

Avaliação de Programas Educacionais: Discutindo Padrões

*Dilvo I. Ristoff **

Resumo: O texto abaixo discute os atributos e padrões para a boa avaliação de programas educacionais elaborados pelo Joint Committee on Standards for Educational Evaluation e apresenta em tradução os quatro atributos e os trinta padrões correspondentes.

Palavras-Chave: Avaliação educacional, avaliação de programas, padrões, meta-avaliação.

Abstract: The text discusses the attributes and standards for good evaluation of educational programs, as developed by the Joint Committee for Educational Evaluation. The text also contains a translation into Portuguese of the four attributes and their corresponding thirty patterns.

Key Words: Educational Evaluation, Program Evaluation, patterns, meta-evaluation

Mulla Nasrudim havia se tornado um favorito do Rei. Começou então a usar a sua posição para demonstrar o modo de agir dos cortesãos.

Certo dia o rei chegou ao palácio com muita fome. Algumas beringelas que lhe foram servidas estavam tão deliciosas que ele ordenou ao chefe de cozinha que as servisse todos os dias.

“Mulla,” perguntou o Rei a Nasrudim, “Este é ou não é o melhor legume do mundo?”

“O melhor, Majestade” respondeu Nasrudim.

Cinco dias depois, quando as beringelas haviam sido servidas pela décima refeição consecutiva, o rei rugiu:

“Tirem esta coisa da minha frente! EU A ODEIO!”

“É o pior legume que existe, Majestade”, concordou Nasrudim.

“Mas Mulla”, disse o Rei, “há menos de uma semana você me disse que era o melhor!”

“É verdade, Majestade, eu de fato disse, mas eu sou um súdito do Rei, não do legume!”

– Idries Shah

É ainda freqüente no meio educacional a percepção de que a boa avaliação é aquela que serve aos interesses do poder estabelecido, dos solicitantes da avaliação ou dos administradores. Vista nesta perspectiva, no entanto, a avaliação corre o risco de negligenciar, ou tratar de forma apenas tangencial, aquilo que, por definição, lhe é central: as considerações de mérito e valor. Ao tornar secun-

dária a questão de mérito e valor, a avaliação torna-se facilmente sinônimo de marketing ou de instrumento de reforço das ações do poder estabelecido. Serve, pois, para confirmar a perspectiva do poder, para reforçar os seus juízos e pré-juízos, ou, como denuncia Nasrudim, na epígrafe acima, para agradar ao chefe. Trata-se, pois, para usar as palavras de Daniel Stufflebeam, não de avaliação propriamente dita, mas de pseudo-avaliação. Daí a importância de definir com clareza o que entendemos por “avaliação” e, em conseqüência, de estabelecer padrões que nos auxiliem na identificação do que seja um bom processo avaliativo.

* Doutor pela University of Southern California, Los Angeles e professor titular da Universidade Federal de Santa Catarina.

Em função do excelente trabalho que o Joint Committee on Standards for Educational Evaluation¹ vem desenvolvendo há vários anos, nos Estados Unidos, já se tornou lugar-comum dizer que boas avaliações de instituições, programas, projetos e materiais educacionais devem ter quatro atributos básicos:

1. Utilidade
2. Viabilidade
3. Propriedade
4. Exatidão

Menos conhecido, no entanto, é o significado que estes termos têm para os seus autores. Além dos quatro atributos acima, o Joint Committee, sob a presidência de James R. Sanders, e com a participação de Daniel Stufflebeam, Oliver W. Cummings, Joy Frechtling e muitos outros, estabeleceu sete padrões para o atributo "Utilidade", três padrões para o atributo "Viabilidade", oito padrões para o atributo "Propriedade" e doze padrões para o atributo "Exatidão". Formam, portanto, através de um conjunto de 30 padrões, uma filosofia do trabalho avaliativo cujo propósito é assegurar que a avaliação seja útil, viável, exata e ética.

Este trabalho pioneiro, é importante destacar, é o coroamento de um esforço coletivo iniciado em 1975. Embora sem tradição de trabalho conjunto, as associações profissionais americanas foram capazes de se sobrepor às profundas diferenças de pontos de vista existentes (Sanders lembra que "um membro, por exemplo, propunha banir os testes padronizados enquanto outros defendiam o seu uso. Também se discutiu se a avaliação institucional ou de pessoal deveria ou não ser incluída"). Em 1981, com a publicação de *Standards for Evaluations of Educational Programs, Projects, and Materials*, pela editora McGraw-Hill, o comitê deixou claro que havia tomado uma decisão importante: deixar a avaliação das instituições para as agências de acreditação. Mesmo limitando o seu campo de atuação, o Comitê a partir daquele ano tornou-se uma organização permanente, sem fins lucrativos, voltada exclusivamente ao desenvolvimento de padrões avaliativos, à sua revisão periódica e à pesquisa na área específica.

