

# O Uso da Avaliação na Reforma Educacional<sup>1</sup>

Thomas Kellaghan

---

## RESUMO

Neste ensaio, o uso da informação sobre o desempenho dos estudantes é examinado no contexto de: (i) exames administrados e controlados externamente; (ii) avaliações nacionais; (iii) avaliações internacionais comparativas; (iv) avaliação de escolas; e (v) avaliação baseada em sala de aula. São apresentados resultados com relação aos efeitos do uso da informação sobre a política educacional, os currículos, o que é ensinado nas escolas e os processos cognitivos dos estudantes. Os efeitos podem ser positivos ou negativos, pretendidos ou não, e são mais evidentes quando ao desempenho do estudante é atribuída uma grande importância. Embora a maioria dos recursos para usar a avaliação com o intuito de melhorar a qualidade da aprendizagem tenha sido investida em avaliações

envolvendo agentes externos, o grande potencial de seu uso parece estar no aperfeiçoamento das capacidades de avaliação dos professores.

### **Thomas Kellaghan**

*Ph.D em Psicologia,  
Queen's University of  
Belfast, Reino Unido*

*Diretor do Centro de  
Pesquisas Educacionais,  
St. Patrick College,  
Dublin City University,  
Dublin, Irlanda*

**Palavras-chave:** Avaliação de Escolas – Avaliação Baseada em Sala de Aula – Avaliações Internacionais de Desempenho de Estudantes – Avaliações Nacionais – Exames Públicos.

## Introdução

Há modismos na área de educação que dominam a agenda reformista por algum tempo. O ônus de aperfeiçoar a eficácia e a eficiência do que ocorre nas escolas é depositado em um aspecto específico da tarefa educacional, o qual pode ser isolado e proposto como uma panacéia para os males educacionais. É possível, por exemplo,

---

<sup>1</sup> Trabalho apresentado na 27ª Conferência Anual da Associação Internacional de Avaliação Educacional. Rio de Janeiro, 7 - 11 de maio de 2001. Tradução por Maria de Lourdes Sette. Tradução de: The use of assessment in educational reform.

pensar na educação centrada na criança, na educação secundária abrangente, na orientação vocacional da educação, no aumento do envolvimento e das escolhas dos pais, e na adaptação do currículo às condições locais. Todos esses aspectos foram, em alguma época, propostos como reformas de grande porte. Algumas delas ficaram confinadas a partes específicas do globo, outras tiveram um apelo mais universal. Uma reforma que dominou a cena educacional nas últimas duas décadas, que atrai o interesse do mundo todo (ou pelo menos o das agências internacionais), é a que deveria despertar o interesse dos membros da IAEA (*International Association for Educational Assessment*), uma vez que ela propõe o uso da avaliação para melhorar os padrões educacionais (Torrance, 1995).

Hoje, desejo aproveitar esta oportunidade para refletir sobre alguns aspectos das propostas de reforma da avaliação, sobre a probabilidade de serem bem-sucedidas, e sobre os problemas associados a elas. As propostas variam consideravelmente, porém todas enfocam os resultados da educação (Kellaghan, 2000). No passado, o foco principal da avaliação da qualidade da educação era nos insumos (taxa de participação dos estudantes, instalações físicas, materiais curriculares, livros e treinamento de professores). E, por mais deficientes que esses insumos sejam em muitos países, eles não são mais o foco principal. Hoje, a questão formulada por muitos analistas e legisladores é: ao passarem pela escola, os estudantes estão adquirindo conhecimentos, habilidades, comportamentos e atitudes apropriados? Uma

das maiores e, provavelmente, uma das mais influentes declarações de preocupação com os resultados encontra-se no documento adotado pela *World Declaration on Education for All* em Jomtien, Tailândia, em março de 1990. Essa declaração afirmava que a oferta de educação básica para todas as crianças seria significativa apenas se estas, de fato, obtivessem conhecimento útil, capacidades de raciocínio, habilidades e valores. Como consequência, o Artigo 4 *World Declaration on Education for All* (1990), afirmava que o foco da educação básica deveria ser "na aquisição verdadeira de conhecimento e em resultados, em vez de apenas na matrícula, na participação contínua nos programas organizados, e na conclusão das exigências de certificação" (p.5).

Três grandes questões referentes aos resultados dos estudantes podem ser identificadas como motivo de preocupação. A primeira é que muitos estudantes, sobretudo aqueles com carreiras escolares curtas, parecem tirar pouco proveito da experiência educacional. A segunda refere-se à percepção de que a qualidade está se deteriorando em muitos países. Isso é atribuído ao fato de que há menos recursos investidos por estudante, devido à diminuição das verbas e ao aumento do número de estudantes. E a terceira questão é que mesmo os estudantes que permanecem no sistema educacional por um longo tempo podem não adquirir as competências adequadas para atender às necessidades da economia competitiva e global do século XXI, a qual é baseada em informação.

Ao atacar os problemas associados à qualidade de aprendizagem do estudante, a maior parte dos recursos foi investida em avaliações envolvendo agentes externos. As razões para tal parecem residir no custo relativamente baixo de tal avaliação (comparado, por exemplo, com o desenvolvimento profissional dos professores); no fato de que ela pode ser exigida de fora para dentro; na possibilidade de ser estabelecida e implementada em um curto espaço de tempo; e na visibilidade dos resultados (Linn, 2000).

Ainda que aceitemos os objetivos dos reformistas (e há muitos que não o fazem), podemos perguntar: a avaliação pode ser o santo graal buscado pelos cavaleiros contemporâneos da reforma educacional? Ou, adotando uma outra metáfora, a avaliação se tornará o talismã que operará milagres no aperfeiçoamento dos padrões educacionais, fornecendo aos estudantes os conhecimentos e as capacidades de que necessitam em uma economia global, garantindo a competitividade econômica das nações? Para tentar responder a estas questões, examinarei os indícios relativos aos efeitos de uma gama de formas de avaliação. Antes de fazê-lo, contudo, é recomendável refletirmos sobre a natureza e a complexidade da avaliação, não apenas para esclarecer aquilo que discutirei a seguir, mas também para avaliar os indícios, uma vez que poderíamos esperar que a forma da avaliação e as condições nas quais é praticada sejam relevantes para uma avaliação de impacto.

