

Meta-Avaliação de uma Avaliação em Larga Escala: Pressupostos e Proposições

Adriana Bauer

Departamento de Pesquisas
Educacionais - FCC
São Paulo, SP, Brasil
adbauer@fcc.org.br

Raquel da Cunha Valle

Departamento de Pesquisas
Educacionais - FCC
São Paulo, SP, Brasil
rvalle@fcc.org.br

Resumo

Este trabalho apresenta e discute uma proposta de meta-avaliação para avaliações em larga escala, a partir de iniciativa meta-avaliativa elaborada para um município do sudeste brasileiro. Inicialmente, a relevância da proposição é justificada, considerando-se sua contribuição para o futuro desse tipo de avaliação; em um segundo momento, apresenta-se a metodologia de elaboração da iniciativa, as dimensões e padrões de meta-avaliação propostos. O texto finaliza com as indicações sobre como se concretiza, em termos metodológicos, a meta-avaliação, compartilhando resultados de um estudo exploratório realizado a partir do modelo elaborado.

Palavras-chave: Meta-avaliação; avaliações em larga escala; indicadores.

1 Introdução

As avaliações externas e em larga escala se disseminaram e consolidaram, no Brasil, ensejando muitas dúvidas, polêmicas e contradições, tanto no que se refere ao seu significado político, quanto no que diz respeito às questões técnicas relativas às medidas educacionais que lhe são inerentes. Em relação a esse último aspecto, no Brasil não existe um órgão regulador que se responsabilize pela qualidade técnica das provas em larga escala, nem que estabeleça possíveis usos de seus resultados, considerando padrões éticos e de justiça social, o que suscita diversos questionamentos sobre a qualidade de seus instrumentos e dos resultados obtidos.

Este trabalho parte do pressuposto de que uma meta-avaliação pode contribuir para a apreciação das iniciativas de avaliação em larga escala em curso, bem como propiciar a obtenção de informações para seu aprimoramento; ou seja, os aprendizados oriundos de uma meta-avaliação de avaliação em larga escala podem ser usados tanto pelas empresas que as operacionalizam, como pelas redes de ensino que as demandam.

Mas o que é uma meta-avaliação? Segundo Furtado e Laperrière, a meta-avaliação “se caracteriza por ter a avaliação como objeto de interesse. É de natureza qualitativa e pretende estabelecer um julgamento de valor, com critérios pré-definidos, com base em informações existentes ou construídas sobre uma determinada avaliação (em curso ou já realizada) para subsidiar eventuais tomadas de decisão” (Furtado e Laperrière, 2012, p. 696).

A avaliação da avaliação (meta-avaliação) tem sido empregada para avaliar a qualidade técnica dos processos avaliativos e a pertinência de suas conclusões. Embora ela esteja presente na literatura atinente desde a década de 1960, são escassas as iniciativas de meta avaliar avaliações em larga escala (RUTKOWSKI, 2017).

Segundo Ory (1992), o movimento de avaliações de estudantes, seja no ensino básico ou no ensino superior, muitas vezes direcionado por propósitos de responsabilização (*accountability*), tem negligenciado a vasta produção teórica e prática desenvolvida no campo da avaliação de programas. Para o autor, aprendizados realizados durante a expansão do campo da avaliação de programas, como a meta-avaliação, por exemplo, podem auxiliar os responsáveis por provas em larga-escala e avaliações de aproveitamento de estudantes, a aprimorar os processos desse tipo de avaliação.

É nesse sentido que se propôs da pesquisa que ora se apresenta, na qual se buscou desenhar uma meta-avaliação formativa que pudesse subsidiar a tomada de decisões dos gestores e avaliadores, visando ao aprimoramento das iniciativas a partir dos aspectos investigados. Ou seja, parte-se do pressuposto de que fazer meta-avaliações de avaliações em larga escala é um esforço necessário para o seu aprimoramento.

