

Usos dos resultados das avaliações de sistemas educacionais: iniciativas em curso em alguns países da América

Adriana Bauer

Resumo

Este artigo é fruto de pesquisa que buscou caracterizar sistemas de avaliação em larga escala desenvolvidos em países da América. Tendo como referência informações obtidas de diversas fontes, analisa tendências de uso dos resultados das avaliações de sistemas educacionais, a partir de seus objetivos declarados. Busca retomar os objetivos das primeiras experiências no campo de avaliação educacional, desde o século 19, e levantar os objetivos anunciados em sistemas atuais, consolidados a partir dos anos de 1990. Como conclusão, destaca a propensão da utilização das avaliações no âmbito de gerenciamento dos sistemas educacionais e prestação de contas. Finalmente, chama a atenção do leitor para algumas críticas que vêm sendo feitas a esses sistemas e seus objetivos.

Palavras-chave: avaliação de sistemas; resultados; gestão educacional.

Abstract

Using the results of educational system assessments: experiences carried on some American countries

This article is based on a research that sought to characterize large scale educational assessments developed in American countries. It gathers referential information from various sources and analyses trends on the usage of the results of system assessments, from their stated objectives. It seeks to recover the objectives of the first experiments in the field of educational assessment since the 19th century and to examine the goals established for the current system assessments, since the 1990s. In order to conclude, it highlights a tendency towards the use of these assessments for the management and accountability of educational systems. Finally, it calls the reader's attention to some critics that have been made to these assessments and their goals.

Keywords: system assessment; results; educational management.

As origens da moderna avaliação de programas e sistemas educacionais¹

Ainda que a avaliação constitua um campo de estudos e pesquisas que se desenvolveu notadamente a partir do século 20, alguns estudiosos são unânimes em afirmar que avaliar, em seu sentido mais amplo, é uma prática tão antiga que remonta aos primórdios da história do homem:

A avaliação, em um sentido amplo, é uma atividade que faz parte da vida humana e está presente no cotidiano dos indivíduos. (Dias Sobrinho, 2002, p. 17).

Ousaríamos dizer que a avaliação surgiu com o próprio homem, se entendermos por avaliação a visão apresentada por Stake – o homem observa; o homem julga, isto é, avalia. (Vianna, 2000, p. 22).

Avaliação não é um conceito novo. Com relação ao aspecto de “examinar e julgar, determinar o valor”, a prática da avaliação sem dúvida precedeu muito sua definição, e suas raízes remontam aos primórdios da história humana. O homem de Neanderthal praticou-a ao determinar que tipos de madeira se prestavam à confecção das melhores lanças, assim como os patriarcas persas ao selecionar os pretendentes mais adequados para suas filhas ou os pequenos proprietários rurais da Inglaterra, que abandonaram seus arcos curtos (bestas) e adotaram os arcos longos do País de Gales [...] (Worthen, Sanders, Fitzpatrick, 2004, p. 36).

¹ Esta sessão tem como referência os estudos sobre história da avaliação de programas educacionais desenvolvidos, principalmente, nos Estados Unidos da América, país em que a discussão sobre avaliação de programas e seus pressupostos teórico-metodológicos tem sido ampla, acompanhada por produção bibliográfica igualmente vultosa. Contudo, apesar da projeção que o campo da avaliação alcançou nos estados americanos, importantes autores, como Madaus, Scriven e Stufflebeam (1983), atestam que uma completa história da avaliação de programas, em nível mundial, ainda está para ser realizada.

Nos sentidos apresentados acima, a avaliação pode ser entendida enquanto prática informal (sem procedimentos sistemáticos ou coleta formal de dados/evidências, normalmente baseada em percepções) e social, usada não só para julgamento, mas também para escolha e seleção. Permeando as escolhas e as decisões humanas, essas avaliações informais assumiam (e ainda assumem) os mais diversos significados e utilidades. Apesar de essa avaliação informal estar presente no cotidiano, ao retomar alguns marcos históricos da moderna avaliação de programas, este estudo irá resvalar na avaliação *formal*, definida como “uso sistemático de informações e critérios acurados para atribuir valores e justificar juízos de valor” (Worthen, Sanders, Fitzpatrick, 2004, p. 60-61).

A ideia de avaliar sistemas e programas educacionais, tão alardeada no Brasil após os anos 1990, remonta, pelo menos, ao início do século 19, desenvolvendo-se a partir de pesquisas empíricas de cunho quantitativo.

Segundo Worthen, Sanders e Fitzpatrick (2004), as primeiras iniciativas de avaliação formal, mais especificamente no que se refere à avaliação de programas e sistemas educacionais, estão relacionadas à pesquisa social, realizada por meio de levantamentos quantitativos já no final do século 17. Tais pesquisas, que viriam a inspirar a avaliação de programas educacionais, teriam sofrido uma forte influência das Ciências Naturais, em franco desenvolvimento àquela época.

Já Madaus, Scriven e Stufflebeam (1983) localizam na Revolução Industrial e nas mudanças sociais que ela impulsionou as origens da moderna avaliação de programas, ressaltando, porém, que, à época de seus escritos, uma completa história da avaliação, retornando ainda mais às suas origens, ainda estava por ser realizada.

De qualquer forma, exemplos de primeiras iniciativas de avaliação são encontrados na Inglaterra, na Irlanda e nos Estados Unidos da América.

Na Inglaterra e na Irlanda, a pressão por reformas dos programas educacionais fez com que ocorressem as primeiras avaliações de instituições escolares, executadas por representantes nomeados pelo rei. As tentativas de reforma educacional e de desenvolvimento de hospitais, orfanatos e saúde pública teriam impulsionado avaliações de natureza informal e impressionista, normalmente feitas por comissões de governo, com o objetivo de “investigar aspectos das áreas sob consideração” (Madaus, Scriven, Stufflebeam, 2003, p. 4). Apesar de também citarem exemplos de comissões relacionadas à área da saúde, os autores apontam já uma preocupação com a avaliação do rendimento dos alunos, que nascia sob a ideia de avaliação de programas educacionais:

Por exemplo, a Comissão Real de Investigação dentro da educação primária na Irlanda, subordinada ao Conde de Powis, depois de receber testemunho e examinar evidências, concluiu que “o progresso das crianças nas escolas nacionais da Irlanda é muito inferior do que deveria ser”. Como solução, a Comissão Powis recomendou, então, a adoção do esquema de pagamento conhecido como “pagamento por resultados” já sendo usado na Inglaterra, por meio do qual os salários dos professores seriam dependentes em parte dos resultados dos exames nacionais em leitura, ortografia, escrita, e aritmética. (Madaus, Scriven, Stufflebeam, 2003, p. 4).

Outra forma de avaliação apontada pelos autores, nos primórdios da história do campo, seria a realizada pelos inspetores escolares, que geravam relatórios anuais sobre as condições das escolas e as realizações dos alunos. Também citam a criação de associações destinadas à investigação social, ainda na Grã-Bretanha, e a organização de investigações de programas sociais patrocinadas pelo governo como importantes influências às avaliações empíricas que se seguiriam com a formalização do campo (Madaus, Scriven, Stufflebeam, 1983, p. 3-22).

O processo de desenvolvimento das avaliações de programas foi influenciado, ainda, pelas transformações ocorridas, a partir do século 19, não apenas na forma de avaliar os alunos no interior das escolas (que anteriormente era feita oralmente, muitas vezes em sessões públicas), mas também mediante uma intensificação na função de organização social da avaliação, que passou a ser vista como instrumento de legitimação das práticas escolares e seleção de alunos e profissionais para cargos públicos, baseada em uma perspectiva de mérito (Dias Sobrinho, 2002).

No que se refere à forma de avaliar os alunos por meio de exames escritos estandardizados, a primeira experiência de que se tem notícia originou-se nos Estados Unidos (Boston, Massachusetts), em 1845. Madaus, Scriven e Stufflebeam (1983, p. 6) apontam, inclusive, que por trás da iniciativa de instituir exames escritos, mais do que a ideia de justiça nos julgamentos, estava presente a ideia de ter dados (as provas), para a comparação das diferenças entre as escolas, que seriam utilizados para pressionar os mestres-escolas favoráveis ao término das punições corporais a seguirem as orientações do Conselho de Educação. De qualquer forma, a iniciativa de Boston é considerada pelos autores como um marco na história da avaliação, pois teria gerado uma tradição, que persiste até hoje, de usar os resultados dos alunos em testes “como a principal fonte de dados para avaliar a efetividade da escola ou do programa instrucional” (p. 5).

Outro exemplo de sistema de avaliação que surgiu influenciado pela concepção de julgar a *performance* seria o *baccalaureát* francês, sistema de certificação que selecionava os alunos que poderiam ingressar no sistema superior e, ao mesmo tempo, legitimava o trabalho realizado no nível anterior de escolaridade, equivalente ao ensino médio no Brasil. Concebido no início do século 19, em 1808, Dias Sobrinho (2002) localiza no *baccalaureát* uma primeira experiência, ocorrida na Europa, de avaliação de sistemas educacionais, que ele denomina de “protótipo dos exames nacionais”.

Madaus, Scriven e Stufflebeam (1983, p. 6), bem como Worthen, Sanders e Fitzpatrick (2004, p. 62), localizam a primeira prática de avaliação de programas educacionais na experiência americana de Joseph Rice (1887-1898), na qual se avaliava o desempenho dos estudantes e se fazia comparações entre as escolas da região.

De qualquer forma, o trabalho de Rice parece ter sido o primeiro estudo longitudinal conhecido em que se procurava avaliar o valor do exercício/repetição no aprendizado da ortografia. A partir dos resultados

que obteve, Rice passou a argumentar que os professores deveriam ser “experimentalistas e pensadores quantitativos”, e seu uso de design de pesquisa para o estudo da aprendizagem dos alunos influenciou a abordagem de Lidquist (1953), Campbell (1969) e outros.