Em 1988, o trabalho teve continuidade com a publicação do livro *The Personnel Evaluation Standards* (Sage) e em 1989 O Joint Committee obteve a aprovação do American National Standards Institute (ANSI), o que na

prática veio a significar que os padrões definidos pelo Joint Committee passaram a ser adotados como padrões nacionais e, como tal, foram internacionalizados. Em 1989, o Joint Committee revisou o seu livro de 1981 e publicou o livro *Program Evaluation Standards*, publicado pela Sage. Os padrões em seguida apresentados foram traduzidos e adaptados da segunda edição, publicada em 1994.

O trabalho do Joint Committee, embora não esteja especificamente voltado para a avaliação institucional, é importante instrumento para nos auxiliar a melhor entender o grande potencial de atividades desenvolvidas nas instituições educacionais, e que podem e precisam ser objeto de avaliação. É importante também por outras razões.

Primeiro, pela definição que proporciona da própria atividade avaliativa. A partir do trabalho do Comitê, avaliação passou a ser definida como "investigação sistemática do valor ou mérito de um objeto" (e.g., um programa², projeto³, material instrucional, etc.). Ou seja, trata-se de uma atividade de pesquisa planejada, organizada, desenhada, sistemática, com objetivo muito específico, qual seja, definir o valor de um objeto *vis-a-vis* os seus possíveis propósitos ou valor de mercado e o seu mérito, entendido como a excelência aferida a partir das qualidades intrínsecas do objeto. Estas e outras definições fazem parte de um rico glossário que acompanha os trinta padrões e que foram resultado das discussões do Joint Committee e que com certeza ajudam a iluminar também a discussão brasileira sobre o assunto.

Segundo, é fundamental destacar que cada um dos 30 padrões vem acompanhado de um número significativo de diretrizes que o contextualiza, define e esclarece. Assim, por exemplo, o Padrão U1 - IDENTIFICAÇÃO DOS GRUPOS DE INTERESSE, vem acompanhado de seis importantes diretrizes, quais sejam:

1. Identificar lideranças e buscar o seu apoio.
2. Contatar representantes de grupos de interesse identificados para ver como percebem a importância da avaliação, como gostariam de usar os resultados e que informação eles considera-

1 Comitê patrocinado por quinze associações profissionais americanas, entre elas a Associação Americana dos Administradores de Escolas, a Associação Americana de Pesquisas Educacionais, a Associação Americana de Avaliação, e a Associação Americana de Psicologia.

2 O Joint Committee define "Programa" como "atividades educacionais que são ofertadas por um período de tempo continuado. Exemplos incluem um programa de leitura de um distrito educacional, um programa de treinamento militar ou industrial, um programa de educação médica, ou programa profissional de educação continuada.

3 O Joint Committee define "Projeto" como "atividades educacionais que são ofertadas por um período de tempo específico. Exemplos de projeto: uma oficina sobre estratégias instrucionais com a duração de três dias, um trabalho de três anos voltado ao desenvolvimento de testes . . . Projetos que se institucionalizam tornam-se Programas".

riam útil. Alertar para o fato de que nem todas as necessidades de informação podem ser atendidas.

3. Discutir com os solicitantes da avaliação os interesses dos vários grupos de interesse.
4. Ficar alerta, durante o processo, sobre a existência de outros possíveis grupos de interesse.
5. Envolver o cliente e os grupos de interesse no planejamento e na condução do processo avaliativo.
6. Não excluir grupos de interesse com base em sexo, etnicidade ou *background* lingüístico.

Como é possível verificar, as diretrizes apontam alguns caminhos práticos que podem ser seguidos para validar um Padrão específico. Não estamos autorizados a traduzir as diretrizes definidas para cada um dos trinta padrões, razão pela qual recomendamos a sua leitura no texto original, pois como o exemplo acima demonstra, as diretrizes explicitam com maior clareza o que exatamente os seus autores imaginaram quando do estabelecimento do Padrão.