## A Natureza e a Complexidade da Avaliação

O termo *avaliação* liga-se ao latim *assidere* que significa “sentar ao lado”, um significado que evoca um quadro um tanto simpático da atividade como sendo aquela em que o tutor senta ao lado da criança e guia seu aprendizado. O significado dessa palavra no uso contemporâneo expandiu-se bastante, claro, e em alguns casos até sugere um antagonismo em vez de um relacionamento de cooperação entre o avaliador e o avaliado. Ademais, hoje, o termo é usado não apenas para descrever atividades relacionadas à aprendizagem de estudantes, mas também, no nível institucional ou do sistema, para descrever o desempenho de escolas ou de sistemas escolares.

A avaliação muda de acordo com várias dimensões. Primeiro: ela varia na forma. A avaliação pode ser escrita, oral, prática, ou por observação. Ela tanto pode ser baseada no desempenho real dos estudantes em uma tarefa, como em uma série de respostas curtas, apresentadas em forma de múltipla escolha ou discursiva. Segundo: a avaliação varia em sua formalidade, abrangendo desde situações nas quais um indivíduo responde a uma série de questões ou estímulos em condições controladas e em um período de tempo limitado, até situações nas quais o comportamento de um indivíduo é observado de uma forma não-intrusiva ao longo de um período de tempo extenso. Terceiro: a avaliação tem várias metas: diagnosticar as dificuldades de aprendizagem do estudan-

te; motivar sua aprendizagem; avaliar a eficiência do ensino; certificar o progresso do estudante; e selecioná-lo para o próximo nível do sistema educacional. Quarto: a avaliação pode se realizar de uma única vez, ou pode ser baseada em dados coletados durante um período de tempo considerável. Quinto: o local de controle da avaliação pode ser interno ou externo à escola. Um sistema externo pode, no entanto, incluir elementos internos (por exemplo, quando a avaliação baseada na escola contribui para a nota do aluno em um exame público), ao mesmo tempo em que um sistema interno pode incluir elementos externos (por exemplo, quando testes padronizados são usados como parte de um sistema de avaliação dos professores para fornecer um padrão de referência exterior à escola). Finalmente, a avaliação é, com freqüência, associada a sanções. Um alto grau de importância está em jogo quando o desempenho tem conseqüências importantes para um estudante, tais como a formatura, a promoção, a seleção para educação mais avançada ou um emprego, e/ou para seu professor ou a escola que frequênta, tais como prêmios em dinheiro, reconhecimento ou *status*. Um grau menor de importância está presente quando decisões importantes não são tomadas com base no desempenho. No entanto, pouca importância não significa nenhuma importância. É difícil imaginar uma situação de avaliação que não tenha conseqüência alguma para alguém. Por exemplo, a avaliação de estudantes por professores em sala de aula normalmente seria considerada como uma questão de pouca importância. No entanto, tal avaliação pode afetar o autoconceito do estudante, seu nível de motivação, ou

sua posição entre os colegas e, se constantemente negativa durante um longo período de tempo, pode ter sérias implicações para sua aprendizagem e atitude geral para com a escola.

Com base na complexidade dessa taxonomia, desejo considerar agora os efeitos da avaliação. Eles serão considerados segundo cinco categorias não total e mutuamente excludentes: exames administrados e controlados externamente; avaliações nacionais; avaliações internacionais; a avaliação das escolas; e a avaliação do estudante pelo professor em sala de aula.

## Exames Administrados e Controlados Externamente

Por exames administrados e controlados externamente entendem-se os exames públicos ou estaduais tradicionais que constituem, há muito tempo, uma característica dos sistemas de educação da África, da Ásia, e da Europa, assim como os testes estaduais que foram introduzidos recentemente nos Estados Unidos. É interessante notar que vários países da Europa Oriental e da ex-União Soviética introduziram, ou estão introduzindo, esta forma de avaliação. Em sua forma tradicional, os exames públicos são, em geral, usados para seleção e, portanto, grande importância é atribuída ao desempenho dos estudantes. Nos Estados Unidos, muitos resultados, que variam de estado para estado, estão ligados ao desempenho, incluindo a promoção e formatura de estudantes, assim como prêmios para professores e escolas que atendem a certos critérios.

Vários analistas acham que os exames externos podem desempenhar um papel na elevação dos padrões acadêmicos (Eisemon, 1990; Heyneman & Ransom, 1992; Ross & Maehlick (Ed.), 1990; Somerset, 1996). Afirma-se que, caso a qualidade e o escopo desses exames sejam satisfatórios, eles fornecem diretrizes aceitáveis para o ensino, levando a um ajuste de processos instrucionais e de aprendizagem o qual, por sua vez, aumenta a qualidade da educação. Acredita-se que o impacto pode ser aumentado se forem fornecidas às escolas informações sobre como seus estudantes se desempenharam em matérias específicas e mesmo em questões ou itens individuais.

De acordo com evidências empíricas e anedóticas, não há grandes dúvidas de que os exames externos de grande importância afetam o ensino e a aprendizagem (Crooks, 1988; Frederikesén, 1984; Kellaghan & Greaney, 1992; Kellaghan, Madaus & Raczek, 1996; Madaus & Kellaghan, 1992). As conseqüências, no entanto, nem sempre são as pretendidas, tampouco são sempre positivas. Primeiro, os exames podem, pelo menos em parte, nortear os esforços ao tornar explícitos os objetivos da instrução e especificar os padrões de desempenho, de modo que os estudantes e professores tenham objetivos bem definidos para seu trabalho. Isso, é claro, pode ser positivo. Se, por exemplo, em um nível muito geral, atribuírem-se notas para desempenho prático em um exame, isso ajudará a assegurar que tais capacidades práticas sejam ensinadas durante o ano todo. Em um nível mais específico, o uso de questões de ordem superior, em

vez de questões que exijam memorização, tende a aumentar a retenção, transferência de aprendizagem, interesse e desenvolvimento de capacidades de aprendizagem. A influência da avaliação nessa situação tende a ser específica ao conteúdo do teste usado; isto é, onde há um efeito, ele se refere principalmente ao tipo de tópicos e capacidades que fazem parte de um teste.