Quase concomitantemente, em 1876, o Canadá também iniciou uma primeira experiência diferenciada de avaliação, no que era conhecido como Upper Canadá (hoje, Ontário), influenciada pela necessidade de expansão da educação em nível secundário (Klinger, De Luca, Miller, 2008). Sob a administração de Egerton Ryerson, Superintendente-Chefe da Educação entre 1846 e 1876, instauraram-se os primeiros exames intermediários (*intermediate examinations*), com o objetivo de selecionar os alunos para o secundário.

Cabe abrir parênteses para considerar que as primeiras avaliações de programas e sistemas educacionais diferem em objetivos e função, se comparadas àquelas que são realizadas hoje em dia. É interessante ressaltar que, apesar de a avaliação dos alunos no interior da escola ser tratada, atualmente, pela avaliação da aprendizagem, subárea também bastante explorada em educação, os exemplos fornecidos por autores de referência permitem supor que as primeiras avaliações de alunos compunham a avaliação dos programas escolares. Posteriormente, com o desenvolvimento de testes e a preocupação com a avaliação do conhecimento, a avaliação de programas e a de aprendizagem foram tomando caminhos próprios e tendem a ser vistas de modo distinto, sendo que as avaliações em larga-escala, não raro, são consideradas incompatíveis com a rotina escolar. Tal fato é aqui mencionado para indicar a necessidade de retomada das origens do campo, que poderiam influenciar, talvez, melhores usos das avaliações de programas e sistemas educacionais.

De qualquer forma, após as primeiras iniciativas de avaliação de programas, a avaliação passou a estar muito relacionada à mensuração de rendimento individual dos alunos, por meio de testes escritos, com forte preocupação em avaliar suas capacidades psíquicas. Essa tendência, que Madaus, Scriven e Stufflebeam (1983) chamam “A idade da eficiência e testagem”, continuou no início do século 20 com o desenvolvimento de metodologias mais sofisticadas de coleta e análise de dados, como ilustram Worthen, Sanders e Fitzpatrick (2004, p. 63):

A década de 1920 viu o rápido surgimento de testes baseados nas normas dos programas, criados para avaliar os níveis individuais de desempenhos. Em meados da década de 1930, mais da metade dos Estados Unidos tinha algumas formas de testes estaduais, e os testes padronizados, baseados nas normas de referência dos programas, inclusive testes para avaliar o aprendizado, assim como testes de personalidade e perfis de interesses, transformaram-se em um grande negócio. Durante esse período, os educadores consideravam as mensurações e a avaliação quase sinônimos, sendo esta última, em geral, vista como um teste final para avaliar o desempenho de um aluno e atribuir-lhe notas.

Nesse sentido pode-se dizer que a avaliação do início do século 20 foi dominada por uma perspectiva psicológica que buscava mensurar

inteligência e conhecimento. Contudo, a avaliação nesse período não se reduz ao estudo, à produção e à aplicação de testes de inteligência. Devido à tendência de gerenciamento científico que passou a dominar as teorias gerais de administração, influenciando também as de administração escolar, passou-se a enfatizar a sistematização, a standardização e a eficiência (Madaus, Scriven, Stufflebeam, 1983, p. 6-7), trinômio recorrente em estudos avaliativos da época.

Ainda segundo os autores supracitados, à tendência de mensuração sistemática e standardizada agregou-se, a partir da década de 1910, uma preocupação com a avaliação das aquisições propiciadas pelos programas educacionais: procurava-se medir e aferir a eficiência das escolas e dos professores. Objetivos educacionais eram traçados e exames eram elaborados para medi-los, havendo um crescente emprego dos testes padronizados para fazer inferências sobre a eficácia dos programas, considerando os resultados obtidos. Datam também, dessa época, os primeiros usos equivocados dos resultados das avaliações:

Eventualmente testes como aqueles em Boston [que] assumiram um caráter de referenciado à norma quando a percentagem de alunos aprovados tornou-se um padrão pelo qual os professores poderiam julgar se suas classes estavam acima ou abaixo do parâmetro geral para a cidade (Ballou, 1916). [...] [Além disso] testes elaborados por pesquisadores como Courtis, Ayers, Thorndike, e outros, que foram orientados para medir um conjunto muito preciso de objetivos instrucionais, [...] geraram dados normativos que possibilitaram um sistema comparar a si mesmo com outro (Tyack, Hansot, 1982). [...] Outro problema associado com estas avaliações iniciais [...] era que os resultados "objetivos" obtidos eram frequentemente usados como propaganda "para construir diques de dados contra marés crescentes de criticismo público". (Madaus, Scriven, Stufflebeam, 1983, p. 7).

O período de reconstrução social e econômica após a 1ª Guerra Mundial e a conjuntura econômica de alguns países, principalmente os Estados Unidos, após a recessão da década de 1930, influenciaram a ampliação do campo da avaliação para além da medida das capacidades individuais:

A avaliação deveria se dedicar ao êxito na escolarização. Comprometia-se, então, com a ideologia da eficiência social. Os objetivos educacionais eram estabelecidos em função do desenvolvimento industrial e com o esforço de recuperação da economia norte-americana. Observa-se, então, uma certa ampliação do campo da avaliação. Continua sendo fundamentalmente um processo de medida, porém, agora, mais preocupada com a adequação dos currículos às exigências sociais e econômicas, com a formulação de objetivos e a medida de seu cumprimento.

A avaliação, nesta fase, ganha um sentido mais operativo: ela passa a averiguar até que ponto os currículos e as práticas pedagógicas estão atingindo os objetivos de levar a escola a ser eficaz e alcançar as metas de eficiência que a economia exigia. Portanto, torna-se mais evidente a sua racionalidade instrumental. Ainda que permanecesse muito centrado nas escolas e, sobretudo, no eixo ensino-aprendizagem, seu campo se ampliou, na medida em que passou a articular os rendimentos

dos alunos às questões de currículos e programas. Sua importância também cresceu como mecanismo de controle e seleção, segundo a lógica utilitarista que então predominava, bem como a aceitação de sua fidelidade e cientificidade, graças ao aperfeiçoamento dos instrumentos de testes e medidas. (Dias Sobrinho, 2002, p. 21-22).

Contudo, observa-se que, diferentemente do que ocorre atualmente, as primeiras experiências em avaliação e testagem e seus usos para controle, seleção e gerenciamento ocorriam de forma localizada e isolada.

Digna de nota é a influência exercida por Ralph Tyler no desenvolvimento do campo da avaliação. Reconhecido como “pai da avaliação educacional” (foi o primeiro a cunhar o termo), Tyler não só trouxe a ideia de avaliação relacionada aos objetivos educacionais (Vianna, 2000, p. 47-66), como também concebeu que a avaliação seria realizada pela comparação dos resultados obtidos – *actual outcomes* – com aqueles intencionados pela intervenção – *intended outcomes* (Madaus, Scriven, Stufflebeam, 1983, p. 8-9), realizando o primeiro estudo longitudinal de largo alcance, o *Eight-Year Study*, ocorrido entre 1932 e 1940.

Os vinte anos que seguem, ou seja, a década de 50 e quase toda a de 60, são dedicados à expansão e ao aperfeiçoamento técnico relacionados às práticas de testagem e ao desenvolvimento de teorias que ajudassem a explicitar os objetivos referentes aos conteúdos do que seria avaliado. Nesse período, por exemplo, são produzidas as taxonomias de Bloom,² que ficaram conhecidas universalmente. Vale ressaltar que, ainda nessas décadas, as iniciativas avaliativas e o desenvolvimento técnico da área ocorriam em nível local, ou seja, as avaliações tinham um alcance limitado, sem projeção estadual ou nacional e praticamente sem envolvimento do governo federal ou de suas agências.

A partir da metade da década de 60, a área sofreu novo grande impulso. Com a promulgação da Lei Educacional de Defesa Nacional de 1958, nos Estados Unidos, em resposta à necessidade de um ensino mais eficiente nas escolas, principalmente na Matemática e nas Ciências, gerou-se uma demanda por avaliações de programas elaborados para a melhoria dos currículos e dos projetos daquelas disciplinas. Como explicam Worthen, Sanders e Fitzpatrick, o lançamento do *Sputnik I*, foguete soviético, em 1957, abalou a crença dos Estados Unidos em sua potencialidade técnica e científica, produzindo uma demanda por avaliações dos currículos e das práticas de Matemática e Ciências, já que o ensino dessas disciplinas deveria ser mais eficaz. Dessa forma, o governo americano passou a investir milhares de dólares para o aprimoramento de programas e o desenvolvimento dos currículos dessas áreas (Worthen, Sanders, Fitzpatrick, 2004, p. 65).

Além disso, o desenvolvimento de projetos nas áreas sociais e da saúde, com o objetivo de equalizar as oportunidades para todos os cidadãos americanos, bandeira dos governos de John Kennedy e Lyndon Johnson, levou a um investimento maciço em programas sociais. Segundo José Dias Sobrinho (2002, p. 26), tais investimentos calcavam-se na crença liberal de que, ao investir em programas que combatessem os diversos

² Benjamim Bloom, durante a década de 1950, liderou uma comissão de pesquisa, com participação de pesquisadores de diversas universidades americanas, que produziu uma estrutura hierárquica dos objetivos educacionais em três distintos níveis – cognitivo, afetivo e psicomotor –, mais conhecida como a Taxonomia de Bloom. Publicada no Brasil, a obra resultante desse trabalho influenciou as discussões sobre planejamento e objetivos do ensino (Bloom et al., 1972).

problemas sociais existentes, seria possível superá-los e propiciar, assim, um melhor desenvolvimento da sociedade.

