

PROGRAMAS DE AVALIAÇÃO EM LARGA ESCALA: ALGUMAS CONSIDERAÇÕES¹

Heraldo Marelim Vianna

Doutor em Psicologia da Educação pela Pontifícia Universidade Católica de São Paulo/PUC-SP. Pesquisador Sênior do Departamento de Pesquisas Educacionais da Fundação Carlos Chagas, São Paulo, SP.

Resumo

A partir de entrevistas não-estruturadas com especialistas, o artigo analisa a complexidade dos procedimentos tecnológicos e gerenciais para elaboração e aplicação de programas de avaliação em larga escala, apresentando uma visão geral e abrangente. O artigo mostra que a exequibilidade de um programa dessa natureza baseia-se na consideração de diferentes aspectos, como a existência de uma estrutura que atenda às expectativas dos interessados no programa, ressaltando que a avaliação em larga escala é uma atividade cooperativa entre técnicos e especialistas, visando à superação de numerosos problemas, que são examinados e discutidos.

¹ Artigo apresentado à 27th Annual IAEA Conference (International Association for Educational Assessment) - May 6-11, 2001, Othon Plaza Hotel, Rio de Janeiro, Brasil.

Introdução

O presente trabalho foi elaborado a partir de entrevistas diretas e não estruturadas com onze especialistas em diferentes áreas e resulta de vivências pessoais do autor, não pretendendo refletir posições institucionais. Analisa a complexidade dos procedimentos tecnológicos e gerenciais para a elaboração e aplicação de programas de avaliação em larga escala, incluindo, no conjunto, avaliação de rendimento, avaliação de sistemas, avaliação em processos seletivos, entre outras modalidades de avaliação, sem, entretanto, detalhar especificidades. É uma visão geral e abrangente. O artigo mostra que a exequibilidade de um programa de larga escala baseia-se na consideração de diferentes aspectos, como a existência de uma estrutura que atenda às expectativas dos interessados no programa. A avaliação em larga escala, acentua o trabalho, é uma atividade cooperativa entre técnicos e especialistas, visando à superação de numerosos problemas, que serão examinados e discutidos a seguir.

Avaliação e o problema da qualidade

Um sistema de avaliação em larga escala não se improvisa, exige experiência, juntamente com uma estrutura básica de funcionamento, controle gerencial e monitoramento que garantam a natureza do produto. É necessário destacar que o programa deve ser baseado em um sistema operacional que assegure a eficiência dos serviços e o valor do produto. O trabalho, nesse tipo de avaliação, precisa ser monitorado de forma a garantir uma produção de qualidade em todos os estágios do seu desenvolvimento. Sem a natureza sistêmica das operações tudo pode resultar em um imenso caos.

É imperativa, no caso de avaliações em larga escala, a existência de setores especializados que ofereçam novas dimensões ao trabalho e apresentem soluções eficientes, rápidas e econômicas. Entretanto, a segmentação institucional pode resultar em fracasso do programa se não houver a integração dos diferentes segmentos e o controle das atividades de forma harmoniosa e interativa. É preciso, insistimos, que haja entre os diversos fazeres a interface do "controller", que elimina ruídos de comunicação, sana atritos e medeia conflitos entre as partes. São ocorrências freqüentes em avaliações como as que discutimos, em que todas as etapas do processo precisam ser controladas e mediadas.

A qualidade antes referida resulta, assim, de um conjunto de adequações e articulações, ressaltando-se, entretanto, que cada programa de avaliação tem suas especificidades e demandas diversificadas.

Início de um programa: as partes contratantes

A licitação, muitas vezes ponto de partida de uma avaliação, ocorre sempre que há o envolvimento de algum dos órgãos dos vários níveis governamentais (federal, estadual e municipal). No caso de avaliações ligadas a instituições privadas há uma simples contratação, decorrente da discussão entre as partes interessadas. A licitação gera um termo de compromisso contratual, que tem regras próprias de natureza impositiva que determinam um contrato de adesão sem admitir discussões. O contrato pode, assim, ser problemático, porque as definições costumam ser vagas e suas proposições revestem-se de grande subjetividade, gerando, dessa forma, múltiplas dificuldades para a interpretação dos interessados e monitoramento das ações. Nas licitações, os prazos extremamente exíguos criam problemas de diferentes naturezas para o gerenciamento dos programas e colocam em risco todo o processo. Por outro lado, muitos contratos decorrentes de licitações exigem a transferência de tecnologia, com a cessão dos pacotes tecnológicos para a instituição contratante.