Terceiro, os padrões em questão são acompanhados de vários registros dos equívocos mais frequentes cometidos pelos avaliadores. Assim, por exemplo, para o PADRÃO DE VIABILIDADE, V3, Custo-Benefício, os erros mais comuns apontados pelo Comitê são os seguintes:

1. Ignorar simplesmente a questão custo-benefício simplesmente porque uma avaliação é exigida e será feita independentemente dos custos;
2. Selecionar um método de avaliação já conhecido sem buscar alternativas que são menos dispendiosas;
3. Iniciar uma avaliação sem garantias de recursos suficientes para concluí-lo;
4. Partir do princípio de que uma alocação inicial suficiente de recursos torna desnecessária a preocupação em demonstrar que a avaliação é econômica;
5. Permitir que preocupações com a economicidade desestimulem os avaliadores e clientes (solicitantes da avaliação) a testarem novos métodos ou a fazerem avaliações que são importantes mas de difícil execução e que podem levar a dados e informações adicionais de grande utilidade;
6. Não admitir que as análises da relação custo-benefício são muitas vezes difíceis de serem completadas porque os resultados considerados em uma avaliação podem ser inúmeros, intangíveis e avaliados de forma diferente por

diferentes grupos, e porque nem sempre são fáceis de medir ou observar.

Novamente, a indicação dos equívocos mais comuns é uma forma encontrada pelos autores de ajudar o leitor a entender o significado de um padrão específico, tornando-se leitura complementar importante.

Por último, o texto ficaria incompleto e de difícil compreensão se os padrões tivessem sido contextualizados e explicitados unicamente a partir das diretrizes e dos equívocos mais comuns. Mais importante, são os exemplos. Por esta razão, o Joint Committee selecionou um exemplo real e concreto para cada um dos 30 padrões, como o exemplo abaixo, relativo ao Padrão V3, Custo-benefício.

“O superintendente de um grande distrito escolar contratou um avaliador para avaliar um novo programa de matemática para o ensino fundamental. Embora parte da avaliação visasse às habilidades computacionais dos alunos da terceira e quarta séries, o desempenho individual dos estudantes não estava em questão. Na verdade, só resultados do distrito como um todo deveriam ser relatados.

Um comitê composto por um professor selecionado de cada escola do distrito escolheu, com base em um modelo consensual, um conjunto de 300 problemas computacionais de um enorme banco de itens. Os itens selecionados foram considerados apropriados para avaliar até que ponto 30 objetivos computacionais do programa de matemática foram alcançados.

Os avaliadores reuniram os 300 itens em um teste que exigiu para a sua aplicação três sessões de duas horas e meia cada. Este teste foi aplicado a todos os alunos de terceira e quarta séries do distrito”.

Além de simplesmente apresentar o exemplo à discussão dos estudiosos do tema, o Joint Committee oferece a sua própria interpretação para esclarecer por que, no caso, o padrão ajuda a revelar as deficiências do processo avaliativo em questão:

“Embora o objetivo (verificar até que ponto os 30 objetivos computacionais do programa de matemática foram alcançados) tenha sido atingido, a avaliação exigiu um investimento em termos de tempo de alunos e professores muito acima do necessário, especialmente se considerarmos que os escores individuais não estavam sendo solicitados.

Em vez de construir um único teste com 300 problemas, os avaliadores poderiam ter aleatoriamente dividido os problemas em 10 subconjuntos de 30 problemas de modo que cada objetivo estivesse coberto por cada um dos subconjuntos. Os 10 subconjuntos poderiam ter sido aleatoriamente distribuídos e aplicados aos estudantes em uma amostragem aleatória de salas de aula, de modo que o estudante que participasse responderia a um único

subconjunto de 30 itens. Uma economia substancial de tempo, para estudantes e professores, teria sido conseguido sem prejuízo à exatidão da avaliação”.

Como é possível verificar, a mera listagem dos trinta padrões, embora um importante começo na discussão da avaliação da avaliação, é insuficiente para entender toda a amplitude e complexidade do trabalho realizado pelo Joint Committee. Mesmo assim, julgamos importante oferecer aos nossos leitores a tradução abaixo, devidamente autorizada pelos autores. No momento em que se proliferam no Brasil novos modelos e novas metodologias de avaliação, a avaliação da avaliação precisa tornar-se parte

de nossas preocupações cotidianas. Do contrário, será uma atividade estéril a serviço dos caprichos do poder estabelecido. Temos a certeza de que o texto abaixo contribuirá de forma significativa com as discussões dos processos avaliativos hoje em uso no país.