No entanto, o poder dos exames de grande importância de nortear os esforços dos professores e estudantes pode provocar efeitos positivos e negativos. Se os exames forem limitados àquilo que é facilmente medido, o que não é facilmente medido será desprezado pelas escolas. Se os exames forem todos escritos, as capacidades práticas e orais provavelmente serão ignoradas. Se forem confinados ao formato de múltipla escolha, as capacidades envolvidas na escrita em prosa contínua serão desprezadas. Se a avaliação for um processo externo e terminal, as capacidades de processamento, planejamento e perseverança receberão pouca atenção. O conteúdo e o formato de um exame podem também encorajar o ensino de exercícios de rotina, o conhecimento adquirido por repetição, e o acúmulo de conhecimento factual em vez do desenvolvimento das capacidades de raciocínio e de resolução de problemas.

Ademais, esforços consideráveis serão dirigidos não apenas para o estudo das matérias que aparecerão no exame, mas também para a preparação dos testes, que pode abranger desde a promoção de conhecimentos sobre testes ao ensino do tes-

te. Portanto, muitos esforços serão envidados em estratégias de auxílio a estudantes para que passem no exame, em vez de dominarem o objeto do estudo e afiarem competências duradouras. Como conseqüência, as deduções possíveis sobre o desempenho com base nas notas dos testes reais podem estar distorcidas. É provável (e quase inevitável) que haja uma melhoria no desempenho dos itens de teste entre os estudantes em cujo ensino os processos de instrução e os itens de teste sejam congruentes. No entanto, os dados são conflitantes e colocam em dúvida a possibilidade deste aperfeiçoamento ser refletido no desempenho em outras medidas de desempenho, ou em medidas de compreensão mais geral (Linn, 2000).

Os efeitos de um sistema de exames públicos não estão limitados aos estudantes que participam do exame. O ensino, e em conseqüência a aprendizagem, pode ser afetado em graus menores. Por exemplo, foi observado o uso do formato de múltipla escolha não apenas nas práticas de avaliação dos professores mas em sua metodologia de ensino em salas de aula em anos escolares anteriores àqueles em que tal formato é realizado. Ademais, os estudantes com potencial para notas baixas podem ser excluídos da situação de teste, tanto por escolha própria quanto por decisão das autoridades escolares. As autoridades escolares podem agir retendo os estudantes em séries inferiores ou providenciando para que estejam ausentes no dia do teste caso acreditem que seu desempenho causará um impacto na avaliação de uma classe, enquanto os estudantes que consideram os exames obstáculos insupe-

ráveis podem abandonar a escola.

Talvez o efeito negativo mais óbvio dos exames de alto grau de importância seja a variedade de formas de práticas fraudulentas, incluindo a "cola", a falsidade ideológica, e a alteração de notas, que estão associadas a eles. Uma Comissão do Punjab, no Paquistão, resumiu em uma única frase uma situação extrema quando observou que "o propósito básico da maioria dos estudantes... é obter notas altas por meios justos ou fraudulentos" (*Punjab Commission for Evaluation of Examination System and Eradication of Malpractices* (1992, p.61).

Finalmente, ao considerarmos os efeitos dos exames externos, parecem relevantes os resultados das experiências e dos estudos de campo sobre os efeitos dos objetivos dos estudantes em seus processos cognitivos. Nesses estudos, os processos cognitivos dos estudantes, o que chamamos de "objetivos de desempenho", isto é, objetivos que refletem a preocupação dos indivíduos em demonstrar suas capacidades e obter julgamentos favoráveis ou evitar avaliações negativas de sua competência, foram comparados com os processos de estudantes que adotaram o que chamamos de "objetivos de aprendizagem". Estes objetivos refletem a preocupação em desenvolver novas habilidades e em aumentar a competência, o entendimento e o domínio de algo novo. Se os resultados desses estudos forem aplicados a exames externos de grande importância, algo que pode ser justificado na medida em que os exames incentivam a busca por objetivos de desempenho, pode-se

concluir que os estudantes que fazem exames externos (i) têm menos probabilidade de adquirir capacidades transferíveis e de ordem superior; (ii) têm menos probabilidade de aplicar o ensino auto-regulado e as estratégias de resolução de problemas, e mais probabilidade de enfatizarem a memorização e o ensaio; (iii) têm menos probabilidade de focar o aperfeiçoamento de sua competência e de investir o tempo e os esforços necessários para fazê-lo; (iv) têm menos probabilidade de ver o fracasso como algo que pode ser remediado por uma mudança de estratégia; e (v) têm menos probabilidade de mostrarem preferência por trabalhos desafiadores e por assumirem riscos.

## Avaliações Nacionais

Desejo agora abordar duas áreas inter-relacionadas, a saber, as avaliações nacionais e as internacionais, as quais receberam atenção no contexto do aperfeiçoamento da qualidade educacional. As duas áreas compartilham muitas características de processo (por exemplo, as relacionadas aos métodos de medição e de amostragem). No entanto, elas diferem em várias formas, mais, obviamente, no fato de que os estudos internacionais permitem comparações entre países (Goldstein, 1996; Greaney & Kellaghan, 1996b, Kellaghan, 1996, 1997; Kellaghan & Grisay, 1995).

Uma avaliação nacional é um exercício planejado para descrever o nível de desempenho, não de estudantes individuais, mas de todo um sistema educacional, ou uma parte claramente definida dele (por

exemplo, alunos da quarta série ou de 11 anos de idade). Uma avaliação pode também ser planejada para fornecer provas de desempenho regional ou distrital, e mesmo de escolas individuais, como nos casos de Grã Bretanha, Chile e França.

A maioria dos países industrializados possui sistemas de avaliação nacional há tempos e, durante a década passada, esforços sérios foram envidados para desenvolver a capacidade de administrar avaliações nacionais nos países em desenvolvimento (Kellaghan & Greaney, 2001). O Projeto de Monitoramento da UNESCO-UNICEF, em conjunto com a Iniciativa de Educação para Todos, foi responsável por iniciar avaliações em quase 50 países enfocando as competências de aprendizagem básica em alfabetização, matemática, e capacidades de vida (consciência e conhecimento de saúde, de nutrição, limpeza, e de higiene) após quatro anos de escolarização (Chinapah, 1997). Vinte e quatro países são da África, doze da Ásia (incluindo quatro ex-membros da União Soviética), cinco do Oriente Médio, três da América Latina e do Caribe (incluindo o Brasil) e três da Europa. O impacto do projeto foi maior na África, onde aproximadamente metade dos países do continente se comprometeu a realizar uma avaliação sob os auspícios do Projeto.

