

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

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

ADRIANA BAUER^I

FABIANA SILVA FERNANDES^{II}

TRADUÇÃO: FERNANDO EFFORI DE MELLO^{III}

REVISÃO TÉCNICA: ADRIANA BAUER

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 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, a seu ver, são seus elementos distintivos.

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

^I Originalmente *Trends and Challenges in Program Evaluation: Interview with Chris Coryn*.

^I Fundação Carlos Chagas (FCC) e Universidade de São Paulo (USP), São Paulo-SP, Brasil; <http://orcid.org/0000-0002-5942-9181>; adbauer@fcc.org.br

^{II} Fundação Carlos Chagas (FCC), São Paulo-SP, Brasil; <https://orcid.org/0000-0002-3458-9963>; fsfernandes@fcc.org.br

^{III} Tradutor freelancer, São Paulo-SP, Brasil; feffori@gmail.com

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.

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

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.

APRESENTAÇÃO

Chris L. S. Coryn é diretor do Programa de Doutorado Interdisciplinar em Avaliação e professor titular de Avaliação na Western Michigan University (WMU). Tem graduação em Psicologia e mestrado em Psicologia Social e Experimental, ambos pela Indiana University. Seu doutorado, em Filosofia da Avaliação, pela Western Michigan University, foi orientado por Michael Scriven, importante teórico cujas contribuições no desenvolvimento da avaliação de programas são inúmeras. Com vasta experiência no campo da avaliação, Coryn tem como interesses de pesquisa a história da avaliação, a pesquisa em teorias da avaliação, métodos de pesquisa quantitativos, qualitativos e mistos, teorias da medição e psicometria e metanálise. Sua trajetória acadêmica é marcada por inúmeros trabalhos de investigação e avaliações em vários domínios, incluindo educação, ciência e tecnologia, saúde e medicina, desenvolvimento comunitário e internacional e serviços sociais e humanos, tendo autorado ou coautorado mais de cem relatórios de pesquisa. Dá aulas, palestras e *workshops* tanto nos Estados Unidos e Canadá quanto em outros países da Europa, Ásia e Oceania. Sua produção científica é extensa, com mais de noventa contribuições entre artigos científicos, verbetes para enciclopédias especializadas, capítulos e livros, dentre os quais se destacam as discussões conceituais, teóricas

e metodológicas sobre avaliação de programas. Por sua contribuição às áreas de pesquisa e ao campo de avaliação, já foi agraciado com quatro prêmios, com destaque para o Research on Evaluation Distinguished Scholar Award, da American Educational Research Association (AERA). É autor, juntamente com Daniel Stufflebeam (*in memoriam*), de *Evaluation theory, models and applications*, publicado pela Jossey-Bass em 2014.

EAE: Considerando o contexto brasileiro, no campo educacional, parece mais comum pesquisar políticas públicas utilizando um referencial de análise de políticas do que teorias e desenhos de avaliação de programas. Quais são as principais diferenças, e quais os princípios comuns, se é que existem, entre esses campos de pesquisa? Além disso, poderia nos dar uma definição de avaliação (no contexto de avaliação de programas) que o senhor considere boa/adequada?

CC: *Os termos avaliação e análise de políticas são frequentemente utilizados como sinônimos, e as distinções entre eles não são necessariamente metodológicas. Ambas se ocupam de objetos de investigação semelhantes. Porém a análise de políticas normalmente se ocupa somente de questões do tipo “o que é?” (ou seja, conclusões descritivas baseadas em fatos) e geralmente não aborda questões do tipo “e daí?” (isto é, conclusões avaliativas baseadas em valores). Valor, valores e valoração são os aspectos centrais que distinguem a avaliação das disciplinas relacionadas, incluindo análise de políticas e pesquisa aplicada, entre muitas outras (por exemplo, avaliação de aprendizado). Por definição – e, portanto, dado que a raiz do termo avaliação é “valu” –, avaliação é o processo de determinar o mérito, o valor ou o significado de algo, ou o produto desse processo. Essa definição não é nem muito ampla nem muito restrita, e é suficiente para abarcar a maioria dos tipos e objetos de avaliação (por exemplo, programas, políticas, profissionais).*

EAE: No Brasil, as políticas e programas são marcados por rupturas e mudanças significativas em suas concepções, por conta das mudanças de poder nos governos. No seu ponto de vista, como conduzir avaliações de programas considerando o desafio imposto por esse contexto político e administrativo dos entes federativos?