Nesse sentido, a avaliação da eficácia dessas políticas e programas, por meio da quantificação dos resultados obtidos, passou a ser foco de atenção dos governos. Era necessário avaliar os programas para conferir maior racionalidade aos investimentos, assegurando credibilidade aos projetos e transparência na gestão dos recursos, já que muitos bilhões de dólares estavam sendo destinados ao financiamento deles. Teorias seminais para o campo de avaliação, produzidas por autores como Donald Campbell e Michel Scriven, são divulgadas à época, e é possível observar grande preocupação com a busca pela “verdade” sobre os resultados dos programas, baseada no interesse público (Shadish, Cook, Leviton, 1991).

Um terceiro fator que contribuiu para o desenvolvimento da avaliação de programas na década de 60 foi a aprovação da Lei do Ensino Fundamental e Médio (LEFM), em 1965, que “propunha um aumento colossal do financiamento federal da educação, com dezenas de milhares de bolsas federais para escolas municipais e estaduais e para universidades e órgãos regionais” (Worthen, Sanders, Fitzpatrick, 2004, p. 69).

A falta de evidências comprobatórias de que financiamentos federais tivessem impactos positivos ou melhorias genuínas no setor da educação fez com que fosse exigida, pelo congresso americano e com o apoio do presidente John Kennedy, a avaliação das atividades realizadas pelos beneficiários dos fundos federais. Os educadores passaram a ter que avaliar suas próprias atividades, sendo a LEFM considerada um marco na história da avaliação de programas e políticas educacionais.

Entretanto, no início desse processo, a falta de pessoal especializado para suprir a demanda por avaliações fez com que muitas delas fossem realizadas a partir de concepções e metodologias inadequadas que, geralmente, se apoiavam nos conhecimentos produzidos em teorias de áreas afins, como a Psicometria e as Ciências Sociais, gerando, entre outros problemas, relatórios de avaliação irrelevantes. Segundo Worthen, Sanders e Fitzpatrick (2004), foi o diagnóstico desse “empobrecimento” teórico e metodológico da área que propiciou a abertura do debate sobre a avaliação de programas, resultando na ampliação da discussão sobre questões conceituais e metodológicas que balizavam a prática dos avaliadores.

Esse movimento deu início ao desenvolvimento de diferentes abordagens filosóficas e metodológicas na área, com efetivos esforços para a construção de conhecimentos empíricos e conceituais que a subsidiassem. Contudo, apesar de começar a existir uma preocupação de técnicos e estudiosos com aspectos que ultrapassavam a mensuração de resultados obtidos pelos programas, observa-se que o foco, neste momento, mantinha-se sobre os dados mais quantitativos da avaliação.

A lógica de racionalidade técnica e econômica que embasou a avaliação de programas e sistemas nas décadas de 60 e 70 do século 20, apesar de persistir na década posterior, foi influenciada pela reconfiguração da sociedade e da economia observada a partir das mudanças tecnológicas,

principalmente nas áreas da informática e das telecomunicações, que influem sobre o sistema capitalista.³ Tais mudanças tiveram implicações tanto na organização quanto no papel dos Estados-Nação, que viram a necessidade de alterar algumas de suas funções, a fim de viabilizar o crescimento econômico interno e superar a crise de Estado que levava ao aumento da desigualdade e à exclusão social.⁴

Assim, a avaliação de programas e políticas educacionais passou a ser ressignificada e a adquirir papel central no âmbito das políticas públicas. Nessa lógica, diversos governos, no Brasil e no mundo, passam a demandar avaliações dos sistemas educacionais, como será visto posteriormente.

Os anos 90 e a avaliação de programas e sistemas educacionais em países americanos: para que servem os resultados da avaliação?

Alicia Bonamino (2002, p. 15-16) explica o contexto de desenvolvimento da avaliação de sistemas no Brasil nos anos 1990:

No final dos anos 80, o reconhecimento da inexistência de estudos que mostrassem mais claramente o atendimento educacional oferecido à população e seu peso sobre o desempenho dos alunos dentro do sistema escolar conduziu às primeiras experiências de avaliação do ensino de primeiro grau.

Já nos anos 90, o sistema de avaliação da educação básica passa a inserir-se em um conjunto mais complexo de inter-relações, em cujo interior operam o aprofundamento das políticas de descentralização administrativa, financeira e pedagógica da educação, um novo aparato legal e uma série de reformas curriculares. Essas inter-relações estão demarcadas pelo encerramento do ciclo de recuperação da democracia política e pela aceitação das novas regras internacionais, derivadas da globalização e da competitividade econômica.

Apesar de se referirem ao caso brasileiro, as considerações de Bonamino espelham a lógica de desenvolvimento de alguns dos sistemas de avaliação dos países americanos que, preocupados com os índices e a qualidade dos serviços educacionais e impelidos pela necessidade de um gerenciamento eficaz dos recursos disponíveis para a educação, valorizam as informações obtidas por meio da aplicação de testes aos alunos como suporte à tomada de decisões na área educacional.

Nessa conjuntura, a avaliação de sistemas passa a ser um ponto de destaque nas propostas de políticas públicas em vários países do globo. Ao explicar a centralidade da avaliação, Patricia Broadfoot (1996, p. 7), ao tratar da função social da avaliação em seu livro *Education, assessment and society*, afirma que a “avaliação educacional em várias formas tem sido de importância central na criação de sistemas educacionais *per se* através da racionalização da provisão educacional e do controle da prática educativa”.

Mediante essa racionalização e controle, o gerenciamento deixa de ser realizado no âmbito de estruturas burocráticas para ser orientado

³ Essas redefinições tiveram lugar, principalmente, nos últimos 25 anos do século 20, acirrando-se com o fim da Guerra Fria e a abertura dos estados soviéticos à economia capitalista.

⁴ Muitos são os autores que analisam as relações entre neoliberalismo, papel do Estado e avaliação de forma bastante detalhada, constituindo-se referência para os que buscam se aprofundar nessa questão. Tendo em vista a existência de trabalhos de excelente qualidade sobre essa temática, optamos por não aprofundá-la nesta reflexão, já que o objetivo é ressaltar as diversas perspectivas que a avaliação de sistemas e programas tem assumido durante seu desenvolvimento. Para um maior aprofundamento nas questões concernentes ao neoliberalismo, Estado e avaliação, conferir as obras de Afonso (1998, 2000), Sousa (1997) e Hidalgo e Silva (2001), que se ocupam especificamente da avaliação da educação básica no Brasil.

pelos resultados obtidos por meio de indicadores de *performance* e do cumprimento de objetivos previamente determinados pelos centros decisores do governo. A avaliação de programas e sistemas educacionais começa a exercer a função de controle sobre os serviços educacionais e prestação de contas tanto ao Estado quanto à sociedade. Como explica Dias Sobrinho (2002, p. 29),

a avaliação assumiu basicamente as características de *accountability*: uma forma tecnocrática de valorar e um procedimento burocrático de exigir o cumprimento de obrigações. É inevitável a conexão entre a *accountability* e a ideologia da eficiência. A responsabilidade, antes entendida nos âmbitos universitários como pertinência e equidade, ou em outras palavras, a prestação de contas à sociedade como um todo, referida como *accountability*, se transforma na exigência de demonstração na obtenção de determinados resultados através do emprego dos meios mais eficientes. É, portanto, a capacidade de prestar contas não à sociedade, mas aos governos e aos clientes.

Segundo Broadfoot (1996), essa *accountability*, compreendida enquanto prestação de contas, pode assumir diversas formas, a depender do país. Ela desenvolve sua argumentação explicando que em alguns países a ênfase pode recair sobre procedimentos de *garantia* de qualidade, como a autoavaliação institucional (por meio da qual a escola se responsabilizará por sanar suas dificuldades particulares) ou a avaliação docente, enquanto que em outros pode incidir sobre os procedimentos de *controle* de qualidade. Independentemente da ênfase dada, o uso de diferentes métodos de avaliação formal para exercer controle, mediante o que a autora chama de “linguagem da *accountability*”, tem se tornado uma característica típica e proeminente das políticas educacionais, nas quais a avaliação assume papel central.

Klinger, DeLuca e Miller (2008) retomam alguns autores para discutir os objetivos da avaliação. A partir de seu estudo, ressaltam que “noções de *accountability* educacional mudam em resposta a movimentos econômicos e políticos” e, conseqüentemente, “os propósitos dos programas de avaliação frequentemente mudam ou se expandem, embora sem mudanças substantivas no programa de avaliação” que está em execução. De qualquer forma, segundo esses autores, haveria três diferentes propósitos para as avaliações em larga-escala: *accountability*, classificação e seleção *gatekeeping* e diagnóstico instrucional:

Accountability relates to public concerns regarding the educational system and student competencies [...]. These assessments are often associated with program and system examinations intent on providing information about specific schools, school-based initiatives, programs, and teachers. Large-scale assessments impact student –, program –, and system-level decisions under the auspices of accountability to stakeholders. They typically provide data supporting program modifications or reflecting the health of the provincial/territorial educational system. The belief is that such assessment programs effect change in policy, curriculum, and practice, especially

in a climate of data driven decision making (Madaus, Kellaghan, 1992). [...] Gatekeeping is used to grant students privileges such as graduation, admission, or grade promotion. At the extreme, results from the large-scale assessments become the sole determiner of a student's success. [...] Large-scale assessments for instructional diagnosis seek to determine what students actually know with respect to a set of criteria, subject expectations, or learning outcomes (Nagy, 2000). Assessment results are provided to teachers and students in a timely manner in order to support and guide instruction and learning.