O International Institute for Educational Planning (IEEP) tem também participado no desenvolvimento de capacidades para avaliações nacionais no Southern Africa Consortium for Monitoring Educational Quality (SACMEQ), que foi estabelecido em 1993 como uma rede de ministérios de edu-

cação na África Meridional e, desde então, foi expandido para a África Oriental. Quatorze países participaram até agora. Cinco questões principais relacionadas com a política nortearam o desenvolvimento da estratégia de avaliação nacional: (i) a necessidade de dados básicos sobre insumos específicos nas escolas; (ii) as condições nas escolas se comparam com os padrões ministeriais; (iii) a equidade na alocação de insumos; (iv) a identificação dos níveis de desempenho em leitura e matemática dos alunos no nível primário superior (6ª série); e (v) a identificação dos insumos educacionais que estão relacionados com o desempenho dos alunos (Ross, 1995).

Além das avaliações nacionais apoiadas pela UNESCO-UNICEF e pelo IEEP, vários países da África realizaram avaliações com o apoio do Banco Mundial e de outras agências.

Houve um desenvolvimento rápido no estabelecimento das avaliações nacionais durante a década de 1990 nos países da América Latina e do Caribe. As avaliações foram associadas, com frequência, ao fornecimento de dados básicos para reformas educacionais. Hoje, praticamente todos os ministérios de educação na região incorporaram avaliações nacionais às suas agendas, e 22 avaliações foram realizadas em 18 dos 30 países (Rojas & Esquivel, 1998). Existe uma semelhança considerável, mas não uma uniformidade, nas áreas de currículo avaliadas. A linguagem/comunicação e a matemática figuram em todas as avaliações; ciências e estudos sociais ou história e geografia em aproxi-

madamente metade. A maioria das avaliações foi realizada em séries da escola primária (3ª série foi a mais popular, seguida pela 4ª e pela 5ª séries). Um censo foi realizado em cinco países (Brasil, Chile, República Dominicana, Panamá, Uruguai). Na Argentina, foram realizados um censo e procedimentos de amostragem. Todos os outros países usaram amostras representativas.

As evidências disponíveis sobre os efeitos das avaliações nacionais são limitadas, em parte, talvez, porque elas não são em geral documentadas publicamente. Quando amostras anônimas de escolas estão envolvidas, espera-se que os efeitos sejam principalmente no nível da política educacional. Quando todas as escolas estão envolvidas, efeitos nos níveis da política nacional e escolar podem ser antecipados. Considerarei os efeitos no nível da escola ao abordar a avaliação escolar. Em um nível político, na França, foram relatados impactos fortes nos níveis regional e nacional. Os indicadores de desempenho para cada região foram publicados; ao mesmo tempo, um novo currículo da escola elementar e as mudanças na organização das escolas secundárias foram considerados como resultantes das informações obtidas em avaliações nacionais.

## Avaliações Internacionais

As avaliações comparativas internacionais se originaram no final da década de 1950, um pouco antes das avaliações nacionais. Seu crescimento foi motivado pela consciência da falta de padrões de desempenho com validade internacional,



por meio dos quais cada país poderia comparar o desempenho de seus estudantes. Além de tais comparações, também era previsto que os estudos capitalizassem a variabilidade existente nos sistemas educacionais, explorando as condições que “um grande laboratório educacional” de estruturas escolares e currículos variados poderia fornecer, não apenas para descrever as condições existentes, mas para sugerir o que pode ser educacionalmente possível (Husén & Postlethwaite, 1996). Portanto, os estudos foram promovidos por terem uma perspectiva de pesquisa (principalmente na tradição do funcionalismo estrutural), assim como implicações políticas mais imediatas.

A maioria das avaliações internacionais foi realizada sob os auspícios da International Association for the Evaluation of Educational Achievement (IEA). Desde 1961, quando um estudo piloto foi realizado, mais de 60 países participaram em estudos de leitura, matemática, ciências, redação, literatura, línguas estrangeiras, educação cívica e proficiência em informática. A participação dos países em desenvolvimento tem sido menor do que a dos países industrializados. Há várias razões possíveis para isso, incluindo limitações financeiras, de infra-estrutura e de recursos humanos, assim como a percepção de que as diferenças nas condições dos sistemas educacionais torna supérflua a documentação das diferenças em seus desempenhos relativos.

Uma vez que todas as avaliações internacionais usam amostras de escolas em países participantes e dados agregados, esperaríamos novamente que seu impacto fosse,

acima de tudo, no nível político. Há vários relatos sobre o uso de resultados de estudos para apoiar a política e a tomada de decisão, que resultaram em mudanças na prioridade concedida a certas matérias e no conteúdo do currículo (Kellaghan, 1996).

O esforço mais significativo para levantar dados sobre o impacto de um estudo internacional encontra-se no Third International Mathematics and Science Study (TIMSS) (Robitaille, Beaton & Plomp (Ed.), 2000). Vinte e nove sistemas educacionais forneceram informações sobre o impacto dos resultados do TIMSS em nível nacional, a partir dos quais algumas conclusões gerais podem ser tiradas. Primeiro, na maioria dos casos, os resultados foram interpretados no sentido de que havia necessidade de melhorá-los. Mesmo em alguns dos países com melhor desempenho, preocupações com alguns aspectos dos resultados (por exemplo, as atitudes ou o desempenho dos estudantes em itens que exigiam capacidades de pensamento de nível superior) foram expressas. Segundo, os resultados parecem ter tido um impacto maior sobre o público em geral e nos círculos políticos, através dos meios de comunicação de massa, do que sobre os educadores. Terceiro, o TIMSS enfatizou para os legisladores o fato de que as práticas educacionais e os currículos em matemática e ciências eram diferentes em outros países, uma constatação que poderia levar a uma revisão de currículo.