CC: *A política sempre foi, e provavelmente sempre será, um importante inibidor da avaliação. Sem me aprofundar muito nas especificidades da política e nas implicações da política para os programas, a resposta simples é que a avaliação e os avaliadores devem*

ser flexíveis e se adaptar a contextos e forças políticas mutáveis. Concepções rígidas e fixas de avaliação quase sempre fracassam.

EAE: A avaliação surge em um cenário de responsabilização [accountability] pública. No Brasil, a discussão sobre as relações entre avaliação e responsabilização é muito controversa. Como o senhor entende essas relações?

CC: *A responsabilização é uma função – ou um propósito – da avaliação. Responsabilização é simplesmente “prestar contas ou ser responsável por algo”, que poderia incluir resultados (ou a falta deles), despesas, atividades e, em última instância, uma divulgação transparente (inclusive das avaliações). Então, em certos sentidos, a responsabilização é um aspecto importante da avaliação, mas não o único.*

EAE: Como o senhor resumiria, em termos teóricos, as principais tendências e movimentos em avaliação de programas?

CC: *Neste momento da breve história da avaliação profissional, o campo está repleto de uma série de “modismos” e “tendências” não necessariamente benéficos ao desenvolvimento da disciplina ou profissão. Isso inclui, por exemplo, as várias abordagens orientadas para a justiça social, o pensamento sistêmico e a teoria dos sistemas, avaliação orientada por teoria e muitos, muitos outros. Fundamentalmente, nenhum desses conceitos é exclusivo da avaliação, estão difundidos em vários campos e disciplinas semelhantes. No entanto, eles compartilham uma característica comum: nenhum está explicitamente preocupado em determinar o mérito, o valor ou o significado; ou, novamente, valor, valores e valoração. Algumas das abordagens populares atualmente em voga são equivocadas, na minha humilde opinião, por conta de preconceitos disfarçados de defesa social, ou a abordagens excessivamente preocupadas com a explicação do fenômeno, por exemplo. Essas preocupações são legítimas e dignas, mas, como disse anteriormente, todas elas esquecem o principal. A avaliação ocupa-se, acima de tudo, com a qualidade (mérito) e a importância (valor). Pense nas implicações subjacentes aos vários sistemas e abordagens orientadas por teoria. Simplificando, já é extremamente difícil determinar se algo funciona (ou seja, descrição causal simples) ou produz efeitos benéficos (ou prejudiciais), quanto mais tentar explicar como, para quem e sob que condições (ou seja, explicação causal). Não estamos de acordo sequer sobre como abordar estas últimas, quanto mais sobre aquelas primeiras.*

EAE: O senhor atua como membro dos Conselhos Editoriais de dois periódicos importantes no campo da avaliação: *American Journal of Evaluation* e *Evaluation and Program Planning*. Além disso, sempre contribuí para o campo com resenhas

de livros sobre projetos de avaliação e pesquisa (entre outros tópicos) e é editor-executivo de uma revista especializada em avaliação. Essa experiência lhe permite conhecer muito bem a produção técnica e científica sobre avaliação de programas. Em sua constante revisão da literatura sobre a avaliação de programas, o senhor vê dificuldades e falhas por parte dos autores na condução de pesquisas avaliativas? Se sim, quais destacaria?

CC: *Os relatórios de avaliação não costumam ser publicados em periódicos revisados por pares e não são considerados como investigações acadêmicas. Tais relatórios são tidos como um elemento essencial da prática, mas não um meio de aumentar ou contribuir para os conhecimentos sobre avaliação. Já as pesquisas SOBRE teorias, métodos e práticas de avaliação aumentaram drasticamente nas últimas décadas. Esse tipo de investigação foi definido como “Qualquer investigação intencional, sistemática e empírica destinada a testar o conhecimento existente, contribuir para o conhecimento existente ou gerar novos conhecimentos relacionados a algum aspecto dos processos ou produtos avaliativos, ou de teorias, métodos ou práticas avaliativas”² (p. 161).³ No entanto, grande parte da literatura atual publicada nos principais periódicos da área consiste em artigos do tipo “como” (por exemplo, como utilizar a análise de escore de propensão) ou “aplicações de” (por exemplo, uma aplicação da avaliação transformadora no contexto x), em vez de pesquisas empíricas sobre teorias, métodos ou práticas avaliativas com potencial de levar a disciplina adiante. Embora os artigos de “como” e “aplicações de” sejam valiosos, eles não examinam empiricamente as afirmações apresentadas em abordagens teóricas, seja para confirmar ou invalidar seu uso na prática (por exemplo, a avaliação de empoderamento realmente resulta em empoderamento?). Como disse certa vez um eminente estudioso da avaliação: “A avaliação em si não deveria estar aberta à investigação sistemática, assim como estão as políticas, programas e práticas que avaliamos?”*

