
OUTROS TEMAS

SISTEMAS ESTADUAIS DE AVALIAÇÃO: USO DOS RESULTADOS, IMPLICAÇÕES E TENDÊNCIAS

SANDRA ZÁKIA SOUSA

sanzakia@usp.br

ROMUALDO PORTELA DE OLIVEIRA

romualdo@usp.br

Professores da Faculdade de Educação da Universidade de São Paulo

RESUMO

O artigo apresenta resultados de pesquisa que analisa sistemas de avaliação implementados por cinco unidades federadas (Bahia, Ceará, Minas Gerais, Paraná e São Paulo), considerando as características vigentes em 2005-2007. Tendo como referência suas especificidades, busca-se explorar como os resultados produzidos vêm informando a formulação e implementação de políticas educacionais, bem como discutir seu potencial para tornar-se um marco da política educacional que efetivamente interfere na gestão das redes de ensino e das escolas. Observou-se que os sistemas tendem a apresentar características semelhantes e tomam como principal referência o Sistema de Avaliação da Educação Básica, de âmbito nacional. Quanto ao uso dos resultados, são tênues ou inexistentes seus impactos ainda que precedidos, na fala de seus propositores, de ampla retórica presente na literatura sobre suas potencialidades. No entanto, a partir do balanço feito pelos próprios gestores dos sistemas, é possível identificar a preocupação de tornar tais sistemas mais efetivos. Assim, já neste estudo, identificam-se iniciativas que procuram transformar a utilização dos dados obtidos em instrumentos de gestão, como as que propõem bônus para professores e funcionários.

POLÍTICAS EDUCACIONAIS – AVALIAÇÃO DE ESTUDANTES – ADMINISTRAÇÃO ESCOLAR – SAEB

ABSTRACT

STATE ASSESSMENT SYSTEMS: THE USE OF RESULTS, IMPLICATIONS AND TRENDS. This article presents the results of research that analyzes the assessment systems implemented by five brazilian states (Bahia, Ceara, Minas Gerais, Parana and Sao Paulo), considering their characteristics in 2005-2007. Taking into account their specificity, an attempt is made

to explore how the results have been used for formulating and implementing educational policies, as well as their potential for becoming a milestone in educational policy that effectively intervenes in the management of schools systems and schools. It was observed that the systems tend to have similar characteristics, their main point of reference being the Basic Education Assessment System, on national level. As for using the results, their impact is at best tenuous, or even non-existent, even though, in the words of their proponents, it is preceded by powerful rhetoric, as expressed in the literature about its potential. However, from an examination of the systems carried out by the managers themselves, it is possible to identify a concern with making them more effective. So, even in this study initiatives have already been identified that seek to transform the use of the data obtained into a management tool, as the ones that propose offering bonuses to teachers and staff.

EDUCATIONAL POLICIES – STUDENTS EVALUATION – EDUCATIONAL ADMINISTRATION – SAEB

A partir da década de 1990, a avaliação de sistemas escolares passou a ocupar posição central nas políticas públicas de educação, sendo recomendada e promovida por agências internacionais, pelo Ministério da Educação e por Secretarias de Educação de numerosos estados brasileiros, como elemento privilegiado para a realização das expectativas de promoção da melhoria da qualidade do ensino básico e superior. Os diversos níveis e modalidades de ensino – da educação básica à pós-graduação – têm sido objeto de avaliação por parte do poder público sob o pressuposto de que a avaliação pode “produzir” um ensino de melhor qualidade (Sousa, 2001, p.90).

As possíveis contribuições da avaliação na promoção da melhoria do ensino têm sido discutidas com base em dois tipos de argumentos:

- a. O primeiro considera a lógica interna dos processos de avaliação, examina seus princípios e procedimentos, procura explicitar os enfoques e critérios adotados, tomando-os como referência para a discussão das potencialidades e dos limites das práticas adotadas. As possibilidades de interferência da avaliação na melhoria da qualidade de ensino estariam presentes na própria lógica intrínseca ao seu delineamento (Oliveira, 2008);
- b. O segundo considera também a utilização dos resultados na condução das políticas educacionais, examinando a validade, a relevância e a oportunidade das iniciativas adotadas como consequência do conhecimento obtido mediante o processo de avaliação.

No Brasil, os debates a propósito das iniciativas do governo federal e de governos estaduais tendem a focalizar os princípios e os delineamentos dos programas realizados, procurando explicitar suas potencialidades ou deficiências e/ou riscos inerentes à sua implementação, explorando, por vezes, concepções de qualidade de ensino implícitas nos programas de diferentes administrações. São ainda escassas as investigações que se propõem a explorar o uso dos resultados nos processos decisórios pelas diversas instâncias do sistema educacional (Sousa, 1997), o que torna oportunas as pesquisas dessa natureza. Sua institucionalização, tanto em âmbito nacional quanto nas unidades federadas, tem implicações nas políticas públicas de educação envolvendo as instituições, o currículo, os profissionais e os estudantes, e evidencia uma redefinição do Estado em relação ao seu papel na área.

No Brasil, a partir de 1995, a implementação de processos avaliativos intensifica-se e a visão de melhoria da qualidade do ensino e do papel do Estado na educação, intrínseca à avaliação em larga escala, assume crescentemente maior importância no debate educacional. Nesse período, agências internacionais passam a estimular e influenciar de modo mais direto o delineamento das propostas de avaliação de sistemas, demonstrando claro interesse sobre a eficácia dos investimentos externos na educação. O Sistema de Avaliação da Educação Básica — Saeb —, de nível nacional, implementado a partir de 1990, passa, por exemplo, a ser parcialmente financiando pelo Banco Mundial e, conseqüentemente, sofre transformações de acordo com a mudança de seus agentes, influências de concepções, pesquisas e técnicas avaliativas internacionais (Bonamino, 2002). A consolidação do Saeb estimulou ainda as propostas de avaliação no âmbito de governos estaduais, iniciativas que passaram a ter centralidade nas políticas em curso.