No contexto estadunidense, a partir da proposta de se realizar meta-avaliações para garantir a qualidade das avaliações de programas, foi necessário

estabelecer padrões para orientá-las, tarefa que ficou sob responsabilidade do *Joint Committee on Standards for Educational Evaluation*. Além desses padrões, internacionalmente conhecidos, existem diretrizes e recomendações para garantir a qualidade técnica de provas e testagens diversas, tais como os *Standards for Educational and Psychological Testing* (2014), que definem padrões relacionados à validade, confiabilidade, precisão e justiça dos testes psicológicos e avaliativos sem, no entanto, definir procedimentos para a análise dos resultados obtidos.

O problema que esta pesquisa buscou enfrentar é que os padrões de meta-avaliação existentes e já consagrados na literatura não abarcam as especificidades da avaliação em larga escala e os padrões de qualidade dos testes psicológicos e cognitivos não discutem procedimentos de meta-avaliação. Foi esta lacuna que a pesquisa ora reportada buscou superar: a proposição de um modelo meta-avaliativo para uma avaliação em larga escala. Cabe destacar que o modelo de meta-avaliação discutido a seguir foi desenvolvido a pedido de uma secretaria de educação da região sudeste do Brasil que possui uma avaliação em larga escala própria.

2 Metodologia

Para a elaboração da proposta, alguns procedimentos metodológicos foram realizados. Partiu-se de uma revisão bibliográfica abrangente sobre meta-avaliação, abarcando a literatura nacional e internacional (em língua inglesa), a fim de verificar se havia a proposição de uma forma específica de realizar meta-avaliações de avaliações em larga escala (ELLIOT, 2011; COOKSY, CARACELLI, 2009; dentre outros). Os trabalhos revisados, apesar de discutirem princípios muito importantes, não propunham como fazer uma meta-avaliação de uma avaliação em larga escala, focando em discussões mais amplas.

Além dos textos acadêmicos, outros dois documentos foram analisados: o "*Standards for Educational and Psychological Testing*", produzido nos Estados Unidos pelo consórcio entre a APA (*American Psychological Association*), a AERA (*American Educational Research Association*) e o NCME (*National Council on Measurement in Education*) e o "*The program evaluation standards: a guide for evaluators and evaluation users*" (YARBROUGH et al, 2010).

A partir do estudo desses documentos, foram identificadas e selecionadas as principais dimensões analíticas, assim como os padrões associados a essas dimensões. Estas foram comparadas com as outras propostas identificadas na revisão bibliográfica. Observou-se a existência de preocupações convergentes no que se refere à garantia da qualidade dos processos avaliativos. Considerando-se essas comunalidades e a especificidade do objeto do estudo, definiram-se as dimensões, utilizadas no processo meta-avaliativo, a saber: confiabilidade e precisão, viabilidade, utilidade e ética e justiça, que são entendidas como:

Confiabilidade e precisão: busca mostrar em que medida as interpretações pretendidas dos resultados articulam-se aos seus usos possíveis, partindo do princípio de que devem ser fornecidas evidências de validade apropriadas para apoiar cada interpretação. Os padrões propostos são: CP1 - Clareza na descrição

da avaliação e dos seus propósitos ao longo do tempo, CP2 - Preocupação em fornecer informações válidas, CP3 - Preocupação em fornecer informações fidedignas e CP4 - Processos de gerenciamento das informações.

Viabilidade: relacionada à efetividade e eficiência da avaliação, busca informações acerca das estratégias de gerenciamento do processo avaliativo, adequação, praticidade e responsividade dos procedimentos utilizados em sua operacionalização, bem como análise acerca da adequação dos recursos. Nesta dimensão foram definidos os seguintes padrões: V1 - Adequação do cronograma às necessidades da rede de ensino, V2 - Clareza e adequação dos procedimentos de realização da prova, V3 - Viabilidade contextual, V4 - Uso eficiente e eficaz dos recursos.