EAE: Cunhados por Scriven, os conceitos de avaliação formativa e somativa já são bem conhecidos no campo da avaliação, frequentemente vistos como finalidades incompatíveis com a prática da avaliação de programas: uma avaliação deve ter um objetivo mais relacionado à melhoria do programa ou à responsabilização. Quais são as características distintivas entre a avaliação somativa e formativa na prática da avaliação de programas? Como os fins formativos e somativos podem ser imbricados no ciclo de políticas públicas?

2 Do original: “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”.

3 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, 2007.

CC: Em essência, os conceitos de avaliação formativa e somativa são simplesmente as funções ou finalidades associadas aos dois conceitos. O primeiro é explicitamente orientado para a melhoria, enquanto o segundo é essencialmente direcionado à responsabilização e à tomada de decisões. Ambas, no entanto, envolvem decisões, e é o uso de informações avaliativas nesse processo que determina a função ou o propósito de uma avaliação (ou seja, como uma avaliação é realmente utilizada). Uma avaliação formativa conduzida com a intenção de melhorar um programa ou política, por exemplo, pode resultar na decisão de encerrar o programa ou a política avaliada (isto é, uma decisão somativa). Assim, os propósitos funcionais claros de formativo e somativo são, às vezes, simples, embora ambos dependam claramente de como uma avaliação é realmente utilizada, e não de usos pretendidos. Mesmo assim, as distinções “formativa” e “somativa” são suficientes para classificar a maioria dos focos avaliativos, apesar de algumas afirmações de que a avaliação é mais do que simplesmente formativa ou somativa (por exemplo, para o desenvolvimento, atributiva, para o aprendizado).

EAE: Outra questão importante na avaliação de programas parece ser a definição de critérios de julgamento. Se avaliar é atribuir valor e envolve o julgamento de um ente avaliado, há algum critério melhor para definir com mais precisão como ele deva ser julgado? Na sua visão, quais são as principais perspectivas teóricas atuais sobre o problema da valoração?

CC: A valoração é, talvez, o grande dilema da avaliação de programas, e não há uma resposta “correta” ou “simples” sobre como os valores devam ser incorporados a uma avaliação. Mesmo assim, a lógica da avaliação de Scriven nos dá uma base sólida para resolver esse problema. Essa lógica e padrão de raciocínio consistem em: (1) identificar critérios – ou seja, os “valores” em relação aos quais um objeto deve ser avaliado (embora simplificando de forma um pouco grosseira); (2) determinar padrões de desempenho para cada critério; (3) medir o desempenho e compará-lo com os padrões; e (4) sintetizar em uma conclusão avaliativa de mérito e valor. As fontes dos valores – ou seja, os critérios – são um desafio constante, e esses valores podem ser obtidos de muitas formas diferentes (veja a *key evaluation checklist – checklist chave de avaliação* –⁴ de Scriven, como referência de fontes de valores/critérios). Os valores (ou critérios) não são necessariamente

⁴ Nota da revisora técnica: trata-se de menção à lista de verificação de avaliação, elaborada por Scriven, que busca subsidiar o usuário com uma listagem de aspectos que deveriam, idealmente, ser considerados em uma avaliação de programas. A *key evaluation checklist* (KEC) tem sido revisitada e reformulada desde sua primeira proposição, em 1971, com mais de 100 edições. Trata-se de um documento com dezoito (18) aspectos, divididos em quatro dimensões (Preliminares, Fundamentos, Subavaliações e Conclusões e Implicações) que deveriam ser considerados no desenho, gerenciamento, monitoramento e avaliação de propostas ou planos de avaliação de políticas, programas, projetos e processos, bem como na análise de avaliabilidade ou na própria metavaliação. As dimensões e aspectos que as compõem são explicados e justificados por meio de notas e referências teóricas. O leitor pode encontrar a última versão do documento, de 2015, em inglês, no site <http://michaelscriven.info/>.

as metas ou objetivos do programa, embora estes devam ser seriamente considerados. As necessidades do destinatário são uma fonte central de valores. Com relação às necessidades, como fonte de critérios ou valores para julgar programas, deve-se distinguir entre as necessidades instrumentais e as de “tratamento” (assim como de “carências”).