Este artigo apresenta os resultados de pesquisa financiada pela Fundação de Amparo à Pesquisa do Estado de São Paulo — Fapesp —, conduzida entre 2005 e 2007, com o propósito de caracterizar e cotejar sistemas de avaliação implementados por cinco unidades federadas (Bahia, Ceará, Minas Gerais, Paraná e São Paulo), buscando compreender seus pressupostos e contornos, bem como analisar se e como os resultados produzidos têm informado a formulação e implementação de suas políticas educacionais.

A pesquisa baseou-se na compreensão de que os sistemas de avaliação precisam ser analisados para além da superfície vislumbrada por suas propostas e seus instrumentos de avaliação, com o intuito de verificar qual sua potencia-

lidade para colocar-se como marco de política educacional que efetivamente interfere nas escolas.

A necessidade de investigações dessa natureza é apontada por Barretto e Pinto (2001) e por Sousa (2002), diante da constatação a que chegam as autoras da existência de um número muito reduzido de estudos sobre os resultados ou impactos das avaliações. Barretto e Pinto, ao mencionarem os trabalhos sobre avaliação de sistemas, ponderam que existe:

...a predominância do discurso oficial nesta temática, principalmente na explicitação das características e finalidades de um sistema permanente de avaliação, capaz de contribuir para a melhoria da qualidade das escolas. [...] Os estudos mais sistemáticos realizados no País pelos próprios sistemas de aferição – Saeb, Paemg, Saresp – parecem encerrar-se em si mesmos, sem que haja um esforço maior de articulação com os demais estudos sobre a escola e o rendimento escolar, não se evidenciando uma reflexão acumulada. (Barretto, Pinto, 2001, p.61)

Há que se considerar ainda que esses sistemas de avaliação emergem sintomaticamente no momento em que se assiste a um questionamento do papel do Estado nas sociedades capitalistas, direcionado à sua capacidade de gerir diretamente os serviços a ele associados historicamente¹ (Sousa et al., 2000; Sousa, Oliveira, 2003). Mais do que um debate acadêmico, a apreciação da natureza e do papel do Estado tem desencadeado mudanças significativas nas políticas públicas, particularmente no processo de privatização da gestão de serviços e atividades estabelecidas na esfera pública. A discussão da avaliação de sistemas abrange um amplo leque de questões de natureza técnica e, para além delas, implicações de ordem política, pois que está intimamente vinculada às políticas públicas de educação.

Esses sistemas de avaliação, não constituindo elementos isolados das políticas educacionais na esfera nacional e internacional, têm conhecido forte desenvolvimento. Destaca-se, no âmbito da educação básica, inicialmente o

1. Essa afirmação pode vir a ser relativizada nos próximos anos, a depender das interpretações que se tornem hegemônicas sobre a recente crise do *subprime* com consequências globais, neste momento, ainda não perfeitamente claras.

Saeb², seguido do Exame Nacional do Ensino Médio — Enem — e do Exame Nacional de Cursos — ENC³. Essa tendência avaliativa ganha também espaço no plano mundial, seja em muitos países singularmente, seja mediante procedimentos que procuram fixar comparações entre países, como é o caso do Programa de Avaliação Internacional de Estudantes — Pisa⁴.

Patrícia Broadfoot ressalta que a matriz dominante relaciona-se fortemente à visão segundo a qual, pelos sistemas de avaliação, se difundiria um processo de gestão de sistema com forte influência da lógica empresarial.

Entretanto talvez, mais importante que esse crescente controle, seja a crescente associação da administração educacional em ambos os países (Inglaterra e França) com uma abordagem da administração empresarial. Essa abordagem tende a dissimular a natureza essencialmente política dos objetivos educacionais por meio de uma ideologia de racionalidade científica. Nesse caso, julgamentos de valor aparecem como meras decisões administrativas, tendo como objetivo maximizar a eficiência. É provável que o efetivo controle educacional implique a existência de uma ordem social pronta a concorrer para o alcance desses objetivos educacionais. As maneiras pelas quais os procedimentos de avaliação contribuem para que isso aconteça são, em última instância, talvez mais impor-

2. Instituído em 1990 pelo Ministério da Educação e do Desporto, efetua avaliações bienais por amostragem em todas as unidades da federação. A partir de 2005, passou a ser constituído por duas iniciativas: a Avaliação Nacional da Educação Básica — Aneb — e a Avaliação Nacional do Rendimento Escolar — Anresc. A primeira é realizada por amostragem das redes de ensino em cada unidade da federação e focaliza as gestões dos sistemas educacionais — mantém características do Saeb. A segunda, conhecida como Prova Brasil, tem foco em cada unidade escolar.
3. O Enem e o ENC (Provão) teriam como objetivo fornecer indicadores, apesar das diferenças em seus procedimentos e métodos, sobre o ensino médio e superior respectivamente, cobrindo um amplo espectro da educação nacional. Excluindo-se a educação infantil — que ainda guarda muita dependência de uma concepção restritiva de assistência social — e os cursos de pós-graduação, que merecem outro tipo de avaliação pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior — Capes.
4. O Pisa avalia estudantes com 15 anos de idade, independentemente do nível de escolarização; é aplicado pela Organização para Cooperação e Desenvolvimento Econômico — OCDE —, entre os países membros, porém possui uma política de ampliação de estudantes de outras nações, o que explica o envolvimento de alunos brasileiros nas provas realizadas no final de 2000 (OECD, 2001, 2001a).

tantes do que o seu papel na imposição dessas metas. De fato, as exigências dos testes de larga escala são de tal ordem que envolvem um considerável acordo acerca dos objetivos educacionais. Pode bem acontecer que a tecnologia dos testes vá além da identificação inicial dos objetivos educacionais, reforçando a tendência, bastante criticada, de reduzir os objetivos curriculares ao que pode ser medido e imprimindo ao processo um caráter próximo ao ideal tecnocrático de gerenciamento. (Broadfoot, 1996, p.123-124, tradução nossa)

Nessa perspectiva, a utilização de resultados de processos avaliativos em larga escala depara-se, no país, com uma intensa disputa ideológica. De um lado, a expectativa de uma gestão de sistema ancorada em uma racionalidade instrumental, com perspectiva de gerar maior eficiência. De outro, a ideia da autonomia dos processos educativos no âmbito da escola, enfatizando processos de autoavaliação. Dessa forma, a questão do uso dos resultados adquire uma nova dimensão. Em que medida a implantação dos sistemas de avaliação tem sido capaz de induzir autorreflexão no âmbito do sistema como um todo e, em particular, na escola? Ou, ainda, em que medida tem sido capaz de provocar alteração na lógica de formulação e implementação de políticas educacionais e nas práticas escolares?