Contudo, parece que as lógicas dominantes são as de *accountability* e *gatekeeping*, principalmente a de *accountability*, que é assumida e enfatizada, inclusive, por agências que têm grande influência sobre os sistemas educacionais, como é o caso da United Nations Educational, Scientific and Cultural Organization (Unesco). No documento *O financiamento e a gestão da educação na América Latina e no Caribe*, produzido em parceria com a Comissão Econômica para a América Latina e o Caribe (Cepal), o conceito de *accountability* é central, sua importância para a educação é defendida e são realizadas indicações sobre como proceder para uma efetiva *prestação de contas*.

Uma maior autonomia da escola reconhece a importância da prestação de contas (*accountability*), cujo objeto é entregar informação à comunidade, às autoridades e à sociedade no geral, saber se o estabelecimento está cumprindo com as metas e funções acordadas. O sistema de prestação de contas se associa à melhoria do desempenho dos alunos não somente porque faculta a atores extraescola – desde pais até instâncias centrais de planejamento e avaliação – para exigir uma oferta pertinente e de qualidade, mas também porque introduz na escola rotinas de compromisso e monitoramento, que impõem maior sentido de responsabilidade pelo funcionamento e pelos resultados. (Cepal, Unesco, 2004, p. 71).

Percebe-se um forte caráter ideológico no debate do *accountability*, em que discursos aparentemente democráticos e legítimos (como a transparência de informações) são utilizados para respaldar ações centralizadoras e antidemocráticas, cuja principal característica é a manutenção do controle sobre o sistema de ensino pelo governo, sem que este assuma, diretamente, a responsabilidade integral pelos serviços prestados. Nesse sentido, observa-se que avaliar o sistema educacional é uma ação essencial a essa lógica de prestação de contas e, por isso, tem se consolidado em vários países do mundo.

No Brasil e em outros países do continente americano, essa centralidade do papel da avaliação passa a ocorrer no final dos anos 1980 e na década de 1990,⁵ quando o panorama internacional atribuiu maior importância à avaliação educacional em larga escala, especialmente com relação aos sistemas responsáveis pela educação básica.

Observa-se, também, que a internacionalização da ideia de avaliar sistemas e programas educacionais deve-se à conjuntura social e

⁵ Refazer o histórico da avaliação educacional no Brasil foge aos objetivos deste trabalho. Contudo, alguns estudiosos já se dedicaram ao assunto e o leitor poderá encontrar boas referências em Vianna (2000), Bonamino (2002) e Sousa e Freitas (2004). Deve ficar claro, porém, que, apesar de o desenvolvimento dos sistemas de avaliação no Brasil ser recente, avaliar sistemas educacionais não é uma novidade, sendo que diversas iniciativas já eram desenvolvidas no Brasil e no exterior. O que mudou, em perspectiva mundial, foram as funções da avaliação e o uso objetivo de seus resultados.

econômica globalizada e à influência de organismos multilaterais⁶ na disseminação da necessidade de avaliar os sistemas de ensino. Essa última questão é ressaltada por Costa (1998, p. 40-41), ao comentar a avaliação educacional no sistema argentino:

A avaliação educacional, que tem décadas de tradição nos países centrais, desde os anos 80 aumentou muito em nível dos sistemas educativos na América Latina, por impulso tanto dos organismos que operam na região – os de educação, ciência e cultura –, como dos organismos que propõem as políticas econômicas.

Essa tendência acentuou-se nos anos 90, chegando a ser incorporada como uma instância institucionalizada nos processos de reformas ou de transformações educacionais impulsionadas pelo Estado, inclusive nas leis que regulam a educação dos países. Citamos como exemplos os casos da Argentina (Lei Federal da Educação – 1993), Bolívia (Lei da Reforma Educacional – 1994) e Colômbia (Lei Geral de Educação – 1994).

Barrera (2000) retrata a disseminação da avaliação de sistemas educacionais no final do século 20, divulgando um levantamento dos sistemas que já estavam consolidados até o início do segundo milênio.

Conforme os dados registrados no Quadro 1, nota-se que, durante a década de 80, apenas cinco países haviam iniciado alguma experiência nacional em avaliação dos sistemas de educação: Panamá, Costa Rica, Chile, Colômbia e Estados Unidos. No período entre 1990 e 1995, esse número eleva-se para 14, com as iniciativas de Brasil e Honduras (1990), República Dominicana (1991), Guatemala (1992), Argentina, Canadá e El Salvador (1993), México (1994) e Paraguai (1995). Finalmente, consolidam-se os sistemas de Bolívia, Cuba, Equador, Nicarágua, Peru e Uruguai (1996) e Venezuela (1998).

Muitos dos programas de avaliação foram desenvolvidos a partir de 1995, quando a V Cúpula Ibero-Americana de Chefes de Estado e de Governo aprovou, em San Carlos de Bariloche, o Programa de Cooperação para o Desenvolvimento de Sistemas Nacionais de Avaliação da Qualidade da Educação.

Também se observa, ao visitar as páginas eletrônicas de diversos ministérios de educação, que os sistemas de avaliação tal como concebidos inicialmente foram modificados ao longo dos anos. Muitos países tiveram uma experiência inicial, patrocinada geralmente pelo Banco Mundial, que serve como antecedente ao desenvolvimento e à consolidação dos sistemas atuais.

⁶ Duas agências multilaterais que possuem grande força coercitiva nas questões educacionais dos países em desenvolvimento são o Banco Interamericano de Desenvolvimento (BID) e o Grupo Banco Mundial. A análise da influência dessas organizações na avaliação paulista extrapola os limites deste artigo. Contudo, o leitor pode conferir trabalhos que se dedicam a esse tipo de reflexão em Coraggio e Torres (1997), Haddad, Tommasi e Warde (1998) e Torres (1998).

Quadro 1 – Primórdios das avaliações de sistemas educacionais nas Américas

País	Nome do sistema de avaliação ⁽¹⁾	Ano de início ⁽²⁾
Argentina	Operativos Nacionales de Evaluación del Sinec (ONE)	1993 ⁽³⁾
Bolívia	Sistema de Medición de la Calidad de la Educación (Simecal)	1996
Brasil	Sistema Nacional de Avaliação da Educação Básica (Saeb)	1990
Canadá	School Achievement Indicators Program (Saip)	1993 ⁽⁴⁾
Chile	Sistema de Medición de Calidad de la Educación (Simce)	1987 ⁽⁵⁾
Colômbia	Sistema Nacional de Evaluación de la Educación (Pruebas Saber)	1988
Costa Rica	Pruebas de Conocimientos	1986 ⁽⁶⁾
Cuba	Sistema de Evaluación de la Calidad de la Educación	1996 ⁽⁶⁾
El Salvador	Sistema Nacional de Evaluación de los Aprendizajes (Sinea)	1993
Equador	Sistema Nacional de Medición de Logros Académicos (Aprendo)	1996
Estados Unidos	National Assessment of Educational Progress State (Naep)	1988 ⁽⁷⁾
Guatemala	Sistema Nacional de Medición del Logro Académico Sistema Nacional de Evaluación, Investigación y Estándares Educativos (Sineie)	1992
Honduras	Sistema Nacional de Evaluación de la Calidad de la Educación (Sinece)	1990
México	Aprovechamiento Escolar – Carrera Magisterial	1994
Nicarágua	Evaluación del Currículo Transformado	1996
Panamá	Programa de Pruebas de Diagnóstico	1985 ⁽⁶⁾
Paraguai	Sistemas Nacionales de Evaluación del Proceso Educativo	1996 ⁽⁶⁾
Peru	Crece con Calidad y Equidad en el Rendimiento (Crece)	1996 ⁽⁸⁾
Rep. Dominicana	Sistema de Pruebas Nacionales	1991 ⁽⁶⁾
Uruguai	Unidad de Medición de Resultados Educativos	1996 ⁽⁶⁾
Venezuela	Sistema Nacional de Medición y Evaluación del Aprendizaje (Sinea)	1998 ⁽⁶⁾

Fontes: Barrera (2000), León e Vasquéz (2009).

Notas:

(1) Os nomes das avaliações do sistema referem-se às avaliações da educação básica.

(2) Os anos de início do sistema referem-se à primeira experiência sistemática conhecida em cada país.

(3) Costa (1998).

(4) Canadá. CMEC. *Assessment: Pan-Canadian Assessment Program*.

(5) Chile. Simce. Pruebas nacionales.

(6) Unesco (2008).

(7) O Naep teve início em 1969; entretanto, mudanças curriculares levaram a transformações na avaliação e o Naep com a sua estrutura atual passou a ser aplicado a partir de 1988 (Mead, 1998).

(8) Peru. Ministerio de Educación. *Que és la UMC?*

Vale ressaltar que, em documento produzido pela Unesco intitulado *National learning assessments by country and region*, fica claro que muitos sistemas foram modificando-se ao longo das últimas duas décadas, nas séries avaliadas, no tipo de avaliação (censitária ou amostral) ou nos conteúdos avaliados.

Esse movimento evidencia a crescente importância atribuída aos sistemas de avaliação, suscitando indagações como: Qual a importância dada a esses programas de avaliação em cada país? Que papel foi relegado a essas ações na política educacional dos diversos países? Houve características comuns entre as experiências dos diversos países?

Para responder tais questões, é necessário refletir sobre os objetivos dessas avaliações e os usos declarados de seus resultados, com o intuito de analisar se há características comuns entre algumas das experiências de avaliação que acontecem atualmente.

Para a construção dessa análise, foi realizada pesquisa em fontes primárias, quando disponíveis, e, principalmente, secundárias. Para a construção dos quadros, utilizaram-se os objetivos de cada sistema divulgados por Barrera⁷ (2000), Wolff (2006) e Ravela *et al.* (2006), que centraram suas análises nos países americanos, reorganizando-os de forma a propiciar uma comparação entre eles, com complementações obtidas nos *sites* relativos à educação em cada país e pesquisas realizadas no *site* do Programa de Promoção da Reforma Educativa na América Latina e Caribe (Preal), todos documentos relacionados nas referências.