## A avaliação das escolas

As informações sobre o desempenho de estudantes são usadas em vários países

para se obter uma apreciação sobre a qualidade do desempenho de cada escola e mesmo de cada professor individualmente. Isso pode ser realizado no contexto de uma avaliação estadual (por exemplo, nos Estados Unidos da América), de uma avaliação nacional (em vários países), de uma análise de resultados de exames externos (públicos) (por exemplo, Quênia) ou por meio do processo menos padronizado de inspeção escolar, o que é comum em muitos países.

Tanto no Texas quanto no Tennessee, o desempenho acadêmico dos alunos nas áreas básicas do currículo (por exemplo, leitura, matemática, ciências, estudos sociais) é avaliado em todas as séries e, com base nesse desempenho, avalia-se a eficiência dos sistemas escolares de cada escola e de cada professor individualmente. Na Inglaterra e no País de Gales, os alunos de 7, 11 e 14 anos são testados em inglês, matemática e ciências como parte da avaliação nacional do currículo. Os resultados são apresentados em tabelas de desempenho publicadas, e as escolas são classificadas segundo a proporção de estudantes que alcançaram níveis estabelecidos. Em muitos países em desenvolvimento, as escolas divulgam para a imprensa os resultados obtidos em exames públicos como uma forma de estratégia promocional.

A publicação de informações sobre o desempenho de escolas na forma de tabelas de classificação é, com frequência, considerado recurso suficiente para melhorar os padrões, uma vez que se presume que gerará competição entre as escolas, os professores, e os estudantes. Em alguns ca-

sos, medidas adicionais são tomadas. No Texas, se metade da classe tiver um desempenho melhor do que aquele do grupo norma, as escolas e os professores recebem prêmios em dinheiro (Webster, Mendro, Almaguer, 1994). Na Inglaterra e no País de Gales, as autoridades locais, baseadas nos resultados dos testes, auxiliam as escolas na implementação de um "plano de ação" elaborado pela escola para resolver problemas. Se os resultados são muito ruins, é dado às escolas um prazo limite para que melhorem e, caso fracassem, podem ser fechadas ou sofrerem intervenção.

Não há dúvida de que, quando os dados da avaliação são usados para verificar o grau de desempenho das escolas, eles podem causar um impacto no comportamento delas. No entanto, seu uso é complexo, está longe de ser isento de problemas, e pode mesmo ser enganador. Primeiro, o desempenho das escolas (e, portanto, sua posição relativa para com o desempenho de outras escolas) variará dependendo do resultado escolar usado. Isso é verdadeiro mesmo quando são comparados os desempenhos em áreas básicas do currículo (Guskey & Kifer, 1990). Segundo, não obstante a medida de resultado usada, os resultados podem variar de ano para ano. A causa disso pode ser as diferenças nos contingentes dos estudantes, assim como as diferenças entre os professores. Terceiro, se existe interesse em julgar a contribuição da escola para o desempenho, enfrenta-se o problema de como permitir a inclusão de fatores que estão fora do controle da escola e que contribuem para esses desempenhos, tais como o desempenho dos alunos por oca-

ção do ingresso na escola, e o grau de apoio familiar que recebem (Goldstein & Thomas, 1996).

Para confrontar a crítica de que as comparações entre escolas são injustas, uma vez que as escolas diferem nas características de seus alunos e nas circunstâncias em que operam, uma entre duas abordagens pode ser adotada. As comparações podem ser confinadas a escolas que sejam semelhantes nas características socioeconômicas de seus alunos, ou medidas podem ser usadas para calcular o "valor" que as escolas agregaram ao desempenho acadêmico dos estudantes durante um certo tempo (*School Curriculum and Assessment Authority, 1994; Webster, Mendro & Almaguer, 1994*). No entanto, na maioria das vezes, não se aplica nenhum mecanismo compensatório com relação às características escolares quando são comparados os desempenhos dos estudantes nas escolas.

Os problemas se exacerbam quando os resultados da avaliação são colocados em um contexto de alta importância, como, por exemplo, quando são publicados os resultados do desempenho das escolas. Como se viu no caso de exames públicos, nessa situação, os professores tenderão a canalizar seus esforços para o que é mensurável, levando a um estreitamento do currículo e à concentração na obtenção de altos níveis de desempenho em testes, em detrimento de formas de aprendizagem mais profundas. Ademais, os recursos e as estratégias educacionais podem ser concentrados nos estudantes limítrofes para produzir o máximo de melhorias no desempe-

no dos estudantes (por exemplo, aumentar o número de notas C obtidas em um exame público), enquanto os estudantes com desempenho inferior ou superior serão negligenciados. Finalmente, os pais podem reagir aos resultados, escolhendo escolas para seus filhos onde isso seja possível e, portanto, aumentando as dificuldades da escola e dos estudantes que permanecem. O Diretor Executivo da Qualifications and Curriculum Authority, que é responsável por revisar o currículo nacional na Inglaterra e no País de Gales, e por administrar os testes nacionais, forneceu um resumo sucinto da situação: "nomear e envergonhar publicamente é doloroso – as escolas sofrem, os professores sofrem, as reputações das escolas sofrem também – mas, no todo, parece uma resposta inteligente para uma situação de crise" (em entrevista com Tell, 1998).

Uma maneira mais benigna de usar os resultados da avaliação escolar é identificar as escolas que necessitam de assistência (como é feito no Uruguai e na França). Recursos adicionais podem então ser fornecidos e as escolas e os professores podem ser auxiliados na identificação de problemas, estipulação de objetivos, elevação de expectativas e desenvolvimento de estratégias apropriadas. Na França, como vimos, foram divulgados os efeitos de uma avaliação nacional nos níveis regional e nacional. Como a avaliação é compulsória, em todas as escolas, para estudantes de 8, 11 e 15 anos, as informações estão disponíveis não só no nível escolar mas também para cada estudante individualmente e, portanto, é possível uma ação em todos esses níveis. Os pais rece-

bem informações sobre os resultados de seus filhos. Ademais, o estudante que esteja com problemas em francês ou matemática é identificado como estando "com dificuldades" e, conseqüentemente, pode receber ajuda adicional, incluindo sessões de dever de casa supervisionado na escola, ou encaminhamento para um programa de recuperação.