Obras citadas:

- Sanders, James R. *The Program Evaluation Standards: How to Assess Evaluations of Educational Programs*. Thousand Oaks: Sage, 1994.
- Shah, Idries. *The Pleasantries of the Incredible Mulla Nasrudin*. New York: Penguin, 1993.

Padrões para Avaliação de Programas

(The Joint Committee on Standards for Educational Evaluation - tradução de Dilvo I. Ristoff)

Boas avaliações de programas, projetos e materiais educacionais, em uma variedade de situações, têm quatro atributos básicos:

Utilidade
Viabilidade
Propriedade
Exatidão.

Os Padrões para Avaliação de programas, estabelecidos por 16 associações profissionais da área de educação, identificam princípios avaliativos que, se trabalhados, devem levar ao aperfeiçoamento das avaliações de programas-avaliações que devem conter quatro atributos que seguem:

1. UTILIDADE

Os padrões de utilidade têm por objetivo garantir que a avaliação atenda as necessidades de informação dos usuários preferenciais.

U1 - IDENTIFICAÇÃO DOS GRUPOS DE INTERESSE: Pessoas envolvidas com a avaliação ou afetadas por ela devem ser identificadas, para que seus interesses possam ser atendidos.

U2 - CREDIBILIDADE DOS AVALIADORES: Para que os resultados tenham o máximo de credibilidade, as pessoas responsáveis pela avaliação devem ser confiáveis e competentes para conduzir o processo avaliativo.

U3 – ABRANGÊNCIA E SELEÇÃO DA INFORMAÇÃO: A informação coletada deve ser suficientemente abrangente de modo a tratar de questões pertinentes à instituição e dar respostas às necessidades e interesses dos solicitantes e outros grupos de interesse especificados.

U4 – IDENTIFICAÇÃO DE VALORES: As perspectivas, procedimentos, e a justificativas usada para interpretar as descobertas devem ser cuidadosamente descritas, de modo a deixar bem claras as bases que sustentam os juízos.

U5 – CLAREZA DOS RELATÓRIOS: Os relatórios da avaliação devem descrever com clareza a instituição sob avaliação e o seu contexto. Devem igualmente descrever os propósitos, procedimentos, e descobertas da avaliação, disponibilizando a informação essencial e tornando-a facilmente inteligível.

U6 – OPORTUNIDADE E DISSEMINAÇÃO DOS RELATÓRIOS: Descobertas significativas ocorridas durante o processo avaliativo devem ser relatadas aos usuários preferenciais, para que as mesmas possam ser utilizadas no momentos oportuno.

U7 – IMPACTO DA AVALIAÇÃO: As avaliações devem ser planejadas, conduzidas e relatadas de modo a encorajar o acompanhamento constante dos interessados, tornando mais provável o uso efetivo da avaliação.

2. VIABILIDADE

Os padrões de viabilidade têm por objetivo assegurar que a avaliação será realista, prudente, diplomática, e econômica.

V1 – PROCEDIMENTOS PRÁTICOS: Os procedimentos avaliativos devem ser práticos, para reduzir ao mínimo os transtornos durante a busca de informações necessárias.

V2 – VIABILIDADE POLÍTICA: A avaliação deve ser planejada e conduzida de modo a levar em conta as diferentes posições dos vários grupos de interesse, possibilitando assim obter a sua cooperação, minimizando e neutralizando possíveis tentativas de algum destes grupos de restringir as operações avaliativas ou de deturpar ou aplicar mal os resultados.

V3 – CUSTO-BENEFÍCIO: A avaliação deve ser eficiente e produzir informações suficientes e valiosas, de modo a justificar os recursos investidos.

3. PROPRIEDADE

Os padrões de propriedade têm por objetivo assegurar que a avaliação seja conduzida legalmente e eticamente, e com adequada atenção ao bem-estar das pessoas envolvidas no processo bem como das pessoas afetadas pelos resultados.

P1 – ORIENTAÇÃO PARA O SERVIÇO: As avaliações devem ser desenhadas de forma a ajudar as instituições a dar atendimento efetivo às necessidades de todos os que dela dependem.