As contribuições presentes na literatura têm incidido na análise da ação reguladora do Estado e de seus efeitos, nos moldes resumidos por Madaus:

1. O poder dos testes e exames para afetar indivíduos, instituições, currículos é um fenômeno perceptivo: se os estudantes, professores ou administradores acreditam que os resultados de um exame são importantes, importa pouco se isso é realmente verdadeiro ou falso – o efeito é produzido pelo que os indivíduos percebem ser.
2. Quanto mais indicadores sociais quantitativos são usados para tomar decisões sociais, mais provavelmente se distorcerão tais processos tentando monitorá-los.
3. Se importantes decisões são supostamente relacionadas aos resultados dos testes, então os professores ensinarão para o teste.
4. Em cada ambiente em que se operam testes, desenvolve-se uma tradição baseada em exames passados, o que eventualmente define “de fato” o currículo.

5. Os professores dirigem particular atenção à forma das questões dos testes (por exemplo, resposta curta, ensaio, múltipla escolha) e ajustam o que ensinam de acordo com ela.
6. Quando os resultados dos testes são o único ou mesmo o árbitro parcial do futuro educacional ou das escolhas de vida, a sociedade tende a tratar o resultado dos testes como o principal objetivo da escolarização, em vez de um indicador útil do desempenho, ainda que falível.
7. Os testes transferem controle do currículo para a agência que controla o exame. (1988, p.94, tradução nossa)

Embora se reconheçam tais limitações, o dimensionamento da qualidade de ensino tem sido apresentado como carro-chefe de várias iniciativas de avaliação de sistema que se apoiam sobremaneira no desempenho dos alunos, quer censitariamente, quer por procedimentos de amostragem da população referida. Por sua vez, não excluindo essa primeira perspectiva, argumenta-se que essas práticas avaliativas seriam indutoras de níveis crescentes de qualidade de ensino. Esses pontos de partida revelam-se, no entanto, extremamente controvertidos pelas dificuldades de definir a qualidade de ensino de forma operacional, inclusive por conta de seus condicionantes sociais e históricos;

...é necessário reiterar que avaliação da qualidade da educação não se limita apenas à verificação do rendimento escolar, que é um momento na caracterização dessa qualidade. O desempenho dos estudantes em pesquisas da qualidade da educação é melhor compreendido e interpretado quando se levantam informações sobre o tipo de ensino que recebem, os procedimentos que vivenciam em sala de aula e no colégio, e ainda sobre as características ambientais da família que determinam o seu comportamento. (Vianna, 1990, p.99)

Essas preocupações ganham relevância quando se constata que os principais processos avaliativos em larga escala, o Saeb e a Prova Brasil, têm o rendimento do aluno como ponto nuclear para as apreciações da educação básica (Sousa, 2001, p.88). Tal constatação não implica um julgamento acabado desses sistemas; trata-se de apontar limitações desses procedimentos, que inclusive podem infundir um “cientificismo” que não lastrearia adequadamente

as políticas que daí derivam (Perrenoud, 1998). Ademais, tais práticas não favoreceriam a consolidação de “um verdadeiro sistema de avaliação”, que, para constituir-se, demandaria uma multiplicidade de iniciativas que têm:

...mais dificuldade de ser entendid[as]o pela sociedade e talvez por isso, tenha[m] menos apelo para quem está em posição de decisão. A iniciativa de avaliação única e censitária desperta maior impacto e emoção. A este respeito, compare-se o Provão e o Saeb. Mas o problema é que nem tudo que provoca impacto e emoção entrega o que promete. (Franco, 2004, p.61-62)

Na literatura investigada há uma concentração de trabalhos que tratam de processos avaliativos em curso nos Estados Unidos nos quais estas preocupações são fortemente reiteradas (Heubert, Hauser, 1999; Kohn, 2000; Orfield, Kornhaber, 2001; Levin, 2001; Mcneil, 2000; Wilde, 2002, entre outros). Também se encontram tais posicionamentos quando se abordam procedimentos em implementação na Europa, como ilustra a manifestação de Roggero, demarcando que:

Considerando o sistema educativo ou uma organização educativa como um sistema complexo, isto é, capaz de proceder a uma autoecorreorganização, tomamos consciência de que ele depende de sua história (re), dos relacionamentos com seu meio ambiente (eco) e de sua identidade interna (auto). Toda ação de avaliação deveria, desde já, levar em conta essas dimensões essenciais que não saberíamos avaliar unicamente por indicadores de performance. (Roggero, 2002, p.43)

As referências evidenciam que ao tempo em que se assiste à ampliação de iniciativas de avaliação em todo o mundo, emergem novas questões de ordem técnica e política que demandam equacionamento no âmbito de políticas educacionais e no de pesquisas que possam revelar tendências, subsidiar críticas e elucidar caminhos.

Um indicador da generalização dos processos avaliativos é a apresentação de um número da *Revista de Educación*, publicado em 2000, pelo órgão de comunicação do Ministério da Educação da Espanha, dedicado às avaliações de sistema:

Una de las preocupaciones de las autoridades educativas de cualquier país es la de conocer la situación de su sistema vigente en la práctica, y más allá de los diseños o de las intenciones legislativas. Por otro lado, la sociedad, en general, cada vez se vuelve más exigente acerca de cómo funciona y qué logros se alcanzan en la marcha de su respectivo sistema educativo. De ahí que los gobiernos *institucionalicen* e *instrumentalicen* mecanismos y *atalayas* desde los que reciben información sobre sus sistemas educativos y, a la vez, ésta la puedan transferir a la ciudadanía. (p.7, grifos nossos)

Se a consciência da importância da avaliação dos sistemas escolares, e particularmente das escolas, generalizou-se, longe estamos, contudo, de algum acordo acerca de “para que avaliar” e de “como avaliar”.