Questão relevante em programas de larga escala refere-se ao desempenho da parte contratante, sobretudo quando instituição privada, que nem sempre define exatamente o que deseja e nem dispõe de conhecimentos que possibilitem uma explicação mais detalhada da metodologia a ser utilizada ou seus fundamentos. Por outro lado, as discussões tornam-se desgastantes pela solicitação de instrumentos especiais de avaliação, que representam aumento nos custos, e dificultam a execução do trabalho. Outros problemas podem ser apontados na fase inicial de uma avaliação em larga escala, como, por exemplo, definição dos objetivos, perfil desejado dos avaliandos, solicitação de diferentes cruzamentos de variáveis, multiplicidade e extensão dos questionários, transmissão de informações incompletas, dificuldade nos trabalhos de equipe entre as partes contratantes e mudanças de parâmetros, entre outros.

A logística dos programas: algumas situações

Os problemas logísticos de uma avaliação como a que consideramos são os mais diversos, dependendo do número de sujeitos e dos instrumentos usados no processo. Várias hipóteses se configuram: (1) número grande de pessoas, locais de aplicação distintos, mas uma única área geográfica e instrumentos específicos; (2) número elevado de sujeitos, em áreas distintas dispersas em territórios diferentes, com instrumentos variados para diferentes fins e (3) um número pequeno de sujeitos, avaliados para diferentes situações com um número bastante diversificado de instrumentos. A operacionalização dessas situações implica aumento de responsabilidade dos "controllers" e demanda eficiência gerencial.

O trabalho a ser desenvolvido é feito sob pressão, não apenas em função dos grandes números e da exigüidade dos prazos, mas também face ao comportamento do contratante dos serviços e à atuação dos interessados nos resultados do processo. A situação é bastante complicada porque nem sempre os avaliandos lêem com a devida cautela o documento definidor das características do processo, que inclui normas e procedimentos a serem seguidos pelos próprios interessados. O desconhecimento desses elementos representa um complicador durante o processo, com a criação de situações que podem ser críticas se não controladas, gerando-se conflitos entre as partes interessadas, que demandam certa criatividade gerencial e intervenção dos "controllers".

Um problema gerencial a ser apontado, ainda que, na visão de leigos, possa ser considerado um aspecto menor, refere-se à montagem de um cronograma de divulgação dos resultados. Ocorrem pressões de diferentes fontes. O cliente deseja conhecer os resultados com o máximo de presteza, esquecendo-se das exaustivas operações numéricas que devem ser processadas. Os participantes, por sua vez, pretendem uma divulgação imediata. Ambos ignoram que esse é um momento crítico no processo e em seu gerenciamento. Apesar da informatização, há possibilidades de erros humanos, impondo-se revisões e verificação da consistência dos resultados. Também ocorrem falhas, ainda que mais raras, no processamento. Além disso, a não observância de parâmetros, impõe a conferência e, muitas vezes, a reconferência dos procedimentos e dos dados que não podem ser omitidas, constituindo-se ambas pontos de preocupação gerencial.

Avaliação – trabalho cooperativo

Um programa da natureza do que ora tratamos exige, no seu planejamento, a definição precisa dos seus objetivos e das metas a alcançar para delineamento das características dos instrumentos e a definição dos recursos humanos, inclusive técnicos com a “expertise” necessária para realizar o empreendimento. Tudo isso, entretanto, precisa ser suplementado por uma estrutura burocrática e logística bem definida para atender às expectativas das partes interessadas e à eficiente ação monitorial.