Quando possível, priorizou-se o uso de materiais oficiais referentes aos primeiros sistemas de avaliação que se concretizaram nos diversos países, a fim de se analisar os objetivos declarados inicialmente. Contudo, durante a pesquisa, pode-se observar que muitos desses objetivos foram modificados ou sofreram acréscimos⁸ ao longo dos anos, e nem sempre foi possível encontrar, nas fontes consultadas, informações relativas aos sistemas antecedentes, o que pode ter gerado imprecisão no preenchimento dos quadros que serão apresentados a seguir.

Para organizar os dados relativos aos objetivos das avaliações obtidos, retomou-se a distinção entre objetivos gerais das avaliações de sistema, apresentada por Klinger, DeLuca e Miller (2008), adaptando-a a quatro categorias:

- 1) Objetivos da avaliação relacionados à *accountability*;
- 2) Objetivos da avaliação relacionados ao diagnóstico instrucional/das aprendizagens;
- 3) Objetivos da avaliação relacionados ao gerenciamento e ao controle educacionais;
- 4) Objetivos da avaliação relacionados com a seleção, classificação, premiação ou certificação.

Em alguns casos, os objetivos declarados nos documentos oficiais ou em produções dos autores de referência poderiam relacionar-se a mais

⁷ Em seu trabalho, Barrera (2000) apresenta o ano de 1994 como o da primeira iniciativa de avaliação na Argentina. Optou-se, nesse caso, pela data anunciada por Costa (1998), pelo fato de ela fazer parte da equipe de avaliação argentina. Assim, sempre que possível, buscou-se confrontar os dados expostos pela autora com aqueles obtidos em *sites* oficiais ou documentos elaborados por membros das equipes de avaliação dos países elencados no Quadro 1. Quando foram encontradas datas divergentes, optou-se pela informação oficial.

⁸ Apontar todas estas mudanças, contudo, não é objetivo deste trabalho: optou-se por restringir a pesquisa às provas e aos sistemas de avaliação referidos no Quadro 1.

de uma categoria, a depender da interpretação que lhes fosse atribuída; nesses casos, realizou-se uma opção baseada no contexto em que a informação foi fornecida no documento consultado.

A série de quadros apresentados é ilustrativa da importância atribuída à avaliação para a melhoria da qualidade da educação e para a tomada de decisões acerca da política educacional, sendo possível ressaltar, a partir da análise das informações contidas, a dimensão gerencial dos sistemas avaliativos, aqui entendida como conceituada por Afonso (2000, p. 18):

Nas organizações, a avaliação é quase sempre um instrumento fundamental em termos de gestão. Particularmente importante na estruturação das relações de trabalho (por exemplo, como instrumento de seleção, promoção e desenvolvimento profissional) a avaliação é também um instrumento importante de controle e de legitimação organizacional. As escolas (enquanto organizações complexas) têm, com algumas similitudes, utilizado a avaliação para esses mesmos fins.

Compreende-se que a dimensão de gestão e gerenciamento do sistema educacional consiste no uso dos resultados da avaliação para a administração, o controle e o direcionamento das questões da educação. Essa perspectiva é ilustrada por objetivos que aparecem nas avaliações, como: produzir informação gerencial, proporcionar tomada de decisões para melhorar a qualidade, promover retroalimentação curricular, avaliar o impacto das políticas e dar continuidade à reforma educativa.

Contudo, também é possível perceber uma intenção de utilizar as avaliações com finalidade de diagnóstico de aprendizagens, muitas vezes para produzir informações visando à retroalimentação do sistema educativo. Os objetivos que poderiam propiciar um uso mais formativo dos resultados são, entre outros: conhecer as aprendizagens; disponibilizar informações para o professor e/ou a escola; avaliar o trabalho docente, os currículos, os textos; promover retroalimentação curricular; avaliar ganhos com reformas curriculares. Vários países declararam como objetivo para os processos de planejamento curricular e avaliação do aprendizado o uso das informações pelo professor. Contudo, nos demais objetivos deste grupo, observa-se que as similaridades entre os países são menos evidentes. Finalmente, no que se refere à categoria de *accountability* (prestação de contas e responsabilização pelos resultados), pode-se observar que diversos países incorporam esse discurso na sistematização de suas metas relacionadas à avaliação. As ideias de medição e monitoramento da qualidade estão presentes nas intenções declaradas pela maioria dos países que adotaram algum tipo de sistema de avaliação.

De forma geral, nota-se que, num momento inicial de consolidação dos sistemas de avaliação educacional, os objetivos declarados voltavam-se mais para gerenciamento e tomada de decisões políticas e prestação de contas e responsabilização (*accountability*), sendo possível observar características semelhantes entre os países.

Quadro 2 – Objetivos da avaliação relacionados com a *accountability*

Objetivos	País																Total				
	Argentina	Bolívia	Brasil	Canadá	Chile	Colômbia	Costa Rica	El Salvador	Equador	Estados Unidos	Guatemala	Honduras	México	Nicarágua	Panamá	Paraguai		Peru	Rep. Dominicana	Uruguai	Venezuela
Medir e controlar a qualidade	X		X		X	X	X	X				X	X					X		X	10
Avaliar a educação pública, a eficácia e a eficiência do sistema escolar											X					X	X				3
Medir, obter informações e comparar o nível de aprendizado que os alunos estão alcançando em diferentes Estados				X						X											2
Estimular a participação de outros setores da sociedade civil nas decisões relativas à educação	X																				1
Fornecer informações para famílias/alunos					X																1
Informar a sociedade civil sobre o quanto os sistemas educacionais estão indo ao encontro das necessidades dos alunos e da sociedade				X																	1
Consolidar um sistema educativo de informação ou um serviço de informações sobre educação		X				X	X					X						X			5
Avaliar o impacto de políticas																			X		1
Avaliar o trabalho docente, o currículo e os textos					X				X												2

Quadro 3 – Objetivos da avaliação relacionados com o diagnóstico instrucional/das aprendizagens

Objetivos	País																	Total			
	Argentina	Bolívia	Brasil	Canadá	Chile	Colômbia	Costa Rica	El Salvador	Equador	Estados Unidos	Guatemala	Honduras	México	Nicarágua	Panamá	Paraguai	Peru		Rep. Dominicana	Uruguai	Venezuela
Medir a aprendizagem; avaliar a qualidade da aprendizagem e o progresso do rendimento escolar; conhecer as aprendizagens		X			X	X	X		X	X	X	X	X	X	X	X			X		13
Conhecer as aprendizagens e os fatores associados; identificar desigualdades e insuficiências; avaliar a motivação dos alunos	X	X				X									X						4
Apoiar a capacitação docente em serviço					X								X								2
Fornecer informações para o professor para os processos de planejamento curricular e de avaliação do aprendizado		X				X	X	X	X			X						X	X	X	9
Disponibilizar informações à comunidade acadêmica para estimular projetos de pesquisa em educação; apoiar investigações						X			X												2
Produzir materiais de apoio, insumos para escolas, professores e supervisores	X	X											X								3
Fornecer informações para famílias/alunos/comunidade para compreensão do projeto pedagógico; promover a valorização da escolaridade					X	X	X														3
Avaliar o trabalho docente, o currículo e os textos									X												1
Promover retroalimentação curricular; avaliar ganhos com reformas curriculares								X							X						2
Promover uma cultura de avaliação; gerar competência técnica na área e atitude positiva em relação à avaliação	X		X																		2
Proporcionar aprendizagem no interior do sistema; promover discussão nacional sobre qualidade; estimular/promover melhoria das aprendizagens							X			X						X			X		4

Quadro 4 – Objetivos da avaliação relacionados com o gerenciamento e o controle educacionais

Objetivos	País																	Total			
	Argentina	Bolívia	Brasil	Canadá	Chile	Colômbia	Costa Rica	El Salvador	Equador	Estados Unidos	Guatemala	Honduras	México	Nicarágua	Panamá	Paraguai	Peru		Rep. Dominicana	Uruguai	Venezuela
Monitorar a qualidade; contribuir para o melhoramento da qualidade	X		X		X	X	X	X				X	X			X		X		X	11
Contribuir para a melhoria da equidade do ensino					X																1
Avaliar a eficácia e a eficiência do sistema escolar; supervisionar falhas e avanços; supervisionar o sistema	X				X						X					X	X				5
Subsidiar tomada de decisões políticas e tomada de decisões para melhorar a qualidade	X	X	X			X	X	X			X	X	X	X		X	X	X	X	X	15
Consolidar um sistema educativo de informação ou serviço de informações sobre educação		X				X	X					X						X			5
Dar continuidade à reforma educativa; implementar e avaliar intervenções educacionais e políticas		X			X																2
Avaliar o trabalho docente, o currículo e os materiais instrucionais; avaliar a escola e os seus mecanismos avaliativos				X	X				X												3
Produzir informação gerencial; melhorar a gestão institucional	X		X						X												3
Promover descentralização do sistema	X																				1
Avaliar ganhos com reformas curriculares								X							X						2
Bonificar os professores; prover incentivos salariais para professores que têm alunos com altos escores					X							X									2

Quadro 5 – Objetivos da avaliação relacionados com a premiação, seleção, classificação ou certificação

Objetivos	País															Totais						
	Argentina	Bolívia	Brasil	Canadá	Chile	Colômbia	Costa Rica	El Salvador	Equador	Estados Unidos	Guatemala	Honduras	México	Nicarágua	Panamá		Paraguai	Peru	Rep. Dominicana	Uruguai	Venezuela	
Bonificar os professores; prover incentivos salariais para professores que têm alunos com altos escores					X								X									2
Monitorar ou certificar o conhecimento dos alunos em relação aos padrões nacionais					X																	1
Avaliar conhecimento/capacidade dos professores													X									1

Objetivos relacionados ao diagnóstico instrucional e às aprendizagens apareceram reportados de forma mais dispersa nas fontes consultadas. Apesar de muitos países declararem a intenção de medir as aprendizagens, avaliar a qualidade da aprendizagem, entre outras, não foram encontradas muitas informações sobre a que serviria o diagnóstico realizado. Também nota-se, pela análise do Quadro 3, a preocupação de alguns países em propiciar informações que subsidiem o trabalho dos professores na sala de aula.

