation is considered a key stage of the policy cycle. But not always evaluation is foreseen at the time of proposing the program. Is it possible to evaluate a program after the end of its implementation? What are the best designs to use in this scenario? Another practical issue relies in the budget and constraints which, sometimes, limit the scope for more comprehensive evaluation. As a specialist in research design, how do you think that is possible to overcome these constraints? Are there best approaches to consider when the time and budget are limited?

**CC:** *Time and budget restrictions are associated with nearly all program evaluations and are (generally) unrelated to the so-called ‘program lifecycle’ (a term that I have never adopted). Programs can be evaluated at nearly any ‘stage’ and the designs necessary are always contextually-dependent. Simply put, there is no ‘best’ or ‘right’ evaluation design. Design choices are informed by a variety of factors, including the ‘stage’ of the program, the resources available to execute the evaluation, the size and scope of the program, the purpose of the evaluation (e.g., formative, summative), the decisions to be informed by the evaluation, interest in the evaluation, and so on and so forth.*

**EAE:** Countries like Brazil and USA, with their continental dimensions, federative organization and diversity may have governmental programs to be implemented in different settings concurrently. Such diversity of contexts might change some prior features of the program or action being evaluated. What are the best methodological approaches to conduct multisite evaluations?

**CC:** *Multi-site evaluations have been the subject of much discussion and debate and the general consensus is that such evaluations are best designed and conducted as if such an evaluation were place-based (i.e., a single site), but with explicit emphasis and attention given to between-site variation (i.e., sites can be either homogenous or heterogenous, and this needs to be carefully considered). Otherwise, the methodological choices are essentially synonymous.*

**EAE:** Since 2008 you are Director of the Interdisciplinary Ph.D in Evaluation (IDPE) and Professor of Evaluation in the same doctorate program; you also were a senior Research Associate in The Evaluation Center (2006-2007), both at Western Michigan University. This professional experience as a practitioner and as a Professor of Evaluation must allowed you to develop a sense of the knowledge required for good practice on evaluation and the necessary knowledge to research in the evaluation field. What do you think are the basic knowledge that every evaluator should have? Is this knowledge the same knowledge required from a researcher in evaluation? Considering your experience in managing the IDPE program, which deals with persons form different background and fields of experience, what aspects cannot be lacking in the training of an evaluator?

**CC:** *My personal position is that competent evaluators, whether scholars or practitioners, should have a deep knowledge of evaluation logic and reasoning, evaluation history, evaluation concepts and its unique vocabulary, evaluation theories, evaluation methods, and evaluation practices. The specific foundational knowledge is extensive and includes both technical and non-technical skills and abilities. Certainly, evaluators require working knowledge of quantitative, qualitative, and mixed-methods of inquiry, measurement theory, research design, statistics, synthesis of facts and values and multiple values and many, many more technical skills. But almost equally important, evaluators should be competent communicators (written, oral, visual), be able to work effectively and efficiently in a variety of circumstances and conditions, have a working knowledge of project management, budgeting, and related skills, an ability to understand other cultures and cultural norms, ethics, and so forth. The American Evaluation Association (AEA) has published various versions of 'evaluator competencies,' many of which I endorse and others which I do not.*

EAE: The evaluative practice can confront the evaluator with ethical issues, either at the time of designing the evaluations, or in conducting them or in reporting the results. What are the main ethical issues posed by the adoption of experimental designs, quasi-experimental designs and nonexperimental designs in the practice of evaluation? In USA is there some guidance to assess and assure best ethical practices in evaluation? If so, could you explain it?

CC: Ethical practice is not generally associated with any particular type of design, though most associate ethical dilemmas with experiments (e.g., withholding a potentially beneficial treatment). In the United States, ethical guidance, policies, and procedures are abundant and largely depend on the research sponsor (e.g., National Institutes of Health), the investigator's institution (most institutions of higher education in the United States have institutional review boards for research with both animal and human subjects).

EAE: In your doctorate, you were concerned about models for evaluating scientific research. Your practice as the editor of the Journal of MultiDisciplinary Evaluation also might confront you with the quality of research issue. Concerning the dissemination of knowledge in evaluation, what do you think that are the best criteria to define if an evaluative research "is good enough" or if it is "intolerable bad", paraphrasing expressions you used in your book with Stufflebeam, Evaluation theories, models and applications?

CC: *The features that characterize evaluation quality have a long history, originating from Scriven's concept of meta-evaluation (i.e., the evaluation of an evaluation) and later in the various versions of The Program Evaluation Standards. Though there is great debate regarding the characteristics associated with what constitutes good or high-quality evaluation, most regard utility, feasibility, propriety, and accuracy as the essential criteria for judging an evaluation's quality. These criteria are quite different than those often associated with social and behavioral research and are usually 'paradigm-specific'. Quantitative inquiry often emphasizes replicability, generalizability, objectivity and similar criteria as important characteristics, whereas certain forms of qualitative inquiry might value transparency and transferability, for example.*



This content is licensed under a Creative Commons attribution-type BY-NC.