Ao nos referirmos a técnicos, englobamos especialistas identificados com a moderna teoria das medidas educacionais (que a partir da década de 60 sofreu grandes transformações), analistas de sistemas, estatísticos com experiência nas novas metodologias de análise e especialistas em diversas áreas substantivas identificados com os diferentes meios e processos de avaliação. Queremos, assim, ressaltar que um programa de avaliação resulta da cooperação e da interação de diferentes capacidades e diversos tipos de vivências. A avaliação em larga escala exige a colaboração de uma equipe para a elaboração de “tailor-made tests”, instrumentos específicos para atendimento de situações particulares, como geralmente são as avaliações em larga escala.

Uma avaliação, qualquer que seja a sua natureza, demanda a pré-testagem dos instrumentos, a fim de adequá-los aos sujeitos integrantes do conjunto avaliado. Isso, naturalmente, exige que se tenha uma amostra representativa, o que em muitas situações constitui-se num quadro bastante problemático por não atender a princípios definidores, dando margem a discussões, com o envolvimento do grande público, e o comprometimento da validade de todo o processo, que passa a não merecer a credibilidade da sociedade.

Queremos lembrar que um programa de avaliação, mesmo não sendo de larga escala, exige o envolvimento de diferentes grupos num trabalho de inter-relações, impondo-se, portanto, um gerenciamento competente, que saiba monitorar os problemas e conflitos de relações humanas, que sempre ocorrem e podem levar ao fracasso das atividades.

Avaliação – o processo gráfico e seus vínculos

A natureza do trabalho executado gera múltiplas ações no processo gráfico do instrumental. Foi superada a fase do fotolito, passando à do laser, que facilitou as operações, permitindo agilização dos trabalhos. A era do off-set também está em declínio, com os avanços da produção gráfica. Isso não significa, entretanto, que os problemas tenham diminuído; ao contrário, os trabalhos passaram a ser em uma escala maior, ficando sujeitos a diversas negociações entre as diferentes equipes atuantes no processo global a fim de assegurar compromissos de prazos. Tudo isso demanda uma ação eficiente dos "controllers" de cada segmento na suas atuações gerenciais, em face à diversidade dos chamados clientes internos responsáveis pela materialização do processo que, conforme acentuamos, não é, em absoluto, "one man/woman show". A área da impressão, para atendimento dos clientes externos, estabelece estreitos vínculos de informações com as de planejamento, logística, medidas, processamento, administração, controladoria e financeira, entre outras, que têm diferentes missões e metas a atingir, justificando-se, pois, a interação das equipes para a realização de vários programas simultaneamente, quase sempre em larga escala. É evidente que costuma ocorrer alguma interferência nas comunicações provenientes de várias fontes, cujo fluxo é muitas vezes interrompido. Há, ainda, acúmulo de prioridades, que gera gargalos de congestionamento no processo, determinando intensa atuação gerencial para solução das situações conflituosas que, em princípio, podem comprometer a qualidade do produto, com repercussões no processo de avaliação, exigindo, desse modo, a intervenção dos níveis superiores de gerenciamento. Uma instituição dedicada a programas de avaliação em larga escala precisa planejar seus trabalhos gráficos e estabelecer prioridades, sabendo que a inclusão de um novo programa no conjunto das suas atividades tem implicações na quantidade de horas de trabalho e na definição de custos, considerando o tempo/máquina e o tempo/recursos humanos.

Ainda do ponto de vista da gerência de um programa, a impressão dos instrumentos tem suas sutilezas, donde a necessidade da participação de pessoal competente em informática, pois, freqüentemente, a diversidade de conteúdos substantivos aplicados a um número reduzido de sujeitos, espalhados na amplidão territorial,

não justifica impressões em off-set, que oneram o processo, impondo-se o uso do laser e até mesmo da impressão por computadores. Outra questão de gerência centra-se na exigência de personalização, ou seja, a identificação dos sujeitos, apesar da sua distribuição aleatória em diferentes áreas de especialização no locais de aplicação. O certo é que o gerenciamento de programas, por sua complexidade, inclusive operacional, demanda cuidados especiais, análise de cada projeto que compõe o programa e previsões para situações futuras.