Casos mais raros foram as declarações de objetivos de alto impacto, como certificações, premiações, etc., o que não permite definir se tais objetivos estão excluídos das intenções dos programas de avaliação ou se, apenas, não são declarados nos documentos consultados. Além do mais, devido à natureza amostral da maioria dos programas nacionais de avaliação, o caráter de seleção, premiação ou certificação que alguns sistemas poderiam intentar fica comprometido. De qualquer forma, a avaliação do trabalho do professor e a sistematização de bônus ou atividades de capacitação docente foram objetivos compartilhados por alguns sistemas.

Observa-se, contudo, que, apesar de alguns dos objetivos declarados serem bastante propícios em contextos de esforço para a democratização da educação e a melhoria de sua qualidade e equidade, muitas críticas a esses sistemas ainda são produzidas. Entender os problemas que pesquisadores têm apontado no que se refere à consolidação deles é o objetivo da próxima seção.

Extrapolando a análise: algumas críticas realizadas aos sistemas nacionais de avaliação

Conforme visto anteriormente, a avaliação de sistemas educacionais não é uma novidade e muitos países a incorporaram nas reformas que se processaram nos últimos 20 anos, enfatizando a dimensão de controle da qualidade de ensino, entendida numa perspectiva de gestão e racionalização dos recursos disponíveis. Observa-se, assim, uma ênfase em três aspectos que, segundo Broadfoot (1996), têm sido objetivos centrais das avaliações: regulação da competição, avaliação da competência e controle (tanto do indivíduo quanto do sistema).

Outro aspecto retratado por essa autora e que ecoa uma preocupação já apontada anteriormente por Costa (1998), ao relatar a experiência argentina, refere-se à ênfase que vem sendo dada à análise técnica da avaliação, deixando-se para o segundo plano a análise dos propósitos e efeitos que esta tem nos sistemas em que é aplicada. Segundo ela:

A maior parte da literatura sobre avaliação de desempenho se preocupa com as técnicas de avaliação. Enquanto muitos estudos se preocupam com o desenvolvimento e o refinamento de diferentes abordagens de medida educacional, muitos outros oferecem poderosas críticas às técnicas de avaliação de desempenho e às falhas dessas mesmas técnicas quando aplicadas. [...]

Então poderosos e intensos debates sobre técnicas concorrentes, os quais têm tradicionalmente caracterizado a pesquisa sobre avaliação educacional auxiliaram a excluir da discussão da questão mais fundamental relacionada aos propósitos e efeitos dessa atividade. (Broadfoot, 1996, p. 12-13).

A preocupação com os propósitos e os efeitos das avaliações, ou com o debate educacional mais amplo, como chama Costa, deveria ser alvo de mais estudos e reflexão, mesmo porque os propósitos da avaliação definem a técnica a ser utilizada e, conseqüentemente, a interpretação e o uso que será feito de seus resultados.

O fato de enfatizar a necessidade de estudos mais aprofundados sobre os efeitos das avaliações e seus propósitos não significa que Broadfoot considere o debate técnico já resolvido. Em sua obra já citada, a autora questiona a adequação dos instrumentos utilizados, colocando em dúvida se são capazes de medir as conquistas em termos de habilidades e competências realizadas pelos alunos, ou se apenas verificam conteúdos conceituais. Ora, tais dúvidas apontam para a necessidade de discussão de aspectos técnicos, mas esta deveria ser realizada em conjunto com a discussão sobre os objetivos dos sistemas, os usos dos resultados e a difusão. Principalmente em um momento em que já se passaram mais de 20 anos da implantação dos primeiros sistemas de avaliação, o debate técnico aliado ao pedagógico poderia ser retomado para que ajustes fossem realizados em prol da consolidação de avaliações mais significativas, ou seja, cujos resultados pudessem alimentar debates e impulsionar o desenvolvimento do sistema educacional.

Outras questões a serem discutidas são a seleção dos conteúdos e habilidades que embasam as avaliações e a definição de níveis ou padrões de rendimento, tanto em seu caráter técnico (o que medir e como medir) quanto pedagógico. Nesse sentido, Broadfoot questiona se as matrizes e os parâmetros de avaliação não estariam orientados por uma visão da elite sobre conteúdos básicos e padrões mínimos, não necessariamente condizentes com as reais necessidades sociais e as possibilidades de desenvolvimento dos alunos.

Por outro lado, a definição de esdandartes baixos de proficiência poderia, com o tempo, gerar um empobrecimento curricular, impulsionando os sistemas a direcionarem os esforços para garantir o ensino do que é “cobrado” nas avaliações, deixando de lado outros conteúdos fundamentais para a educação básica das crianças e dos adolescentes em idade escolar.

Além desses questionamentos elaborados por membros da academia, mais distantes dos órgãos responsáveis pelo desenho e gerenciamento das avaliações, existem críticas que vêm sendo formuladas por pesquisadores e colaboradores envolvidos, direta ou indiretamente, com as equipes de avaliação de diversos países.

No seminário internacional “La dimensión política de la evaluación de la calidad educativa en Latinoamérica”, realizado em 2002 pela Unesco, cujo objetivo era fazer um balanço das avaliações implementadas na América Latina, ficou evidente que, apesar dos esforços realizados especialmente pelos países latino-americanos para a melhoria da qualidade de ensino, os dados obtidos por meio dessas avaliações mostram que pouco tem se alterado o quadro educacional da maioria dos países que as empregam enquanto instrumento de gestão. Essa constatação causa, segundo Tedesco (2003), certo mal-estar entre os gestores e os tomadores de decisão, traduzido por críticas e tensões que vêm à tona quando se procura debater a questão da avaliação sistêmica.

Tentando entender essas tensões, Iaies (2003) alerta para o fato de que é imprescindível que se supere o debate eminentemente técnico que dominou a área nos últimos 15 anos, atentando para a perspectiva política subjacente à avaliação. Para o autor, urge recuperar o sentido de avaliar os sistemas educacionais, e, para isso, é necessário assumir que os objetivos desse tipo de avaliação estão estreitamente vinculados a posturas político-filosóficas dos gestores.

Subentende-se, a partir da argumentação, que os resultados das provas não têm apresentado o impacto esperado na melhoria da qualidade do sistema educacional, porque as avaliações configuram-se em modelos cuja prioridade é *medir* o rendimento dos alunos, e não *avaliar*⁹ o sistema como um todo. A inexistência de clareza sobre os objetivos da avaliação e a dificuldade de definir e entrar em consenso sobre padrões de qualidade claros – que permitam comparar longitudinalmente os resultados obtidos e possam ser utilizados para analisar as possíveis mudanças que se operam a partir das políticas e programas implementados – fazem com que a preocupação recaia sobre os resultados das provas e sua dimensão

⁹ A diferenciação entre avaliação e medida tem sido alvo de atenção de inúmeros pesquisadores da área e pode ser encontrada nos trabalhos de Vianna (2000).

técnica, não se enfatizando as análises contextuais que permitiriam um melhor entendimento da situação educacional e uma intervenção mais efetiva. Como explica Iaies (2003, p. 18):

Os sistemas educativos deixaram de trabalhar para melhorar a qualidade e a equidade educativa e passaram a trabalhar para o melhoramento dos resultados das avaliações. [...] Os dispositivos e seus produtos têm tido maior impacto na construção do imaginário educativo da sociedade, que na transformação das estratégias educativas.

Nesse sentido, pode-se dizer que as decisões tomadas a partir dos resultados obtidos, muito relacionadas à redefinição de conteúdos curriculares na formulação de indicadores de uma “boa educação” e na definição de habilidades e competências a serem desenvolvidas nos alunos, parecem não ecoar na escola, não fazer sentido para aqueles que, em última instância, fazem as reformas acontecerem, ou seja, os professores e os técnicos escolares. Essas avaliações, em geral, não atentam para outras demandas que as escolas enfrentam, como as necessidades sociais do público que tem acessado a escolarização formal (e que, anteriormente, estava excluído dela), as características de formação dos profissionais que se dedicam à educação, a garantia de condições de educabilidade mínimas para que os estudantes possam produzir suas aprendizagens (Iaies, 2003).

Dessa forma, haveria uma distância entre as considerações técnicas que embasam a avaliação e o debate político-educacional que precisa ser enfrentado tanto na escola como nos níveis centrais pelos gestores do sistema. Como provoca o autor:

Construíram-se indicadores que se definiram tecnicamente, e que consideram quase com exclusividade as habilidades acadêmicas. Nossos índices não consideram o aumento das taxas de escolarização, a capacidade do sistema para homogeneizar atores de uma sociedade cada dia mais segmentada, dar conta dos novos públicos que a escola tem sido capaz de abrigar, da capacidade de conter outras realidades sociais, etc. E essas definições implicam uma tomada de postura ideológica, utilizar umas variáveis e abandonar outras; o certo é que a experiência dos 90 faz pensar mais em uma “não tomada” de posição política, no sentido de que os tomadores de decisões não se posicionaram nesse ponto. (Iaies, 2003, p. 20-21).