Utilidade: busca captar se as partes interessadas na avaliação (gestores de nível central, diretores escolares, pais e responsáveis etc.) consideram os processos e produtos da avaliação valiosos para atender às suas necessidades de informação. É composta pelos seguintes padrões: U1 - Credibilidade do avaliador, U2 - Atenção aos interessados, U3 - Adequação da comunicação dos resultados aos interessados, U4 - Estratégias para evitar usos equivocados dos resultados, U5 - Preocupação de que o processo avaliativo e os resultados sejam significativos para os interessados, U6 - Comunicação oportuna.

Ética e Justiça: tem por objetivo verificar em que medida uma avaliação é transparente em seus processos e imparcial na produção e divulgação das informações. Os padrões propostos são: EJ 1 - Transparência na divulgação da documentação, EJ 2 - Acompanhamento e análise da execução da prova por equipes internas à secretaria de educação (SE), EJ 3 - Acompanhamento e análise por equipes externas à SE, EJ 4 - Explicitação dos valores subjacentes à proposição avaliação e EJ5 - Direitos humanos e respeito à diversidade.

Cabe salientar que tais dimensões e padrões foram estabelecidos considerando a literatura da área, mas também as necessidades dos técnicos do núcleo de avaliação da secretaria de educação que solicitou a pesquisa, em uma perspectiva dialogada.

Além disso, as dimensões e padrões estabelecidos foram utilizados para a construção de uma matriz de indicadores, que passou por três fases de validação, com diferentes especialistas e com os técnicos da secretaria de educação (SE) em tela. A cada painel de discussão, a proposta foi revisada e aprimorada, a partir das indicações recebidas por especialistas e técnicos em avaliação.

Em termos operacionais, foram coletadas informações em fontes distintas para o cálculo dos indicadores: documentos e legislação atinentes à avaliação em larga escala em tela, entrevistas realizadas com técnicos da SE e supervisores de diretorias de ensino e informações coletadas diretamente junto a diretores escolares e coordenadores pedagógicos por meio questionário administrado por meio eletrônico. Para cada indicador, foi atribuída pelo menos uma fonte de dados. Os dados coletados em documentos e entrevistas foram classificados, pelos membros da equipe de pesquisa, de acordo com a seguinte escala: 1 - não existe dentro dos limites da pesquisa; 2 - insuficiente; 3 - básico; 4 - adequado; 5 - avançado. Os indicadores cuja “nota” foi definida a partir de uma interpretação do pesquisador foram calculados com peso 1. Os dados que foram coletados diretamente

(questionário) foram calculados com peso 2. Para cada indicador foram atribuídas classificações por dois membros da equipe de pesquisa que, posteriormente, discutiram as diferenças de interpretação e chegaram a um único resultado (1, 2, 3, 4 ou 5) por indicador.

Assim, para cada um dos 58 indicadores foi calculada uma nota final e apresentada também a interpretação qualitativa correspondente à essa nota. Os pontos de corte utilizados na determinação da categoria qualitativa dos indicadores foram:

CRITÉRIO UTILIZADO :	
nota final	interpretação
até 1,50	não existe*
de 1,51 a 2,50	insuficiente
de 2,51 a 3,50	básico
de 3,51 a 4,50	adequado
4,51 ou mais	avançado

* Não existe dentro dos limites da pesquisa

Com essas decisões metodológicas, foi possível apresentar os resultados da meta-avaliação, eminentemente qualitativa, de forma sistematizada e quantitativa, o que permite inclusive a análise comparativa ao longo do tempo. Cada dimensão obteve uma pontuação, que propicia que a secretaria de educação perceba quais são os aspectos mais urgentes para se aprimorar, a partir dos resultados da meta-avaliação. Os resultados foram plotados em gráficos de radar, que permitem facilmente identificar os indicadores que devem ser aprimorados.