Avaliação – situações conflituosas

As solicitações de programas em larga escala, pensando no caso brasileiro, quase sempre partem de instituições governamentais; contudo, instituições de direito privado também realizam avaliações envolvendo um número elevado de indivíduos. As discussões contratuais costumam ser extremamente difíceis do ponto de vista gerencial. No primeiro caso, envolvendo instituições públicas, o ponto de partida é uma licitação, como vimos anteriormente, que por ser um documento de adesão não admite alterações e deve ser observado estritamente. O envolvimento de instituições privadas em processos de avaliação resulta de negociações bastante conflituosas. Assim, surgem solicitações no sentido de avaliar diferentes indivíduos, com vistas a objetivos conflitantes, usando os mesmos instrumentos; outras, dentro do espírito iniciado na Inglaterra durante a era Thatcher, pretendem, em função dos resultados, fechar instituições – o que realmente não se concretiza dentro do contexto brasileiro - ou, ainda, atribuir prêmios ou bônus e fazer o “ranking” das melhores instituições. Tudo isso traduz uma concepção ultrapassada de avaliação e gera certo desconforto naqueles que, por sua formação e atuação profissional, acreditam que a avaliação seja parte de um processo para identificar problemas que devem ser superados por meio de decisões próprias. Outro aspecto que constitui sério impasse gerencial, e problematiza a ação dos “controllers”, centra-se no excesso de solicitações não previstas contratualmente que não podem ser atendidas por impedimentos financeiros, mas que, “noblesse oblige”, acabam sendo atendidas, o que representa um desgaste gerencial.

Avaliação e testes "ad-hoc"

Problema gerencial extremamente complexo centra-se na deficiência do banco de itens institucional em certas áreas substantivas, impossibilitando que, em tempo recorde, itens sejam construídos, pré-testados e analisados para a construção de testes "ad-hoc". A atribuição gerencial decorre da necessidade da contratação e treinamento rápido de novos especialistas para a elaboração dos itens e constante monitoramento dos trabalhos. As situações se tornam mais complicadas quando as instituições contratantes apresentam tabelas de especificação de conteúdos que fogem inteiramente à realidade educacional, gerando problemas ligados à validade de conteúdo. Outras vezes, a complexidade aumenta, quando são solicitados instrumentos que meçam exclusivamente competências e não conteúdos, ignorando que, por intermédio dos conteúdos, medimos algo mais, habilidades e competências, conforme Benjamin Bloom.

As situações muitas vezes se complicam e se revelam difíceis de gerenciar e monitorar, com a solicitação de instrumentos diversos dos estabelecidos contratualmente, dificultando a operacionalização do programa. Além da mensuração de aspectos cognitivos, solicitam-se, por vezes, medidas ligadas a representações sociais dos professores, capacidade de liderança dos diretores, habilidades sociais, interesses, auto-estima, atitudes, entre outras mensurações que nem sempre estão previstas contratualmente e nem se ajustam aos objetivos do programa inicialmente definido. Há, com freqüência, preocupação de testar o máximo de atributos, avaliar tudo por intermédio de um único projeto, nos limites de um número limitado de itens e de instrumentos. Esquecem-se que os instrumentos abrangem amostras de comportamento que nos possibilitam inferir sobre determinados traços julgados importantes. À gerência cabe, juntamente com técnicos especialistas no assunto, exercer, assim, as funções de monitoria, a fim de evitar esse detalhismo concentrado em aspectos não inteiramente relevantes para a situação ou que levam a conclusões nem sempre valiosas para os objetivos da avaliação. Além das restrições anteriores, queremos situar a questão dos questionários, pelo número elevado de questões, muitas vezes repetitivas, enfocando assuntos já explorados em outras avaliações sobre a mesma população de sujeitos.