As avaliações nacionais, mesmo quando projetadas para fornecer informações que nortearão decisões sobre a alocação de recursos, podem, no entanto, também ter conseqüências negativas, embora não pretendidas. Por exemplo, no Chile, alguns professores enfocam seu ensino na preparação para a avaliação, e são conhecidos por "ajudarem" os estudantes na avaliação de maneira pouco ortodoxa. Neste país, foram criadas tabelas informais que classificam as escolas em "boas" e "más"; e professores com resultados ruins foram desligados da escola em que trabalhavam (Olivares, 1996).

## A avaliação baseada em sala de aula

A avaliação da aprendizagem dos estudantes em sala de aula é especialmente importante quando consideramos como a avaliação pode ser usada para melhorar a qualidade da educação. Primeiro, a avaliação em sala de aula ocupa um papel central no processo de aprendizagem de todos os estudantes. Segundo, ela é uma característica de todos os sistemas educacionais, embora outras áreas de avaliação não sejam. E, terceiro, sua função é

principalmente formadora; é usada para que decisões sobre o ensino sejam tomadas; ocorre durante o aprendizado (em vez de no momento em que se pressupõe que a aprendizagem esteja concluída); e é planejada para auxiliar os estudantes na tarefa de adquirir conhecimentos e capacidades. Uma publicação britânica oficial aponta algumas regras básicas de avaliação em ensino e aprendizagem:

[A avaliação] pode fornecer uma estrutura na qual os objetivos educacionais possam ser estabelecidos, e o progresso dos alunos mapeado e expresso. Ela pode gerar uma base para o planejamento das medidas educacionais em resposta às necessidades das crianças. Ao facilitar o diálogo entre professores, ela pode ampliar as capacidades profissionais e auxiliar as escolas, como um todo, a fortalecerem o ensino em todo o currículo... (*Great Britain. Department of Education and Science and Welsh Office, 1988*).

Apesar de sua importância central para a aprendizagem, na verdade não sabemos muito sobre como os professores avaliam seus alunos ou como eles usam seu conhecimento para orientar o processo de aprendizagem. Isso não é surpreendente uma vez que, para a maioria, as avaliações dos professores são subjetivas, informais, imediatas e contínuas, envolvendo um monitoramento contínuo do comportamento do estudante, do desempenho escolar, e da sensibilidade para a instrução. Os professores usam vários tipos de evi-

dências. No entanto, eles parecem não prestar muita atenção ao que pode ser chamado de informações "objetivas", por exemplo, as informações obtidas de testes de desempenho, padronizados e referenciados a normas, os quais são muitas vezes administrados nas escolas de forma rotineira. Tampouco o uso dos testes parece causar impacto no desempenho do estudante, exceto quando os testes fornecem informações diagnósticas (Kellaghan, Madaus & Airasian, 1982).

Nosso conhecimento das práticas de avaliação dos professores é ampliado em certa medida pelos dados coletados no TIMSS nos quais os professores de 8<sup>o</sup> série em 38 sistemas educacionais foram solicitados a responder a um questionário que perguntava sobre a frequência com que usavam informações de avaliação para vários propósitos (Madaus & Kellaghan, *no prelo*). A grande maioria, na maior parte dos países, relatou que usava essas informações para dar notas aos alunos, fornecer informações aos alunos sobre seu desempenho, diagnosticar problemas de aprendizagem e planejar as lições futuras. O uso de informações de avaliação para cientificar os pais, para indicar alunos para cursos e para classes separadas por coeficiente de inteligência, enquanto prática muito comum em alguns países, simplesmente não acontecia em outros.

Esperava-se que as práticas de avaliação dos professores causassem um impacto na aprendizagem dos estudantes, e alguns indícios de que o impacto pode ser positivo, a maioria deles coletados nos Estados Unidos da América e na Grã

Bretanha, estão disponíveis (Black & Wiliam, 1998; Crooks, 1988). Talvez seja surpreendente que as análises intrapaíses dos dados do TIMSS, nos quais o uso das informações de avaliação por cada professor em sala de aula foi correlacionado com as notas médias de matemática de seus alunos, revelaram um grau muito pequeno de significância. Esses resultados talvez reflitam a precariedade dos instrumentos de estudo, assim como a falta de variação nas práticas de avaliação dos professores, e não devem levar à conclusão de que as práticas de avaliação dos professores não têm, pelo menos, o potencial de causar impacto sobre a aprendizagem dos alunos.

## Conclusão

Para concluir, vale a pena novamente refletir sobre a complexidade da avaliação no processo educacional. A avaliação tem vários sentidos, é usada em uma gama de contextos, é baseada em muitas técnicas, e serve a vários propósitos. Pode ter efeitos negativos e positivos, conseqüências pretendidas e não pretendidas. Pode ser usada conscientemente para alterar a prática educacional, pode afetar a prática mesmo quando ninguém pretende fazê-lo deliberadamente. Tais considerações, obviamente, deveriam levar-nos a hesitar antes de prescrever a "avaliação" como uma panacéia para lidar com problemas educacionais de grande porte. A possibilidade de a avaliação poder atender ao objetivo mais modesto de melhorar o desempenho do estudante, ou não, dependerá de vários fatores que deveriam ficar claros quando tentamos responder a questões como: Que forma de avaliação pode ter o maior

impacto? Que tipo de aprendizagem desejamos encorajar? Que providências são necessárias para melhorar a capacidade do sistema para fornecer tal tipo de avaliação? Que tipos de problemas devemos antecipar e que providências poderiam ser tomadas para mitigá-los? Como deveriam ser usadas as informações derivadas de uma avaliação? Pode uma forma proposta de avaliação ter conseqüências não pretendidas? Como ela se enquadrará no ensino do currículo existente e em outras práticas de avaliação? Embora os propósitos principais das diferentes categorias de avaliação possam ser especificados, o fato de que as categorias interagem, intencionalmente ou não, também precisa ser reconhecido. Assim sendo, os exames externos e as avaliações nacionais que permitem uma avaliação do desempenho das escolas causam impacto nas práticas dos professores (incluindo suas práticas de avaliação) e na aprendizagem dos alunos.