3 Resultados

A fim de preservar o anonimato da rede de ensino que demandou o processo de meta-avaliação, os dados apresentados foram reconstruídos com dados simulados, a partir dos reais resultados obtidos. A aplicação dos procedimentos metodológicos anteriormente explicados propiciou o cálculo de uma nota para todas as dimensões e padrões propostos. Em geral, a maioria dos indicadores obteve nota insuficiente (42,1%) ou básica (42,1%), o que indica que o Sistema de Avaliação em tela precisa se adequar para atender aos parâmetros estabelecidos.

O exemplo apresentado na Tabela 1 refere-se ao padrão Ética e Justiça, pois devido ao limite de páginas desse texto não é possível apresentar os resultados em todos os atributos e padrões elaborados.

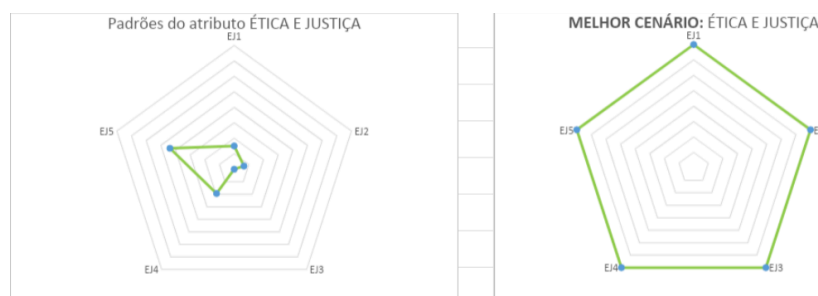
A Figura 1 exemplifica como os resultados foram apresentados graficamente, de forma simplificada, a fim de constituir ferramenta útil aos técnicos de avaliação envolvidos do processo de meta-avaliação.

Tabela 1 – Notas finais dos padrões utilizados na dimensão ética e justiça simulados para a Prova Fictícia P-Alfa.

	Padrão	Nota	Avaliação
EJ1	Transparência na divulgação da documentação da P-Alfa	1,75	Insuficiente
EJ2	Acompanhamento e análise da avaliação por equipes internas à SE	1,33	Não existe
EJ3	Acompanhamento e análise da avaliação por equipes externas à SE	1,00	Não existe
EJ4	Explicitação dos valores subjacentes à proposição da P-Alfa	1,97	Insuficiente
EJ5	Direitos humanos e respeito à diversidade	3,19	Básico

Fonte: Elaboração dos autores.

Figura 1 - Representação gráfica das notas obtidas na Prova Fictícia P-Alfa (simulado) nos padrões do atributo Ética e Justiça



Fonte: Elaboração dos autores a partir da base de dados simulados da pesquisa. 2025.

4 Conclusões e Considerações Finais

Com o desenho metodológico estabelecido foi possível alcançar o objetivo geral da pesquisa e propor padrões e indicadores adequados às dimensões propostas, que pudessem ser calculados de forma independente.

A partir dos resultados obtidos, entende-se que o meta-avaliador pode selecionar os padrões e indicadores mais relevantes para o aspecto das avaliações em larga escala que deseja analisar, propondo aprimoramentos de acordo com as necessidades e possibilidades políticas e orçamentárias do momento.

5 Referências

- COOKSY, L. J., CARACELLI, V. J. Metaevaluation in Practice: Selection and Application of Criteria, *Journal of MultiDisciplinary Evaluation*, Vol. 6, n. 11, 2009. Disponível em: <http://digital.library.unt.edu/ark:/67531/metacrs9145/>.
- ELLIOT L. G.. Meta-avaliação: das abordagens às possibilidades de aplicação. *Ensaio: avaliação e políticas públicas em Educação* [Internet]. 2011, Oct; vol 19 n. 73, p. 941–64.
- RUTKOWSKY, D. *Improving international assessment through evaluation*. 2017. Manuscrito.
- YARBROUGH, D.B., SHULA, L.M., HOPSON, R.K., CARUTHERS, F.A. *The Program Evaluation Standards: a guide for evaluators and evaluation users* (3rd. ed). Thousand Oaks, CA: Corwin Press, 2010.