Avaliação e questões de processamento

Um dos importantes subprodutos de um programa, como o que ora é considerado, consiste na geração de dados com informações resultantes de análises estatísticas e de cruzamentos de variáveis, o que não é viável quando ocorrem omissões de planejamento, ausência de monitoramento e falhas de gerenciamento. A análise dos dados e o excesso de informações solicitadas têm suas implicações gerenciais. As mais diversas análises costumam ser solicitadas e dados são pedidos sem que haja, posteriormente, utilização em relatórios e pesquisas. Certos dados pretendidos, ainda que simples, exigem procedimentos que implicam o uso de programas estatísticos complexos, sem que o solicitante considere as implicações na divulgação dos dados para certo nível de clientela e a elevação dos custos no projeto.

Sem um gerenciamento adequado, associado a constante monitoramento, com a intervenção dos "controllers", seria impossível o cadastramento de todos os sujeitos. Informações oriundas de diferentes fontes, usando grande diversidade de equipamentos capazes de gerar dados consistentes, requerem um trabalho minucioso. Dados cadastrais imprecisos geram diferentes problemas e comprometem aspectos logísticos da avaliação e elementos administrativos fundamentais para a coleta de informações. Isso, por sua vez, perturba a cronologia dos prazos, gerando mudança de datas e postergando a divulgação dos resultados. Tudo influencia, desde o simples erro de digitação, que pode ter conseqüências desastrosas, a dados deficientes, provenientes de fontes várias, que devem ser convertidos a um único banco de dados. Sem a atuação eficiente dos "controllers", monitorando em várias áreas diferentes aspectos do projeto, corre-se o risco de ter dados comprometidos capazes de provocar a nulidade de todo o processo. A informatização dos dados é um trabalho diversificado, realizado com cautela, sob a supervisão da gerência e constante monitoramento, pois cada evento tem as suas peculiaridades e necessita de diferentes materiais, sendo impossível reaproveitar ferramentas, em função das diferentes lógicas de cada programa.

A ação da informática em um projeto de avaliação, como o ora considerado, sob a gerência do responsável e a intervenção de monitores, parte da verificação de todo o seu equipamento para evitar dificuldades ligadas às máquinas, na fase de execução, que podem eventualmente ocorrer, e falhas humanas. Há, assim, todo um

procedimento gerencial e de monitoramento bastante cauteloso e que tem continuidade durante o uso de diferentes e indispensáveis pacotes estatísticos, visando sempre à consistência dos resultados que, posteriormente, serão submetidos a outros tratamentos.

Avaliação e problemas de terceirização

Ainda que hoje em dia a tendência seja no sentido da terceirização de serviços, apenas algumas fases de uma avaliação de larga escala admitem esse procedimento, por uma questão de prudência e bom senso, pois, algumas vezes, além do sigilo dos dados, há problemas outros tais como interesses institucionais que recomendam cautela, em face à dificuldade na formação de um banco de dados, pela inobservância de normas comuns pelas várias fontes transmissoras em projetos que cobrem grandes áreas. Há necessidade de controle gerencial e constante monitoramento, acompanhado de estudos, para a freqüente revisão dos dados que muitas vezes são transmitidos por pessoas que desconhecem o uso adequado dos meios de passagem e o fazem de forma descuidada, sem um mínimo de documentação que indique os procedimentos, sobretudo quando os dados são provenientes de diferentes fontes. Esse é um problema sério, que exige gerenciamento e cautelas no caso de terceirização. A dependência de fontes externas pode resultar, desse modo, em dados pouco fidedignos, em face a vários tipos de falhas, que não devem, em princípio, ocorrer, sendo exemplo típico a transmissão de informações fragmentadas, uma vez que todos os elementos são igualmente importantes para a realização do processo. A ocorrência de fatos como os apontados determina freqüentes modificações do cadastro, dificultando o trabalho dos analistas. Acrescente-se a isso, o envio de informações conflitantes, divergentes, tornando-se difícil a elaboração dos resultados e o cumprimento dos prazos, com prejuízo para a gerência do programa.