EAE: Atualmente, algumas abordagens em avaliação de programas são voltadas ao aprimoramento da utilização, de modo a melhorá-los, ou, ainda, a subsidiar decisões sobre seu desenvolvimento. Essas abordagens podem desafiar o papel geralmente atribuído aos avaliadores, uma vez que, às vezes, o avaliador se envolve no processo decisório e no próprio gerenciamento do programa. Como o senhor entende o papel do avaliador na prática avaliativa de hoje, considerando suas opiniões sobre as melhores abordagens avaliativas para o século XXI?

CC: *A recente tendência do avaliador como parte interessada do programa, ou como tomador de decisões – em grande parte decorrente da noção de avaliação para o desenvolvimento de Patton – é preocupante. Afinal, vários estudiosos da avaliação (inclusive Patton) afirmaram por muito tempo que tais atividades tendem a criar várias formas de preconceito e a inibir severamente a capacidade de realizar avaliações imparciais e não enviesadas. Pode-se argumentar que os avaliadores têm habilidades e conhecimentos importantes, os quais são relevantes para a concepção e administração dos programas. Mesmo assim, essas pessoas não devem avaliar programas em que tenham um interesse envolvido ou um conflito de interesses.*

EAE: A natureza da avaliação pode ser muito controversa. Algumas pessoas podem argumentar que a avaliação é uma pesquisa aplicada, baseada nos conhecimentos e procedimentos de outras áreas disciplinares. Outros afirmam que a avaliação é um campo disciplinar em si, com seu próprio raciocínio teórico. Michel Scriven entende a avaliação como uma transdisciplina. Entre os interesses de pesquisa do senhor estão estudos sobre teorias de avaliação. Assim, como definiria a natureza da avaliação e, mais especificamente, a natureza da avaliação de programas? O que são teorias de avaliação? As pesquisas sobre teoria da avaliação têm importância?

CC: *Em alguns aspectos, os debates sobre a natureza disciplinar da avaliação são irrelevantes à prática real da avaliação e, na verdade, só interessam aos estudiosos da avaliação. Se a avaliação é ou não pesquisa aplicada de ciências sociais, essa questão, assim como as definições de avaliação, não tem consenso; embora eu discorde fortemente. A pesquisa aplicada, por um lado, é uma pesquisa voltada a fornecer uma solução para*

um problema percebido. Já a avaliação está preocupada se a solução é meritória e vale a pena. As teorias no campo da avaliação são apenas prescrições sobre como as avaliações devem ser praticadas. Não são teorias explicativas ou preditivas, da forma como entendemos o significado e o uso coloquial do termo. Dado que tais teorias são prescritivas, é extremamente importante investigá-las empiricamente, a fim de averiguar até que ponto as alegações teóricas são confirmadas ou refutadas, e também seguidas na prática avaliativa (veja a pesquisa de Miller e Campbell sobre avaliação de empoderamento⁵ e a pesquisa de Coryn et al. sobre avaliação orientada por teoria).

EAE: Há um debate controverso, não apenas no Brasil, mas também internacionalmente, sobre o uso de Ensaio Controlado Randomizado (ECR) como padrão-ouro em avaliação de impacto. O senhor poderia explicar seus pontos de vista sobre os aspectos positivos e negativos em torno dos ECRs e outros desenhos de avaliação fundados na prática baseada em evidências?