Evidentemente, há uma articulação entre os motivos que levam à avaliação e à forma como ela é realizada (Oliveira, Sousa, 1999). A adoção de determinados processos avaliativos sinaliza que modificações se quer induzir. Mesmo que não se trate de “prêmios” e /ou “punições” para os bem ou mal avaliados, o simples fato de realizar uma avaliação significa que foi definida uma situação desejável.

Ao realizar determinado processo avaliativo, espera-se, explicitamente, verificar quão distante se está da situação desejável e, a partir daí, definir elementos para modificar a situação em direção ao padrão desejado. Além disso, a avaliação pode, também, possibilitar a emergência de propostas de redirecionamento ou transformação da situação avaliada. Não é possível pensar a avaliação dissociada da ideia de modificação. Para isso, talvez, a mais importante questão seja como criar um envolvimento tal que impulsiona os membros da instituição a se engajarem no processo de transformação.

Os critérios ou padrões podem, facilmente, simplificar problemas complexos pela utilização de indicadores com potencial de provocar mudanças, mas não necessariamente induzir a situação desejada. A observação indica a necessidade de uma reflexão mais cuidadosa sobre os usos (e abusos!) do Índice de Desenvolvimento da Educação Básica – Ideb (Oliveira, 2007).

No caso dos sistemas estaduais de avaliação, cabe indagar acerca de suas finalidades, seus procedimentos e consequências no que diz respeito à melhoria da qualidade do ensino e, até mesmo, saber qual o sentido atribuído ao compromisso com a qualidade do ensino anunciado nas políticas educacionais formuladas no âmbito dos estados. Pois qualidade

...não é “algo dado”, não existe “em si”, remetendo à questão axiológica, ou seja, dos valores de quem produz a análise de qualidade. A emergência de critérios de avaliação não se dá de modo dissociado das posições, crenças, visão de mundo e práticas sociais de quem os concebe. É um conceito que nasce da perspectiva filosófica, social, política de quem faz o julgamento e dela é expressão. Portanto, os enfoques e critérios assumidos em um processo avaliativo revelam as opções axiológicas dos que dele participam. (Sousa, 1997, p.26)

Assim, ao se fixar no uso dos resultados dos sistemas de avaliação, este estudo investiga não só o que tem sido feito, mas inquirir acerca de suas eventuais consequências.

METODOLOGIA

Na primeira fase da pesquisa, foram consolidadas informações acerca dos sistemas de avaliação de cada estado (Bahia, Ceará, Minas Gerais, Paraná e São Paulo), disponíveis nos respectivos sítios das Secretarias de Educação, bem como em documentos produzidos por essas instituições sobre seus sistemas de avaliação. Em alguns casos, foram ainda consideradas informações fornecidas por profissionais envolvidos com a gestão estadual em resposta a consultas realizadas por correio eletrônico.

Visando compreender como se inserem as propostas de avaliação nas políticas educacionais formuladas pelos estados, foram colhidas informações sobre elas, buscando apreender as diretrizes e prioridades estabelecidas pelos governos e o lugar reservado à avaliação. Para tanto, consultaram-se documentos das secretarias e ainda produções acadêmicas, particularmente dissertações e teses, que tinham como objeto de investigação essas políticas.

O período delimitado para a coleta de informações variou de um para outro estado; tomou-se como referência para início do levantamento documental o ano em que foi implantado o sistema de avaliação no respectivo estado, abrangendo, sempre que possível, dados até 2006.

Essa primeira etapa da pesquisa resultou em uma descrição preliminar dos sistemas de avaliação, posteriormente complementada com informações coletadas quando da realização da pesquisa de campo, que consistiu na segunda fase da investigação.

O levantamento de informações foi feito por meio de entrevistas orientadas por um roteiro semiestruturado. As entrevistas tiveram como propósito principal conhecer como têm sido utilizados os resultados da avaliação e compreender, na perspectiva dos gestores e técnicos, a importância atribuída à avaliação para a formulação e implementação de políticas educacionais. Também, por meio das entrevistas, buscou-se a complementação e/ou o esclarecimento de dados acerca das características dos delineamentos das propostas estaduais, quanto aos seus princípios, abrangência e procedimentos.

As informações foram sistematizadas e analisadas quanto às evidências acerca do sentido e finalidades que a avaliação vem adquirindo nas políticas educacionais. A partir do mapeamento de cada estado, tendo como referência o conjunto de discursos de seus gestores e técnicos, procedeu-se a uma análise comparativa entre os estados, que permitiu indicar tendências e elencar perspectivas que se podem vislumbrar com base nos relatos das práticas atuais de avaliação dos sistemas educacionais.

SISTEMAS DE AVALIAÇÃO: PECULIARIDADES E RUMOS

É no início dos anos 1990 que se registra a implantação dos sistemas de avaliação nos estados em estudo, sendo que dentre estes o último organizado foi o da Bahia, em 1999. Minas Gerais foi um dos primeiros estados do país a instituir sistemática própria de avaliação, em 1992. No entanto, o sistema com as características hoje vigentes foi delineado a partir de 2000. O Estado do Ceará tem iniciativas de avaliação já no início dos anos 1980.

Os objetivos da avaliação, declarados nos documentos oficiais, tendem a afirmar como expectativa que a avaliação venha a subsidiar os diferentes níveis do sistema na tomada de decisões com vistas à melhoria da qualidade do ensino. Desse modo, espera-se que os resultados da avaliação venham a ser apropriados pelos gestores e equipes centrais e regionais das Secretarias de Educação bem como pelas escolas, havendo menção ainda, por alguns estados, quanto à expectativa de que a comunidade escolar como um todo – envolvendo alunos e pais – venha a se inteirar dos resultados obtidos pelas escolas. Ou seja, há referências de que a avaliação deve iluminar e trazer consequências para a formulação e/ou reformulação das políticas educacionais em relação às suas diretrizes, prioridades e metas, focalizando questões relativas à gestão,

ao currículo e à infraestrutura, assim como às propostas e práticas das escolas, particularmente em relação a aspectos que concernem ao desenvolvimento do currículo. Especificamente o Paraná, em 1995, estabelece como objetivos de seu sistema “conhecer o desempenho das escolas quanto à sua organização, gestão e articulação com a comunidade”. O Ceará visa, em 2003, “implementar o processo de formação dos Conselhos Escolares, para que esses possam desenvolver, de forma competente, sua função avaliativa”.