Avaliação e novas metodologias

A partir dos programas de avaliação nacional, instituídos pelo Ministério da Educação – MEC, no Brasil, no início dos anos 90, as metodologias clássicas de análise dos resultados foram sendo substituídas de modo gradativo pela Teoria da Resposta ao Item, pelo

menos nos centros de investigação ligados à avaliação em grande porte. A análise clássica, entretanto, não desapareceu inteiramente, continua a ser feita para um primeiro entendimento dos resultados obtidos. Outras análises também são feitas, como a análise do viés dos itens e o estudo da relação entre diversas variáveis. Apesar do controle gerencial e do constante monitoramento de todo o processo, muitas vezes diferentes projetos se superpõem e é impossível a realização de outros estudos estatísticos mais elaborados, como, por exemplo, o uso de modelos hierárquicos lineares. As instituições interessadas deixam de entender a demora dos resultados, acreditam que os computadores fazem tudo e rapidamente. Há fatores adversos como a não convergência dos dados, a modificação dos parâmetros, entre outros, como o fato de que pacotes estatísticos fechados são usados e muitas vezes não resolvem certas situações, obrigando os analistas a desenvolverem programas específicos para a solução de casos especiais; outras vezes, análises que carecem de finalidade são solicitadas, sendo desprezadas ou mal aproveitadas informações ricas de sentido. Fazem-se solicitações quase que burocráticas, esquecendo que há cálculos extremamente complexos a fazer em função do número de dados e que o pedido determina a elevação dos custos.

Análises complexas podem ser feitas, mas é evidente que há necessidade da filtragem de algumas variáveis. O retorno do aproveitamento das informações do programa é demorado e parcial. Apesar da ação atuante da gerência, da participação dos "controllers" e constante monitoramento, uma conclusão é possível: malgrado o desenvolvimento das máquinas e a sofisticação dos pacotes estatísticos, sente-se que os dados não são devidamente explorados e aproveitados.

Avaliação, estrutura de empresa e divulgação

Um sistema de avaliação como o que estamos considerando existe dentro de uma estrutura de empresa; dessa forma, deve-se agir como tal, gerenciando e controlando inclusive as despesas, que são de duas naturezas:- as despesas de manutenção e outras necessárias ao funcionamento institucional; e as despesas do programa, que são mais fáceis de controlar e devem seguir de forma estrita os procedimentos de controle. A facilidade de gerenciamento e controle das despesas do projeto resultam da existência de um orçamento com as devidas

especificações e de relatórios específicos para cada fase da programação, tendo em vista a prestação de contas ao final.

Assim, uma avaliação em larga escala deve ser planejada em todos os seus momentos, que são gerenciados, acompanhados pelos "controllers" e monitorados no sentido de obter o máximo de qualidade. Ao término, o programa passa por uma fase crítica, a da divulgação dos resultados, o que deve ser feito considerando a diversidade do perfil das várias clientelas interessadas. Dessa forma, diferentes relatórios, dirigidos a cada um dos públicos, devem ser elaborados, sob pena de comprometimento da validade de todo o processo, pela falta de compreensão das informações, e o surgimento conseqüente de dúvidas quanto à qualidade da avaliação, fato esse que reduz substancialmente o impacto esperado e não se alcançam mudanças nas formas de pensar e agir, como seria desejável.

Conclusões

O trabalho mostrou a importância da infra-estrutura de sustentação das atividades de um programa de larga escala e a atuação integrada de técnicos e especialistas, sem a qual é impossível um instrumental válido e que garanta a confiabilidade (*reliability*) dos resultados. Ficou comprovada a necessidade da fixação dos objetivos da avaliação para a concretização das metas a atingir no final do programa. Mostraram-se diferentes situações que devem ser harmoniosamente resolvidas para evitar momentos problemáticos ao longo do processo. A avaliação é um trabalho integrado de diferentes grupos, conforme a ênfase apresentada, e ficou clara a questão da produção gráfica e suas interligações com diversas áreas em busca da convergência dos objetivos. Os problemas de medidas, as questões de processamento e a problemática da terceirização mereceram análises e discussões ao longo de todo o processo. Debates sobre as questões metodológicas entre as partes contratantes foram destacadas, e procurou-se ligar a validade do processo à divulgação dos resultados da avaliação.