Patrícia Broadfoot já havia denunciado essa distância entre as propostas de avaliação, as necessidades socioeducacionais e a dinâmica escolar em sua obra de 1996 sobre o papel social da avaliação. Segundo a autora, faz-se necessário aprofundar o debate acerca do papel que as avaliações têm assumido na sociedade atualmente, principalmente quando essas avaliações relacionam-se a políticas e programas educacionais. Talvez essa possa ser uma das razões pelas quais essas propostas não têm impactado nos sistemas de forma significativa, mesmo quando aliadas a outras ações, dentro de uma reforma educativa mais ampla.

Ao apreciar os sistemas de avaliação implementados por diversos países latino-americanos, Iaies (2003) mostra que, apesar da diversidade

de modelos utilizados, há aspectos convergentes entre eles, quando o objeto de análise e comparação são os resultados obtidos pelos sistemas desde sua implantação. Revela ainda que, de maneira geral, não houve uma mudança significativa no rendimento dos alunos nos testes, e arrisca afirmar que essa invariabilidade não é apenas quantitativa, mas também qualitativa. Ou seja, além do número de acertos nos testes não ter variado muito, o tipo de teste em que os alunos têm um nível de acerto mais elevado também não teria mudado.

Outro aspecto que pode ser observado nos sistemas, segundo esse autor, é que as equipes técnicas se ocupam mais na produção da informação, obtida mediante os instrumentos utilizados, do que na exploração do potencial das informações para a mudança educativa. Ressalta ainda que, na maioria dos países, há um problema na comunicação das informações obtidas. Mesmo nos raros sistemas de avaliação em que há maior preocupação com uma análise refletida dos resultados e seu uso para a tomada de decisão, a elaboração e a divulgação dessa análise é tão demorada que acaba sendo pouco aproveitada para a reelaboração de estratégias pedagógicas ou a reorientação de políticas educacionais. Para Iaies (2003, p. 28-29), a forma como esses resultados têm sido comunicados gera uma sensação de deterioração do sistema educativo e de conseqüente imobilidade ou impossibilidade de este sistema melhorar qualitativamente.

Outro fato apontado na literatura relaciona-se aos usos objetivos dos resultados da avaliação e aos efeitos da avaliação propriamente dita. Os estudos concentram-se na análise da avaliação das aprendizagens ou, ainda, na compreensão da avaliação sistêmica, sem enfrentarem profundamente uma questão que é essencial: *o uso da avaliação*. Um exemplo dessa lacuna refere-se ao uso da avaliação como fonte de dados para a elaboração de programas de formação de professores. Por meio do estudo bibliográfico realizado, observa-se que a relação entre os usos dos resultados da avaliação e sua influência no aprimoramento da formação oferecida aos professores tem sido pouco explorada pelos estudiosos da educação e áreas afins. Mesmo importantes trabalhos que refletem sobre as possibilidades do uso dos resultados da avaliação visando à melhoria da qualidade de ensino (Vianna, 2003) não chegam a aprofundar a relação existente entre esses resultados e as possibilidades que apontam à formação dos professores, ficando clara a necessidade de estudos que busquem iluminar essa relação.

Dentre as muitas dimensões desse debate, encontram-se a influência da divulgação dos resultados sobre o seu uso, a relação entre os objetivos da avaliação sistêmica e às informações por ela produzidas e a discussão política e ideológica da aplicação dos resultados, já apontada anteriormente.

Considere-se primeiramente a dimensão que relaciona os usos dos resultados à sua divulgação e comunicação. Retomando a discussão já iniciada por Iaies, Pablo Halpern Britz (2003) explica que seria importante minimizar o caráter político-ideológico que a divulgação dos resultados

tem assumido na maioria das avaliações que ocorrem em países da América Latina. Isso porque, tecnicamente, as generalizações e a divulgação dos resultados brutos não se sustentam, já que as avaliações são desenhadas para obter informações mais gerais, que auxiliem na gestão dos sistemas educacionais como um todo. Poucos são os países que se dedicam ao desenvolvimento de instrumentos que permitam a obtenção de informações que possam ser utilizadas em nível local. Halpern Britz (2003, p. 94) vê problemas na divulgação dos resultados da forma como vem sendo realizada e aponta, como possíveis consequências, a limitação do uso da informação e a tendência de recuo na divulgação de dados para os meios de comunicação em massa:

Um olhar mais pessimista sobre esta tensão entre os debates técnico e público na educação assinala que a situação se fará crescentemente complexa. A massificação dos indicadores impedirá de circunscrever seu uso ao âmbito para o qual foram desenhados: cada vez mais será usado como um *ranking* para tomar decisões imediatas pela família, cada vez terá mais sentido transformá-lo em base de um debate político.

Esta desnaturalização dos alcances dos indicadores finalmente pode levar ao imobilismo: estariam dispostas as autoridades a seguir dando transparência a sua gestão sem que os instrumentos se voltem contra si? Estariam dispostos os instrumentos de comunicação a não disporem destes dados?

Ainda nessa linha de argumentação, José Joaquín Brunner traz uma importante contribuição que pode iluminar a discussão sobre a problemática do impacto das avaliações nas escolas ao analisar a questão da divulgação dos resultados das avaliações, especialmente nos jornais e periódicos. Segundo o autor, os periódicos não atentam para a complexidade dos resultados e o contexto em que são produzidos:

[Os periódicos] se limitam a ressaltar unidimensionalmente os produtos finais destes processos – a saber, os resultados enquanto pontuação – e a organizá-los em um esquemático *ranking* de países ou tipos de estabelecimento, suprimindo toda referência ao contexto onde estes resultados se obtêm. Com isso, se banaliza a informação que aportam os estudos nacionais e internacionais de medida do rendimento escolar e, o que é mais grave, se distorce a opinião pública e se limitam seus efeitos. (Brunner, 2003, p. 81-82).

Seria ingenuidade negar que a opinião pública tem forte influência sobre as decisões de gestores e políticos. Tanto os jornais quanto a literatura educacional, de tempos em tempos, apresentam exemplos nos quais gestores educacionais voltaram atrás na proposição de ações por estas não terem sido bem recebidas por professores e pela população educacional.¹⁰

É importante observar que, apesar de muitos pesquisadores se dedicarem ao estudo das avaliações de sistema e aos problemas dela decorrentes, essas reflexões parecem ter pouco impacto entre os gestores das políticas, visto que a incorporação da crítica e o uso dos conhecimentos

¹⁰ Seria interessante um aprofundamento, por meio de estudos empíricos, sobre a influência que a divulgação dos resultados tem sobre a opinião pública e a relação entre esta e os usos dos resultados das avaliações e o impacto delas nos programas educacionais.

elaborados para a reorientação das avaliações parecem estar fora do rol de intenções dos técnicos responsáveis pelo gerenciamento das avaliações. Broadfoot (1996) ilumina essa questão identificando uma desarticulação, um distanciamento entre os acadêmicos da área de avaliação e os responsáveis pela definição das políticas, que se mantêm isolados em seus respectivos campos de atuação.

Apesar disso, é possível localizar os limites dos usos dos resultados das avaliações relacionados principalmente às dimensões técnica, política e ideológica, conforme apontado anteriormente. Fica claro que o debate sobre avaliação deve ser encarado em sua complexidade, a fim de que os sistemas já consolidados possam se desenvolver e produzir informações que permitam superar os usos políticos e ideológicos que vêm sendo feitos dos resultados. Esse debate contribuiria efetivamente para iluminar a problemática educacional, possibilitando a realização de ações direcionadas para a melhoria da educação, o que exige uma análise que ultrapasse a comparação de resultados quantitativos sobre níveis de aprendizagem e considere, também, aspectos curriculares, de infraestrutura e de formação docente, dentre outros.

Referências bibliográficas

AFONSO, Almerindo. *Políticas educativas e avaliação educacional*. Portugal: Universidade do Minho, 1998.

_____. *Avaliação educacional: regulação e emancipação – para uma sociologia das políticas educativas contemporâneas*. São Paulo: Cortez, 2000.

ARGENTINA. *Programa de evaluación de la calidad educativa*. Disponível em: <www.mejujuy.gov.ar/ecalidad>. Acesso em: 2 mar. 2010.

BARRERA, Susana. Políticas educacionales, calidad de la educación en latinoamérica y aporte de la evaluación como herramienta de gerencia social. In: INSTITUTO INTERNACIONAL DE INTEGRACIÓN DEL CONVENIO ANDRÉS BELLO. *El proceso de la integración educativa – Mesa redonda “Convergencia Mercosur-CAN”*. La Paz, Bolívia, 2000. p. 79-114.

BLOOM, Benjamin S. et al. *Taxonomia dos objetivos educacionais*. 2 v. Porto Alegre: Globo, 1972.

BOLÍVIA. *Sistemas educativos nacionales: Bolívia*. Cap. 17. Disponível em: <<http://www.rieoei.org/quipu/boli17.pdf>> . Acesso em: 03/03/2010.

BONAMINO, Alicia C. *Tempos de avaliação educacional: o SAEB, seus agentes, referências e tendências*. Rio de Janeiro: Quartet, 2002.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). *Saeb: perguntas mais frequentes*. Disponível em: <http://www.inep.gov.br/basica/saeb/perguntas_frequentes.htm>. Acesso em: 2 mar. 2010.

BROADFOOT, Patricia. *Education, assessment and society: a sociological analysis*. Buckingham: Open University Press, 1996.

BRUNNER, José Joaquín. Límites de la lectura periodística de resultados educacionales. In: IAIES, Gustavo et al. *Evaluar las evaluaciones: una mirada política acerca de las evaluaciones de la calidad educativa*. Buenos Aires: IPE/UNESCO, 2003: 67-84.

CANADA. Council of Ministers of Education (CMEC). *Assessment: Pan-canadian Assessment Program*. Disponível em: <<http://www.cmec.ca/Programs/assessment/pancan/Pages/default.aspx>>. Acesso em: 3 mar. 2010.