O principal indicador dos sistemas de avaliação para aquilatar a qualidade de ensino tem sido a medida da proficiência dos alunos, obtida por instrumentos de testagem. Há menção à coleta de informações acerca das escolas e dos alunos. Essas informações, no entanto, não têm sido exploradas de modo articulado ao rendimento apresentado pelos alunos nas provas. Ou seja, na documentação coletada não há registros de interpretação dos resultados de modo contextualizado, procurando identificar os fatores associados ao nível de proficiência dos alunos.

Observou-se, ainda, que é usual a referência à intenção de se criar nos sistemas de ensino uma “cultura de avaliação”, conceito que, no entanto, não é explicitado ou discutido de modo a permitir uma apreensão clara do seu significado. Mesmo a análise dos procedimentos utilizados na implementação da avaliação não fornece pistas que apoiem inferências nesse sentido.

Ainda no que tange a objetivos, registre-se que o Ceará, em 1992, estabeleceu como meta “criar índice de qualidade de cada escola” e o Paraná dedicou-se ao “estabelecimento de metas para cada escola”. Observou-se que a perspectiva de estabelecer metas com base nos resultados da avaliação, embora não conste dos documentos de outros estados, está presente nos depoimentos coletados nas entrevistas; há uma tendência de associar aos resultados, compromissos específicos a serem assumidos pela escola e também pelas outras instâncias do sistema. Essa configuração ficou mais claramente explicitada no modo pelo qual têm sido conduzidas as ações em Minas Gerais.

A aplicação das provas de rendimento escolar nos estados pesquisados tem sido censitária. Entretanto, há uma exceção em São Paulo, em 2002; apesar de a aplicação ter caráter censitário em relação às escolas, foi amostral em relação aos alunos. Além de abranger todas as escolas estaduais, os sistemas de avaliação têm contemplado escolas de redes municipais e particulares que desejem participar das avaliações. Em Minas Gerais e Ceará, há participação de

redes municipais e em São Paulo, Bahia e Paraná também as escolas particulares são convidadas a participar.

Em todos os estados, houve delimitação quanto às séries e disciplinas focalizadas nas avaliações. A aplicação das provas tem seguido uma periodicidade anual. Nos estados do Paraná e Bahia, nos anos recentes, essas passaram a ser bianuais. São avaliados, em geral, os alunos da 4ª série, da 8ª série e da 3ª série do ensino médio nos componentes curriculares Português e Matemática. Além destes, no Paraná foram realizadas avaliações de Ciências, Física, Química e Biologia e, em São Paulo, nas últimas avaliações examinadas, houve provas de Leitura, Escrita e Matemática.

Sobre a construção das provas, os documentos indicam, com diferentes níveis de detalhamento, os referenciais que são utilizados, os passos seguidos para a elaboração dos instrumentos e os responsáveis pelas atividades previstas no fluxo de trabalho estabelecido pela Secretaria de Educação. São aqui destacados alguns elementos que permitem caracterizar tendências das iniciativas no âmbito de cada estado em análise.

A Secretaria de Educação da Bahia, contando com a participação de professores de escolas públicas e privadas, definiu as matrizes para elaboração das provas tendo como referência os Parâmetros Curriculares Nacionais, os livros e as práticas didáticas consideradas usuais no estado pelos professores participantes. Essas matrizes foram validadas por especialistas, que avaliaram o conjunto de descritores propostos quanto ao seu valor e pertinência. A escolha dos itens para elaboração das provas coube à equipe do projeto de avaliação da Secretaria de Educação.

No Ceará e em São Paulo, os itens das provas eram elaborados com base nos referenciais curriculares do respectivo estado e nos Parâmetros Curriculares Nacionais, contando com a participação de professores e técnicos da rede. Contou-se com especialistas das áreas de conhecimento para revisão dos itens.

O Paraná e Minas Gerais estruturaram a matriz de descritores com base no currículo vigente nas escolas da rede, nos Parâmetros Curriculares Nacionais e, ainda, procuraram compatibilizá-la aos descritores adotados no Saeb. Também nesses estados, os professores da rede e técnicos da Secretaria de Educação participaram da elaboração da matriz, contando com consultores externos.

As informações sobre o processo de elaboração dos itens das provas permitem supor que há tendência de que esses sejam semelhantes, considerando que utilizam para sua elaboração referências comuns, quais sejam: os Parâmetros Curriculares Nacionais e o Saeb. Caberia verificar se há traços específicos nas propostas curriculares estaduais que as diferenciam significativamente das orientações nacionais, capazes de justificar a organização de descritores específicos pelos estados ou, mesmo, em que medida a participação de professores da rede, sob o pressuposto de articular os descritores às propostas vigentes nas escolas, tem resultado em expectativas de proficiência dos alunos diferentes das previstas no Saeb.

Além dos testes, foram utilizados questionários para coleta de informações acerca dos alunos. Alguns estados fizeram também uso de instrumentos de levantamento de informações sobre a escola e seus profissionais.

A julgar pelas informações disponibilizadas, os sistemas tomam o Saeb como referência e procuram articular suas escalas de proficiência e matriz de referência. Alguns dos programas utilizam como ferramenta analítica a Teoria da Resposta ao Item – TRI –, com três parâmetros. Essa característica é explícita para o caso do Ceará em 1998 e 2002, bem como a utilização da Teoria Clássica da Medida. A ferramenta também só é mencionada explicitamente por São Paulo para os anos de 2000 e 2001. Em 2000 este estado menciona igualmente a utilização de Teoria da Resposta ao Item, mas não acrescenta detalhes. No caso da Bahia, não dispomos dessa informação para 1999 e 2000, e a escala originalmente construída em 2001 não corresponde diretamente à escala do Saeb, tendo sido construída uma escala de equivalência, com base em um “julgamento social”, com assessoria do American Institute for Research – AIR. Segundo o material coletado, “seu intuito era permitir associações do tipo: um *score* médio ‘x’ na avaliação de desempenho corresponderia a um desempenho médio localizado dentro do nível e da escala Saeb” (Bahia, 2003, p.71). Em Minas Gerais, em 2000, 2001 e 2002, foi utilizada como ferramenta a TRI com três parâmetros. No Paraná é apenas mencionado o procedimento de “análise estatística”, sem mais detalhes.