É necessário reconhecer que o uso de testes de alto grau de importância, tanto os exames públicos quanto as avaliações nacionais e estaduais, para melhorar a qualidade da educação, carrega consigo a probabilidade de provocar conseqüências indesejáveis, embora não pretendidas. Isso não significa dizer que os procedimentos de avaliação em exames públicos em situações em que não haja alternativa óbvia não devam ser melhorados. Certamente existe espaço para melhorias. Os instrumentos de exame em muitos países são de baixa qualidade. Eles não refletem adequadamente o currículo escolar; as formas de avaliação empregadas por eles são extremamente limitadas; eles tendem a focar

a memorização e o reconhecimento em vez de capacidades de nível de raciocínio superior; não refletem a vida cotidiana dos alunos; e são projetados para selecionar os alunos para estudo mais avançado em vez de certificar os desempenhos de todos os candidatos, com todas as conseqüências implícitas para os alunos que não são selecionados. Esforços foram envidados para abordar estas questões em alguns sistemas de exame, os quais esperaríamos que trouxessem alguma contribuição para a qualidade da aprendizagem dos alunos. No entanto, parece irônico que, no momento em que padrões mais altos de desempenho, tais como o entendimento conceitual aprofundado, o interesse na aprendizagem e o compromisso com ela, a valorização da educação e a confiança do estudante em suas capacidades e atributos, estão sendo propostos para todos os estudantes, os meios que estão sendo adotados para atingir esse objetivo provavelmente inibem em vez de facilitar sua realização.

Embora haja pouca evidência empírica de que a avaliação tem um impacto sobre o desempenho do estudante (a maioria dos efeitos está relacionada às condições de aprendizagem), o maior potencial de uso da avaliação para melhorar a qualidade da educação parece estar no aperfeiçoamento das capacidades de avaliação dos professores (as quais exibem muitos pontos fracos). Outras categorias de avaliação podem aspirar a afetar a aprendizagem, mas podem esperar fazê-lo apenas indiretamente. Embora o propósito principal da avaliação escolar seja chegar a conclusões sobre a efi-

ciência das escolas, dos exames (públicos) externos para selecionar e certificar alunos, e das avaliações nacionais e internacionais para descrever o desempenho de um sistema em vez dos desempenhos de cada estudante em particular, a avaliação baseada em sala de aula é a única categoria de avaliação que tem como objetivo principal e imediato facilitar a aprendizagem do estudante.

Se consideramos central o papel do professor na avaliação, então parece razoável investirmos recursos nesta área para desenvolver as capacidades de avaliação dos professores com base em resultados de pesquisas e em uma análise das boas práticas. Seria de se esperar que os professores necessitassem apoio considerável neste processo. Ademais, outros fatores precisam ser considerados, tais como currículos inadequados, manuais e turmas grandes, qualquer um dos quais pode inibir os esforços dos professores para usarem avaliação com o intuito de melhorar a aprendizagem do estudante. É necessário reconhecer a probabilidade de que o investimento no desenvolvimento do professor exija mais recursos do que os investimentos em outras formas de avaliação.

Uma vez que o compromisso com os estudos nacionais e internacionais sobre o desempenho estudantil parece forte e, provavelmente, continuará sendo no futuro próximo, deve-se refletir sobre seu aperfeiçoamento, tanto do ponto de vista técnico quanto do uso dos resultados. Ademais, maior consideração precisa ser dada aos contextos nos quais estão inseridos. Os sistemas educacionais diferem substancialmente em suas estruturas institucionais e na forma pela qual os poderes decisórios e as responsabilidades entre professores, escolas e órgão (s) responsáveis pela educação são alocados. Uma avaliação do papel e do valor dos exames em larga escala exigiria o exame da operação do sistema de avaliação, como ele se relaciona com as estruturas de responsabilidade e de incentivos no sistema, até que ponto mudar aquelas estruturas é desejável e viável, e o provável impacto das mudanças no desempenho do aluno. Ademais, se não forem desenvolvidos os procedimentos para traduzir as informações, derivadas de programas de avaliação em larga escala, em práticas na sala de aula, seu potencial para influenciar a qualidade da educação pode não se realizar e, portanto, poderemos esperar que as reformas atuais sigam o caminho de muitos esforços reformistas anteriores, ou seja, venham a ser substituídos por outra inovação na década seguinte.

## ABSTRACT

*In this essay, the use of the information about the students' performances is analyzed in the context of (i) examinations administered and controlled externally; (ii) national evaluations; (iii) international evaluations; (iv) school evaluations; and (v) evaluations based on classes. Results are presented related to the effects of the use of the information on the educational policy, in the curriculum, what is taught at school, and the cognitive processes of the students. The effects can be positive or negative, intended or not, and they are more evident when we attribute to them a great importance on the students' performance. Although the major part of the resources to use the evaluation with the purpose of improving the quality of the learning process, they were inverted in the evaluations that involve external agents, the big potential of its use seems to be in the improvement of the capacities of the teachers' evaluation.*

**Keywords:** School evaluation – Evaluations based in classes – International evaluations of the students' performance – National evaluations – Public examinations.

## RESUMEN

*En este ensayo, el uso de la información sobre el desempeño de los estudiantes es examinado en el contexto de (i) exámenes administrados y controlados externamente; (ii) evaluaciones nacionales (iii) evaluaciones internacionales; (iv) evaluaciones de escuelas; y (v) evaluaciones basadas en las clases. Son presentados resultados con relación a los efectos del uso de la información sobre la política educativa, los currículos, lo que se enseña en las escuelas, y los procesos cognitivos de los estudiantes. Los efectos pueden ser positivos o negativos, pretendidos o no, y son más evidentes cuando se le atribuye una gran importancia al desempeño del estudiante. Aunque la mayoría de los recursos para usar la evaluación con el intuito de mejorar la calidad del aprendizaje se haya invertido en evaluaciones envolviendo agentes externos, el gran potencial de su uso parece estar en el perfeccionamiento de las capacidades de evaluación de los profesores.*

**Palabras-clave:** Evaluación de escuelas – Evaluación basada en el aula – Evaluaciones internacionales de desempeño de estudiantes – Evaluaciones nacionales – exámenes públicos.