CC: *Os argumentos a favor e contra o uso dos ECR são bem documentados, e a resposta é quase sempre “depende”. Embora muitos resistam ao uso dos ECR para fins de avaliação, há casos em que tais desenhos são a escolha adequada. Testes farmacêuticos são exemplos claros de uma aplicação apropriada dos ECR. No caso de outros tipos de intervenções sociais em assuntos humanos, a resposta não é tão evidente. Dito isso, os ECR respondem apenas a um tipo de pergunta, a saber, o tipo “causal”. Em minha experiência, tanto os partidários do uso dos ECR quanto os que se dizem contrários geralmente não têm um entendimento real sobre eles e, portanto, não têm base para afirmar que tais desenhos devam ou não ser usados. A lógica e o raciocínio subjacentes aos ECR são complexos e exigem uma compreensão de probabilidade, equacionamento, inferência contrafactual e muitos assuntos relacionados. Simplesmente descartar os ECR como “muito caros” ou “muito demorados” é um erro de raciocínio. A verdadeira questão é apenas se um ECR é a escolha metodológica apropriada. Essencialmente, deve-se considerar a realização de um ECR apenas quando as questões básicas forem causais, existirem recursos suficientes (incluindo monetários, de tempo e mão de obra qualificada), houver grande probabilidade de o programa ou política continuar operando por um longo período, a alocação randomizada de um “tratamento” for viável (e ética), e não houver um grande número de ECR já em curso no programa ou política (por exemplo, é realmente necessário conduzir*

⁵ Nota da revisora técnica: As referências bibliográficas dos trabalhos citados são: MILLER, R. L., CAMPBELL, R. Taking stock of empowerment evaluation: An empirical review. *American Journal of Evaluation*, v. 27, n. 9, p. 296-319, 2006; e CORYN, C. L. S., NOAKES, L. A.; WESTINE, C. D.; SCHROETER, D. C. A Systematic Review of Theory-Driven Evaluation Practice From 1990 to 2009. *American Journal of Evaluation*, v. 32, n. 2, p. 199-226, 2011.

mais um ECR promissor?), entre outras considerações (por exemplo, políticas). Nem todas as avaliações são voltadas para inferências causais, mas, por sua própria natureza, todos os ECR são.

EAE: Normalmente, a avaliação é considerada um estágio fundamental no ciclo de uma política, mas ela nem sempre está prevista no momento da proposição do programa. É possível avaliar um programa após o final de sua implementação? Quais os melhores desenhos a se utilizar nesse cenário? Outra questão prática refere-se ao orçamento e a restrições que, às vezes, limitam o escopo de uma avaliação mais abrangente. Como especialista em desenhos de pesquisa, como o senhor pensa ser possível superar essas restrições? Existem melhores abordagens a ser consideradas quando o tempo e o orçamento são limitados?

CC: *As restrições de tempo e orçamento estão associadas a quase todas as avaliações de programa e (geralmente) não estão relacionadas ao chamado “ciclo de vida do programa”. Os programas podem ser avaliados em praticamente qualquer estágio, e os desenhos necessários sempre dependem do contexto. Simplificando, não há um desenho “melhor” ou “correto” de avaliação. As escolhas de desenho são informadas por uma variedade de fatores, incluindo o “estágio” do programa, os recursos disponíveis para executar a avaliação, o tamanho e o escopo do programa, o propósito da avaliação (por exemplo, formativa, somativa), as decisões a ser informadas pela avaliação, o interesse na avaliação, e assim por diante.*

EAE: Países como o Brasil e os EUA, com suas dimensões continentais, organização federativa e diversidade, podem ter programas governamentais a ser implementados simultaneamente em diferentes cenários. Essa diversidade de contextos pode alterar algumas características originais do programa ou ação que está sendo avaliada. Quais são as melhores abordagens metodológicas para se realizar avaliações multilocais?

CC: *As avaliações multilocais foram objeto de muita discussão e muito debate, e o consenso geral sobre a melhor forma de desenhar e conduzir tais avaliações é tomá-las como se fossem baseadas em localidade (isto é, em local único), mas com grande ênfase e atenção à variação entre os locais (ou seja, os locais podem ser homogêneos ou heterogêneos, e isso precisa ser cuidadosamente considerado). No mais, as escolhas metodológicas são essencialmente sinônimas.*

EAE: Desde 2008, o senhor é diretor do Doutorado Interdisciplinar em Avaliação (IDPE) e professor de Avaliação no mesmo programa; também foi pesqui-

sador associado sênior do Evaluation Center (2006-2007), ambos na Western Michigan University. Essa experiência profissional como praticante e professor de avaliação deve ter-lhe permitido desenvolver um sentido tanto dos conhecimentos exigidos para uma boa prática avaliativa quanto dos conhecimentos necessários à pesquisa no campo da avaliação. O que o senhor considera serem os conhecimentos básicos que todo avaliador deveria ter? Esses conhecimentos são os mesmos necessários a um pesquisador de avaliação? Considerando a sua experiência na gestão do programa IDPE, que lida com pessoas de formações e áreas de experiência distintas, quais aspectos não podem faltar na formação de um avaliador?