Os estados previam a divulgação dos resultados com vistas à sua utilização por diferentes instâncias e participantes das redes de ensino. No entanto, no desenvolvimento da pesquisa, não se obteve acesso a registros atualizados, o que faz supor que a publicização dos resultados não é feita

com a periodicidade prevista, ou seja, com a mesma frequência com que se aplicam os testes.

No Paraná, era prevista a divulgação dos resultados obtidos pelos alunos para diretores de escolas e secretários municipais de educação. Havia também o que denominam relatório de “análise pedagógica dos resultados da avaliação”, que apresentava considerações e indicações aos professores, às equipes pedagógicas e aos presidentes das associações de pais e mestres, com base nos resultados obtidos pelos alunos. Era ainda elaborado o “boletim da escola”, que trazia uma descrição simplificada do seu desempenho, divulgado aos pais dos alunos.

São Paulo previa a divulgação dos relatórios com os resultados obtidos pelos alunos no sítio da Secretaria de Educação na *internet* e em relatórios distribuídos às regionais, responsáveis pela sua discussão junto às escolas a elas vinculadas. Também a Bahia e Minas Gerais previam a divulgação de relatórios da avaliação, distribuídos a todas as escolas da rede, com as informações acerca dos resultados gerais do estado, das regiões, dos municípios e de cada unidade escolar, seguidas de análise pedagógica dos itens da prova. O mesmo ocorria no Ceará, que fez ainda referência à realização de seminários regionais para discussão dos resultados.

Quanto às consequências esperadas, decorrentes da utilização dos resultados da avaliação, o mais enfatizado pelos estados foi a expectativa de que estes viessem a subsidiar as iniciativas de formação continuada dos professores com vistas ao aprimoramento das práticas pedagógicas. Evidenciou-se, nos depoimentos coletados por meio das entrevistas, que os resultados da avaliação tendiam a ser considerados no planejamento das ações de formação continuada, privilegiando conteúdos em que os alunos mostraram menor domínio.

No caso do Ceará, associava-se aos resultados uma premiação. A partir de 2002, foi instituído o prêmio educacional Escola do Novo Milênio: Educação Básica de Qualidade no Ceará, destinado às escolas e aos alunos mais bem colocados no Sistema Permanente de Avaliação da Educação Básica do Ceará – Spaece.

Em São Paulo, em 2001, os resultados do Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo – Saresp – foram utilizados pela Secretaria de Educação para determinar a possibilidade de os alunos darem

continuidade à trajetória escolar ou fazerem a recuperação, sendo prevista sua reprovação em caso de aproveitamento insatisfatório nessa atividade.

Em linhas gerais, observa-se, pelos registros consultados, grande similaridade nos delineamentos adotados pelos sistemas de avaliação. Há que ressaltar, no entanto, o movimento realizado pela Secretaria de Educação do Ceará, com vistas a integrar ao Spaece – que se caracteriza por ser uma avaliação em larga escala –, a Avaliação Institucional, que trata de implementar a autoavaliação da escola, em todos os seus aspectos, e é realizada por todos os segmentos da comunidade escolar. Assim, a partir de 2000, o Spaece contemplou tanto a avaliação em larga escala quanto a avaliação institucional que, até então, vinha sendo realizada como atividade paralela ao Spaece. Considerando as características dos delineamentos adotados nos estados em análise, a configuração que assume a sistemática de avaliação no Ceará parece a mais inovadora no sentido de articular diversas vertentes, dimensões e fluxos ao processo avaliativo.

No Paraná ocorreu um movimento diferente do Ceará, o que se constatou quando da visita *in loco*. A decisão foi deixar de implementar a avaliação em larga escala passando à proposição apenas de uma sistemática de autoavaliação institucional, que previa análise das ações implementadas por todos os níveis do sistema de ensino.

Também o delineamento do Projeto de Avaliação Externa do Estado da Bahia, de acordo com informações divulgadas pela Secretaria da Educação – SEC – na *internet* (<http://www.aval.ufba.br/projeto.asp>), apresentava uma especificidade, integrando as ações do Educar para Vencer – programa estratégico, assumido pelo governo do estado –, composto por dois sistemas de avaliação em larga escala: a Avaliação de Desempenho e a Avaliação de Aprendizagem, que cumpriam propósitos específicos. No entanto, conforme informação das entrevistas, essas dimensões não se concretizaram de modo articulado.

O estudo dos documentos relativos aos sistemas de avaliação dos estados permite supor que a implementação da avaliação dos sistemas de ensino é uma tendência que se tem consolidado no cenário educacional brasileiro, envolvendo inicialmente as redes públicas estaduais e se estendendo, gradativamente, para as redes públicas municipais e, até mesmo, em caráter voluntário, para escolas particulares. No entanto, no contato direto com os gestores dos estados, foi possível notar que, se por um lado, a avaliação integra suas iniciativas,

por outro, se está evidenciando a necessidade de repensar o seu escopo, visando torná-la mais útil à gestão e à promoção da qualidade da educação. Também, há uma sensação no âmbito dos estados de duplicação de iniciativas, considerando que o governo federal, a partir de 2005, passou a realizar uma avaliação das redes de ensino não mais de caráter amostral, a Prova Brasil.

As propostas de avaliação dos estados, cotejadas nesta investigação, em seus desenhos originais apresentavam características comuns entre si. Entretanto, o que se constatou pelas entrevistas foi um movimento que começa a ocorrer no âmbito de Secretarias Estaduais de buscar abordagens e perspectivas avaliativas que se mostrem mais consequentes no sentido de subsidiar decisões e ações e induzir a produção dos resultados esperados.