P2 – ACORDOS FORMAIS: As obrigações das partes envolvidas no processo avaliativo (o que deve ser feito, como, por quem, quando) devem ficar registradas por escrito, de modo a exigir que as partes adiram às condições do acordo ou que formalmente as renegociem.

P3 – DIREITOS DOS SUJEITOS: As avaliações devem ser desenhadas e conduzidas de modo a respeitar e proteger os direitos e o bem-estar dos sujeitos.

P4 – INTERAÇÕES HUMANAS: Os avaliadores devem respeitar a dignidade e os valores humanos em suas interações com as pessoas associadas à avaliação, de modo a não intimidar ou magoar os participantes.

P5 – AFERIÇÃO COMPLETA E JUSTA: A avaliação deve ser completa e justa em seu exame e registro das forças e fraquezas de uma instituição, permitindo reforçar os seus pontos fortes e lidar com os seus aspectos problemáticos.

P6 – APRESENTAÇÃO DE RESULTADOS: As partes formalmente envolvidas na avaliação devem garantir que as pessoas afetadas por ela tenham acesso a todas as descobertas bem como as limitações pertinentes, devendo ficar assegurado o seu direito legal de ser informado dos resultados.

P7 – CONFLITO DE INTERESSES: Os conflitos de interesse devem ser discutidos abertamente e honestamente, para que eles não comprometam os processos e resultados da avaliação.

P8 – RESPONSABILIDADE FISCAL: A alocação e os gastos do avaliador devem refletir procedimentos adequados e responsáveis, além de serem prudentes e éticos, para que os gastos sejam apropriados e justificáveis.

4. EXATIDÃO

Os padrões de exatidão têm por objetivo assegurar que a avaliação revele e transmita informações tecnicamente adequadas sobre os aspectos que determinam valor e mérito da instituição sob avaliação.

E1 – DOCUMENTAÇÃO DA INSTITUIÇÃO: A instituição sob avaliação deve ser descrita e documentada claramente e com exatidão, para que a instituição seja identificada com clareza.

E2 – ANÁLISE CONTEXTUAL: O contexto no qual a instituição existe deve ser examinado detalhadamente, de modo a identificar as suas prováveis influências sobre a instituição.

E3 – PROPÓSITOS E PROCEDIMENTOS DESCRITOS: Os propósitos e procedimentos da avaliação devem ser monitorados e descritos detalhadamente, para que possam ser identificados e aferidos.

E4 – FONTES DE INFORMAÇÕES DEFENSÁVEIS: As fontes de informações utilizadas na avaliação de uma instituição devem ser detalhadamente descritas, para que a adequação da informação possa ser aferida.

E5 – INFORMAÇÃO VÁLIDA: Os procedimentos de coleta de informação devem ser escolhidos ou desenvol-

vidos e implementados de modo a assegurar que a interpretação final seja válida para o uso pretendido.

E6 – INFORMAÇÃO CONFIÁVEL: Os procedimentos de coleta de informação devem ser escolhidos ou desenvolvidos e implementados de modo a assegurar que a informação obtida é suficientemente confiável para o uso pretendido.

E7 – INFORMAÇÃO SISTEMÁTICA: A informação coletada, processada, e relatada num processo avaliativo deve ser sistematicamente revisada e todos os erros devem ser corrigidos.

E8 – ANÁLISE DE INFORMAÇÕES QUANTITATIVAS: As informações quantitativas em uma avaliação deve ser apropriadamente e sistematicamente analisadas, para que perguntas avaliativas possam ser efetivamente respondidas.

E9 – ANÁLISE DE INFORMAÇÕES QUALITATIVAS: As informações qualitativas em uma avaliação deve

ser apropriadamente e sistematicamente analisadas, para que perguntas avaliativas possam ser efetivamente respondidas.

E10 – CONCLUSÕES JUSTIFICADAS: As conclusões tiradas a partir da avaliação devem ser explicitamente justificadas, para que os grupos de interesse possam aferi-las.

E11 – RELATÓRIO IMPARCIAL: Os relatórios devem prever salvaguardas contra distorções produzidas por sentimentos pessoais e por preconceitos de alguma das partes para com a avaliação, para que os relatórios reflitam com justeza as descobertas avaliativas feitas.

E12 – META-AVALIAÇÃO: A própria avaliação deve ser avaliada formativa e somativamente a partir destes e de outros padrões, para que a sua condução seja apropriada e, uma vez concluído o processo, para que os grupos de interesse possam examinar minuciosamente as suas forças e fraquezas.