CC: *Minha opinião pessoal é que os avaliadores competentes, sejam eles acadêmicos ou praticantes, devem ter profundo conhecimento da lógica e do raciocínio avaliativos, da história da avaliação, dos conceitos de avaliação e seu vocabulário exclusivo, dos métodos, das práticas e das teorias sobre avaliação. O conhecimento fundamental específico é extenso e inclui aptidões e habilidades técnicas e não técnicas. Certamente, os avaliadores necessitam ter conhecimento prático sobre métodos de pesquisa quantitativos, qualitativos e mistos, teoria da medida, desenhos de pesquisa, estatística, síntese de fatos e valores e de múltiplos valores, e muitas outras habilidades técnicas. Além disso, quase tão importante quanto esses aspectos, os avaliadores devem ser comunicadores competentes (em termos escritos, orais, visuais), ser capazes de trabalhar de forma eficaz e eficiente em uma série de circunstâncias e condições, ter um conhecimento prático de gerenciamento de projetos, orçamento e habilidades relacionadas, capacidade de entender outras culturas e normas culturais, éticas e assim por diante. A American Evaluation Association (AEA)⁶ publicou várias versões das “competências do avaliador”, muitas das quais eu endosso, outras não.*

EAE: A prática avaliativa pode confrontar o avaliador com questões éticas, seja no momento de desenhar as avaliações, de conduzi-las ou de relatar os resultados. Quais são as principais questões éticas colocadas pela adoção de desenhos experimentais, quase experimentais e não experimentais na prática da avaliação? Nos EUA, há alguma orientação sobre como avaliar e assegurar melhores práticas éticas na avaliação? Se sim, o senhor poderia explicar?

⁶ Nota da revisora técnica: A AEA tem contribuído com o debate das competências básicas do avaliador e dos princípios éticos subjacentes à prática avaliativa. Uma versão atualizada dos Princípios-Guia para os Avaliadores pode ser encontrada em: <https://www.eval.org/p/cm/ld/rid=51>.

CC: A prática ética geralmente não está associada a nenhum tipo específico de desenho, embora a maioria das pessoas associe dilemas éticos a experimentos (por exemplo, reter um tratamento potencialmente benéfico). Nos Estados Unidos, orientações, políticas e procedimentos éticos são abundantes e dependem, em grande medida, do patrocinador da pesquisa (por exemplo, os Institutos Nacionais de Saúde – NHI) e da instituição do pesquisador (a maioria das instituições de ensino superior nos Estados Unidos tem conselhos institucionais de revisão para pesquisas, tanto com animais quanto com seres humanos).

EAE: Em seu doutorado, o senhor se preocupou com desenhos de avaliação de pesquisas científicas. Sua prática como editor do *Journal of MultiDisciplinary Evaluation* também pode confrontá-lo com a questão da qualidade da pesquisa. Com relação à disseminação do conhecimento em avaliação, quais critérios o senhor considera os melhores para definir se uma pesquisa avaliativa “é boa o suficiente” ou se é “intoleravelmente ruim”, parafraseando expressões que o senhor utilizou em seu livro com Daniel Stufflebeam, *Evaluation Theory, Models and Applications*?

CC: Os traços que caracterizam a qualidade da avaliação têm uma longa história, que remonta ao conceito de metavaliação de Scriven (ou seja, avaliação da avaliação) e, posteriormente, às várias versões de *The Program Evaluation Standards*. Apesar do extenso debate sobre as características associadas ao que constitui uma avaliação boa ou de alta qualidade, consideram-se, majoritariamente, a utilidade, a viabilidade, a adequação e a precisão como critérios essenciais para avaliar a qualidade de uma avaliação. Esses critérios são bem diferentes dos frequentemente associados a pesquisas sociais e comportamentais e geralmente são “específicos de paradigma”. A investigação quantitativa muitas vezes enfatiza a replicabilidade, a generalização, a objetividade e critérios similares como características importantes, enquanto certas formas de pesquisa qualitativa podem valorizar a transparência e a transferibilidade, por exemplo.