Nessa direção, registram-se iniciativas que apresentam especificidades. Uma delas refere-se à implantação da autoavaliação institucional, com base na compreensão de que tal sistemática teria poder, entre outras atribuições, de gerar envolvimento e compromisso de todos com a melhoria da qualidade da educação. A experiência do Ceará, que não é recente e vem sendo mantida no bojo da avaliação de larga escala, e a do Paraná, implantada em 2005, são exemplos.

Outro tipo de iniciativa, que caminha em direção diferente, é a de intensificar a relação entre resultado da avaliação e algum tipo de consequência. Isso pode ocorrer desde a elaboração de um plano específico de ação que será submetido à negociação, acompanhamento e controle, até a premiação por bons resultados. Em Minas Gerais se evidenciou com maior clareza o movimento de implantar uma dinâmica de trabalho com a rede que intensificou a relação entre avaliação, planejamento e controle. Quanto à premiação associada à avaliação, o Ceará foi o estado em que essa iniciativa foi mais evidenciada, no entanto, é uma perspectiva que apareceu subjacente à manifestação de outras Secretarias. Há propostas que advogam sejam premiados docentes e escolas em razão do desempenho.

São muitas as expectativas declaradas pelos gestores acerca das contribuições esperadas das avaliações, que vão desde consequências na formulação e execução das políticas educacionais até influências na gestão e no currículo das escolas. É possível afirmar que se está vivendo um momento propício para a constituição de novas dinâmicas de avaliação.

Ao que parece, o que se busca são alternativas de avaliação capazes de contribuir para a formulação de políticas e gestão do sistema. Há que reconhecer que está havendo acúmulo de grande quantidade de informações como resultado dos processos em curso, que chegam a constituir sistemas de informação sem precedentes em nossa educação. Esses sistemas se encontram, ainda, marcadamente focados no desempenho dos estudantes. Informações sobre variáveis de contexto escolar são coletadas, mas parecem pouco exploradas em seu potencial explicativo acerca de eventuais diferenças de rendimento escolar identificadas no interior do sistema.

Não há evidências, portanto, de que as informações estejam sendo utilizadas de maneira consistente em um processo de avaliação sistêmico, o que se compreende pela relativa juventude do conjunto de iniciativas aqui analisadas.

Por fim, cabe registrar o caráter de referência que o Saeb adquiriu com os gestores estaduais, quer seja como modelo do ponto de vista de desenho, quer seja como parâmetro para a proposição de alternativas.

SISTEMAS DE AVALIAÇÃO: TENDÊNCIAS E PERSPECTIVAS

A investigação realizada permitiu identificar características presentes na condução da avaliação de sistemas educacionais no Brasil. Dentre as revelações trazidas com o estudo, destacam-se, a seguir, algumas delas, que expressam de modo mais evidente desafios a serem enfrentados não só pelos gestores dos sistemas, mas também por pesquisadores que vêm se dedicando ao campo da avaliação educacional, particularmente à avaliação de políticas educacionais.

CONFIABILIDADE DOS RESULTADOS

Nos últimos anos, melhorou muito a capacidade técnica disponível no Brasil para realização de processos de avaliação em larga escala, particularmente no que diz respeito à teoria da medida. Exemplo disso é a generalização da utilização da Teoria da Resposta ao Item, sendo utilizada tanto com um parâmetro quanto com dois ou três, da capacidade de elaboração e testagem de itens, bem como a consolidação de algumas instituições com capacidade de realizar tais análises, ainda que seja correta a ponderação do representante do

Banco Mundial, entrevistado nesta pesquisa, que afirma ainda não haver um “mercado” nesta área no Brasil.

Permanecem, entretanto, alguns problemas que podem comprometer a confiabilidade dos resultados. Um deles diz respeito a práticas relacionadas ao processo de aplicação dos instrumentos, pois ainda é comum a utilização de professores da própria escola para a sua aplicação, ao mesmo tempo em que os resultados da avaliação impactariam salários, bônus ou recursos repassados para a escola. Outro problema diz respeito ao interesse político de secretários e governantes que veriam num resultado positivo da avaliação uma chancela às suas iniciativas de política educacional.

DESCONTINUIDADE DAS PROPOSTAS

Dos cinco estados estudados, apenas um deles, o Ceará, tem dado continuidade ao delineamento inicial previsto para a avaliação do sistema, caracterizando-se as alterações introduzidas como busca de aprimoramento da sistemática de avaliação implementada. Nos demais, seja por mudança na administração, seja em decorrência da iniciativa do governo federal de avaliar todas as escolas, via Prova Brasil, não tem havido continuidade na sistemática inicialmente concebida.

Desses elementos, o referente à descontinuidade por mudança de gestão é consequência de característica geral da administração pública brasileira. Ao mudar o governo e, às vezes, apenas o Secretário de Educação, alteram-se as iniciativas anteriores. No caso dos sistemas de avaliação, isso tem ocorrido pela simples supressão do sistema anteriormente implementado, ou pela alteração de seus objetivos, procedimentos ou usos, corrompendo, assim, sua integridade e inviabilizando o recurso à análise de séries históricas.

Há que registrar, também, que a construção dos sistemas estaduais de avaliação deveu-se em grande parte à indução do governo federal e à sua iniciativa de criar o Saeb, de caráter amostral. Em muitos casos, ainda, financiamentos do Banco Mundial previam em suas cláusulas o desenvolvimento de sistemas de avaliação por parte dos estados. Nessa perspectiva, tecnicamente, vários dos sistemas estaduais foram concebidos em complementariedade ao Saeb, o que significa que suas características (matriz curricular, metodologia de

análise, banco de itens etc.) seriam originalmente compatíveis ou assim foram se constituindo no decorrer do tempo⁵.

Quando o governo federal, unilateralmente, toma a iniciativa de realizar uma testagem censitária, a Prova Brasil, apresentam-se dois tipos de problemas. De um lado, os estados passam a questionar-se acerca da validade de manutenção de sistemas próprios, com seus altos custos e complexidade logística, para realizar a mesma função que o governo federal já realiza. De outro, se estabelece uma frequência muito alta de aplicação de provas, o que inviabiliza, até mesmo, o tempo necessário para que seus resultados sejam absorvidos, analisados e transformados em ações por parte das redes de ensino.