**Referências Bibliográficas**

- BLACK, P., WILIAM, D. Assessment and classroom learning. *Assessment in Education*, n.5, p.7-74, 1998.
- CHINAPAH, V. *Handbook on monitoring learning achievement: towards capacity building*. Paris: UNESCO, 1997.
- CROOKS, T.J. The impact of classroom evaluation practice on students. *Review of Educational Research*, n.58, p.438-81, 1988.
- EISEMON, T.O. Examination policies to strengthen primary schooling in african countries. *International Journal of Educational Development*, n.10, p.69-82, 1990.
- FREDERIKSEN, N. The real test bias: influence of testing on teaching and learning. *American Psychologist*, n.39, p.193-202, 1984.
- GOLDSTEIN, H. International comparisons of student achievement. In: LITTLE, A., WOLF, A. (Ed.). *Assessment in transition. learning, monitoring and selection in international perspective*. Oxford: Pergamon, 1996.
- GOLDSTEIN, H., THOMAS, S. Using examination results as indicators of school and college performance. *Journal of the Royal Statistical Society*, n.159, p.149-63, pt.1, 1996.
- GREANEY, V., KELLAGHAN, T. *Monitoring the learning outcomes of education systems*. Washington, DC: World Bank, 1996a.
- GREANEY, V., KELLAGHAN, T. The integrity of public examination in developing countries. In: GOLDSTEIN, H., LEWIS, T. (Ed.) *Assessment: problems, developments and statistical issues*. New York: Wiley, 1996b.
- GREAT BRITAIN. Department of education and Welsh Office. *Task group on assessment and testing: a report*. London: Her Majesty'S Stationery Office, 1988.
- GUSKEY, T.R., KIFER, E.W. Ranking school districts on the basis of state wide results: Is it meaningful or misleading? *Educational Measurement: Issues and Practice*, v.9, n.1, p.11-16, 1990.
- HEYNEMAN, S.P., RANSOM, A.W. Using examinations and testing to improve educational quality. In: ECKSTEIN, M.A., NOAH, H.J. (Ed.) *Examinations: comparative and International studies*. Oxford: Pergamon, 1992.
- HUSÉN, T., POSTLETHWAITE, T.N. A brief history of the International. Association for the Evaluation of Educational Achievement (IEA). *Assessment in Education*, n.3, p.129-41, 1996.

KELLAGHAN, T. IEA studies and educational policy. *Assessment in Education*, n.3, p.143-60, 1996.

KELLAGHAN, T. Seguimiento de los resultados educativos nacionales. In: ALVAREZ H.B., RUIZ-CASARES, M. (Ed.). *Evaluación y reforma educativa: opciones de política*. Washington DC: U.S. Agency for International Development, 1997.

KELLAGHAN, T. Outcome evaluation. In: STUFFLEBEAN, D.L., MADAUS, G.F., KELLAGHAN, T. (Ed.). *Evaluation models: viewpoints on educational and human service evaluation*. Boston: Kluwer Academic Publishers, 2000.

KELLAGHAN, T., GREANEY, V. *Using examinations to improve education: a study in fourteen African countries*. Washington DC: World Bank, 1992.

KELLAGHAN, T., GREANEY, V. The globalisation of assessment in the 20th century. *Assessment in Education*, n.8, p.87-102, 2001.

KELLAGHAN, T., GRISAY, A. International comparisons of student achievement: problems and prospects. In: *Measuring what students learn: mesurer les résultats scolaires*. Paris: Organization for Economic Cooperation and Development, 1995.

KELLAGHAN, T., MADAUS, G.F., AIRASIAN, P.W. *The effects of standardized testing*. Boston: Kluwer-Nijhoff, 1982.

KELLAGHAN, T., MADAUS, G.F., RACZEK, A. *The use of external examinations to improve student motivation*. Washington DC: American Educational Research Association, 1996.

LINN, R.L. Assessments and accountability. *Educational Researcher*, v.29, n.2, p.4-16, 2000.

MADAUS, G.F., KELLAGHAN, T. Curriculum evaluation and assessment. In: JACKSON, P. (Ed.) *Handbook of research on curriculum*. New York: Macmillan, 1992.

MADAUS, G.F., KELLAGHAN, T. (in press). Teachers' sources and uses of assessment information in TIMSS. In: BEATON, A.E., ROBITAILLE, D.F. (Ed.) *TIMSS secondary analyses*.

OLIVARES, J. Inclusive national testing: Chile's 'quality of education assessment system'. In: LITTLE, A., WOLF, A. (Ed.) *Assessment in transition: learning, monitoring and selection in international perspective*. Oxford: Pergamon, 1996.

PUNJAB COMMISSION FOR EVALUATION OF EXAMINATION SYSTEM AND ERADICATION OF MALPRACTICES. *Report: report submitted to Mr. Ghulam Hauder Wyne, Chief Minister of the Punjab*. Lahore: Author, 1992.

ROBITAILLE, D.F., BEATON, A.E., PLONP, T. (Ed.). *The impact of TIMSS on the teaching and learning of mathematics and science*. Vancouver: Pacific Educational Press, 2000.

ROJAS, C.C., ESQUIVEL, J.M. *Los sistemas de medición del logro académico and Latino América*. Washington DC: World Bank, 1998.

ROSS, K.N. (Ed.) From educational research to educational policy: an example from Zimbabwe. *International Journal of Educational Research*, v.23, n.4, 1995. (whole issue).

ROSS, K.N., MAEHLCK, L.(Ed.) *Planning de quality of education: the collection and use of data four informed decision-making*. Oxford: Pergamon; Paris: UNESCO, 1990.

SCHOOL CURRICULUM AND ASSESSMENT AUTHORITY. *Value-added performance indicators of schools*. London: Author, 1994.

SOMERSET, A. Examination and educational quality. In: LITTLE, A., WOLF, A. (Ed.). *Assessment in transition: learning, monitoring and selection in international perspective*. Oxford: Pergamon, 1996.

TELL, C. Whose curriculum? A conversation with Nicholas Tate. *Educational Leadership*, v.56, n.2, p.64-9, 1998.

TORRANCE, H. The role of assessment in educational reform. In: TORRANCE, H. (Ed.) *Evaluating authentic assessment: problems and possibilities in new approaches to assessment*. Buckingham: Open University Press, 1995.

WEBSTER, W.J., MENDRO, R.L., ALMAGUER, T.D. Effectiveness indices: a 'value-added' approach to measuring school effect. *Studies in Educational Evaluation*, n.20, p.113-45, 1994.

WORLD DECLARATION ON EDUCATION FOR ALL. Adopted by the World Conference on Education for All. *Meeting Basic learning Needs*. New York: UNICEF House, 1990.