CHILE. Ministerio de Educación. Sistema de Medición de Calidad de la Educación (Simce). *Presentación: Que és el Simce*. Disponível em: <www.simce.cl>. Acesso em: 1º mar. 2010.

COLOMBIA. Ministerio de Educación Nacional. *Educación básica y media*. Disponível em: <<http://www.mineduccion.gov.co/1621/article-156285.html>>. Acesso em: 2 mar. 2010.

COMISIÓN ECONÓMICA PARA AMÉRICA LATINA Y EL CARIBE (Cepal); ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, A CIÊNCIA E A CULTURA (Unesco). *Financiamiento y gestión de la educación en América Latina y el Caribe* [versión preliminar]. Documento preparado para ser presentado en el 30º periodo de sesiones de la Cepal realizado en San Juan, Puerto Rico, del 28 de junio al 2 de julio de 2004. Disponível em: <www.cepal.org/id.asp?id=15462>.

CORAGGIO, José Luis; TORRES, Rosa Maria. *La educación según el Banco Mundial: un análisis de sus propuestas y médios*. Buenos Aires: CEM, Minõ y Dávila, 1997.

COSTA, Marta Elena. A avaliação de desempenho na área de língua na Argentina: a utilização de provas objetivas. *Série Idéias*, FDE, São Paulo, n. 30, p. 39-88, 1998.

DIAS SOBRINHO, José. Campo e caminhos da avaliação: a avaliação da educação superior no Brasil. In: FREITAS, Luiz Carlos de. *Avaliação: construindo o campo e a crítica*. Florianópolis: Insular, 2002. p. 13-62.

EL SALVADOR. *Sistema Nacional de Evaluación de los Aprendizajes (Sinea)*. Disponível em: <http://www.mined.gob.sv/downloads/Documentos/sinea2005_0_.pdf>. Acesso em: 1º mar. 2010.

FREITAS, Luiz Carlos de (Org.). *Avaliação: construindo o campo e a crítica*. Florianópolis: Insular, 2002.

HADDAD, Sérgio; TOMMASI, Livia de; WARDE, Miriam Jorge. *O Banco Mundial e as políticas educacionais*. São Paulo: Cortez, 1998.

HALPERN BRITZ, Pablo. Indicadores y sus usos en educación: una tensión no resuelta. In: IAIES, Gustavo et al. *Evaluar las evaluaciones: una mirada política acerca de las evaluaciones de la calidad educativa*. Buenos Aires: IPE, Unesco, 2003. p. 85-96. Disponível em: <www.oei.es/evaluacioneducativa/evaluar_evaluaciones_iipe.pdf>.

HIDALGO, Ângela Maria; SILVA, Ileizi L. F. (Orgs.). *Educação e Estado: as mudanças nos sistemas de ensino do Brasil e do Paraná na década de 90*. Londrina: Ed. UEL, 2001.

IAIES, Gustavo. Evaluar las evaluaciones. In: IAIES, Gustavo et al. *Evaluar las evaluaciones: una mirada política acerca de las evaluaciones de la calidad educativa*. Buenos Aires: IPE, Unesco, 2003. p. 15-36. Disponível em: <www.oei.es/evaluacioneducativa/evaluar_evaluaciones_iipe.pdf>.

KLINGER, Don A., DE LUCA, Christopher; MILLER, Tess. The evolving culture of large-scale assessments in Canadian education. *Canadian Journal of Educational Administration and Policy*, n. 76, July 3, 2008. Disponível em: <<http://www.umanitoba.ca/publications/cjeap/articles/klinger.html>>.

LEÓN, Adelina; VÁZQUEZ, Silvia A. Report on educational evaluation in Argentina and the surrounding debates. February 2009. Disponível em: <<http://www.idea-network.ca/userfiles/image/file/EvaluationReportArgentinaCTERA.pdf>>. Acesso em: 6 mar. 2010.

MADAUS, George; SCRIVEN, Michael; STUFFLEBEAM, Daniel. Program evaluation: an historical overview. In: MADAUS, George; SCRIVEN, Michael; STUFFLEBEAM, Daniel (Ed.). *Evaluation models: viewpoints on Educational and Human Services Evaluation*. Boston: Kluwer-Nijhoff, 1983. p. 3-22.

MEAD, Nancy. Programas Mínimos e Testes de Avaliação: a experiência norte-americana. *Série Idéias*, FDE, São Paulo, n. 30, p. 137-146, 1998.

MÉXICO. Secretaria de Educación Pública. *Programas: evaluación externa de programas federales*. Disponível em: <http://sep.gob.mx/wb/sep1/programas_federales_sujetos_a_evaluacion_externa>. Acesso em: 4 mar. 2010.

_____. *¿Qué es el INEE?*. Disponível em: <http://www.inee.edu.mx/>. Acesso em: 4 mar. 2010.

PARAGUAI. Sistema Nacional de Evaluación del Proceso Educativo. *Pruebas nacionael*. Disponível em: <http://www.oei.es/quipu/paraguay/prueba_nacional.pdf>. Acesso em: 5 mar. 2010.

PERU. Ministerio de Educación. *¿Qué es la UMC?* Disponível em: <http://www2.minedu.gob.pe/umc/index2.php?v_codigo=34&v_plantilla=2>. Acesso em: 5 mar. 2010.

RAVELA, Pedro et al. Los proximos pasos ¿Como avanzar en la evaluación de aprendizagens en América Latina? In: ARREGUI, Patrica (Ed.). *Sobre estándares y evaluaciones en América Latina*. Santiago: Grupo de Trabajo sobre Estándares y Evaluación, Preal, 2006. p. 53-122. Disponível em: <<http://www.thedialogue.org/page.cfm?pageID=32&pubID=388>>.

SHADISH, William; COOK, Thomas; LEVINTON, Laura. *Foundations of program evaluation: theories of practice*. Newbury Park, CA: Sage, 1991.

SOUSA, Sandra Maria Zakia Lian . Avaliação do rendimento escolar como instrumento de gestão educacional. In: OLIVEIRA, Dalila A. de (Org.). *Gestão democrática da Educação: desafios contemporâneos*. Petrópolis, RJ: Vozes, 1997. p. 264-284.

_____. 40 anos de contribuição à avaliação educacional. In: COSTA, Albertina de O.; MARTINS, Ângela Maria; FRANCO, Maria Laura P. B.(Orgs.). *Uma história para contar: a pesquisa na Fundação Carlos Chagas*. São Paulo: Annablume, 2004. p. 171-200.

_____. Ensino fundamental: avaliando as políticas de avaliação educacional. Texto apresentado no Seminário "Avaliar para quê? Avaliando as políticas de avaliação educacional", realizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), em Brasília, no dia 8 de abril de 2003.

SOUSA, Sandra Maria Zákia Lian; FREITAS, D. N. T. Políticas de avaliação e gestão educacional Brasil, década de 1990 aos dias atuais. *Educação em Revista*, UFMG, Belo Horizonte, n. 40, p. 165-185, 2004.

TEDESCO, Juan Carlos. Prólogo. In: IAIES, Gustavo et al. *Evaluar las evaluaciones: una mirada política acerca de las evaluaciones de la calidad educativa*. Buenos Aires: IPE, Unesco, 2003. p. 11-14. Disponível em: <www.oei.es/evaluacioneducativa/evaluar_evaluaciones_iipe.pdf>.

TORRES, Rosa Maria. Tendências da formação docente nos anos 90. In: WARDE, M. (Org.). *Novas políticas educacionais: críticas e perspectivas*. São Paulo: PUC, 1998. p. 173-191.

UNESCO. National learning assessments by region and country. In: EDUCATION FOR ALL: GLOBAL MONITORING REPORT. *Education for All by 2015: will we make it?*. 2008. Disponível em: <<http://www.unesco.org/education/gmr2008/annexes/annex3.pdf>>. Acesso em: 6 mar. 2010

UNITED STATES. Department of Education. *An introduction to Naep: National Assessment of Educational Progress*. 2008. Disponível em: <http://nces.ed.gov/nationsreportcard/pdf/about/introduction_to_naep_2008.pdf>. Acesso em: 6 mar. 2010.

URUGUAY. Administración Nacional de Educación Primaria. *Estudio de evaluación de impacto de la educación inicial en el Uruguay*. Montevideo, 2000. Disponível em: <http://www.oei.es/quipu/uruguay/evaluacion_inicial.pdf>. Acesso em: 3 mar. 2010.

VIANNA, Heraldo. *Avaliações nacionais em larga escala: análises e propostas*. São Paulo: FCC, 2003.

_____. Avaliação: considerações teóricas e posicionamentos. In: _____. *Avaliação educacional: teoria, planejamento, modelos*. São Paulo: Ibrasa, 2000. p. 21-46.

_____. Avaliação educacional: uma perspectiva histórica. *Estudos em Avaliação Educacional*, São Paulo, n. 12, p. 7-24, jul./dez. 1995.

WOLFF, Laurence. Las evaluaciones educacionales en América Latina: avance actual y futuros desafíos. In: ARREGUI, Patricia (Ed.). *Sobre estándares y evaluaciones en América Latina*. Santiago: Grupo de Trabajo sobre Estándares y Evaluación, Preal, 2006. p. 13-52. Disponível em: <<http://www.thedialogue.org/page.cfm?pageID=32&pubID=388>>.

WORTHEN, Blaine; SANDERS, James; FITZPATRICK, Jody. *Avaliação de programas: concepções e práticas*. São Paulo: Gente, 2004. p. 15-96.

Adriana Bauer é doutoranda vinculada à Faculdade de Educação da Universidade de São Paulo, na área Estado, Sociedade e Educação.

dri_bauer@yahoo.com.br // dri.bauer@usp.br

Recebido em 16 de novembro de 2009.

Aprovado em 19 de abril de 2010.