É possível inferir que alguns dos gestores dos sistemas estudados já teriam abandonado a implementação de seus próprios sistemas se tivessem confiança na perenidade da Prova Brasil. De toda forma, o arranjo estabelecido padece de um problema estrutural. A única finalidade de um sistema censitário é sua utilização na gestão do sistema de ensino com implicações para cada escola, ou até mesmo para professores e alunos, uma vez que a coleta de informações é satisfatoriamente respondida por um sistema amostral, de custo significativamente mais baixo. Na medida em que os estados não pactuaram com o governo federal a utilização da Prova Brasil para fins de gestão do sistema, ocorre um problema de equacionamento quanto ao seu uso, além da já mencionada superposição de avaliações⁶.

Caso haja maior permanência nas iniciativas de forma a criar-se uma cultura e procedimentos administrativos e técnicos mais estáveis, e caso a avaliação seja assumida como função de Estado e não de governo, a utilização ampla de suas potencialidades pode ser alcançada.

5. Observou-se intercâmbio de experiências entre os sistemas estudados, o que significa dizer que algumas iniciativas foram tomadas em decorrência de crítica a experiências não bem-sucedidas em outro sistema ou apropriadas devido aos seus bons resultados. Exemplo disso é a utilização, por parte de Minas Gerais, da experiência de certificação de diretores desenvolvida na Bahia para formatar sua iniciativa de certificação.

6. Há que se considerar, no entanto, que os resultados da Prova Brasil passaram a integrar o Ideb que, em conjunto com o Censo Escolar, relativos à aprovação, constituem referência para a definição de metas a serem alcançadas pelas redes públicas de ensino até 2021.

ÊNFASE NA RESPONSABILIZAÇÃO DAS ESCOLAS

Na formulação dos objetivos dos sistemas, menciona-se a expectativa de que os resultados do sistema venham a subsidiar a tomada de decisões por parte de todas as instâncias da rede de ensino, dos gestores e equipes centrais, regionais e escolas, havendo referência, em alguns casos, até mesmo, à intenção de que a comunidade escolar como um todo venha a se inteirar dos resultados da avaliação. Ao buscar exemplos concretos de uso desses resultados, verifica-se que as iniciativas tomadas indicam majoritariamente a escola como grande usuária, tendendo-se a responsabilizá-la pelos resultados.

Os próprios gestores, no entanto, responsáveis por iniciativas nessa direção, reconhecem que as escolas têm dificuldade até mesmo de ler e compreender os resultados produzidos pelo sistema de avaliação. Nesse sentido, ocorrem, inclusive, esforços da direção para traduzi-los em uma linguagem compreensível. Isso inclui seminários para explicitação dos significados dos resultados, e a produção de materiais sintéticos a serem remetidos às escolas.

Não faz parte, no entanto, da cultura e da dinâmica da organização do trabalho escolar pautar seu planejamento em resultados de avaliação. Desse modo, as iniciativas referidas ainda encontram pouco eco na dinâmica da escola.

Na ótica dos seus profissionais, ao que parece, a avaliação entendida como significativa é a avaliação do aprendizado do aluno, que ocorre no âmbito de competência de cada professor. As tentativas, como a feita por São Paulo, de utilizar os resultados do sistema de avaliação como elemento definidor da aprovação do aluno, foram entendidas pelas escolas como uma perda de autonomia e de poder.

Assim sendo, explicita-se uma diferença de compreensão entre o que os gestores tentam fazer com os resultados dos sistemas de avaliação, responsabilizando a escola pelo seu uso, e o que esta faz de fato, praticamente, desconhecendo-os.

O uso dos resultados dos sistemas de avaliação por parte dos gestores é escasso ou inexistente. Nesse nível, observa-se a tendência de que tais resultados sejam compreendidos como apenas um indicador (a ser justaposto a outros, tais como evasão, repetência etc.), não informando políticas específicas. O gerenciamento do sistema apoia-se nas estruturas burocráticas e não se orienta pelos resultados de desempenho escolar. O investimento maior – humano e

em recursos financeiros – é feito na direção da produção de informações e não em sistemáticas que estimulem e apoiem o uso dos resultados. A noção de *accountability* – tanto em seu sentido de prestação de contas à sociedade quanto no de responsabilização – ainda é pouco explorada. Parece que até mesmo os pais de alunos pouco interagem com a avaliação.

TERCEIRIZAÇÃO DAS ATIVIDADES

As Secretarias de Educação estudadas mantêm, em diferentes pontos de seu organograma, equipes responsáveis pela avaliação de seus sistemas. Entretanto, dos cinco estados pesquisados, dois deles, Ceará e São Paulo, contam com equipes consolidadas, com capacidade técnica de interpretação dos resultados e com clareza de seus possíveis usos. Integram essas duas equipes técnicos incorporados há tempos ao sistema de ensino e que têm garantido uma compreensão institucional de seus sistemas de avaliação, mesmo que estes tenham se transformado ao longo do tempo por decisões de natureza política. Em Minas estabeleceu-se uma dinâmica tal que, no próprio núcleo dirigente da Secretaria, incorporam-se conhecimentos técnicos acerca da avaliação e se espera que esta venha a ser parte importante do eixo principal da política educacional em implementação. Nas demais secretarias, o problema da descontinuidade de pessoal e de demanda na equipe de avaliação manifesta-se, a ponto de, em um dos estados pesquisados, nenhum membro da equipe de avaliação ser capaz, ao menos, de abrir o banco de dados disponibilizado pela empresa contratada para a elaboração, aplicação e análise das provas.

Independentemente da configuração específica das equipes de avaliação, a complexidade e a logística requeridas para a implementação de um processo avaliativo têm levado à contratação de instituições e/ou empresas para a aplicação dos testes, o que configura uma série de limitações. De um lado, as equipes de avaliação não se apropriam completamente do desenho do sistema de avaliação, ficando na dependência dos tempos da instituição contratada e, mais do que tudo isso, têm de confiar nos resultados e procedimentos que lhes são apresentados. De outro, não se incorpora ao patrimônio técnico das equipes de avaliação a competência na área de medida e avaliação educacional. É, sem dúvida, preciso equacionar a distribuição de competências entre a equipe