



Ensaio: Avaliação e Políticas Públicas em
Educação
ISSN: 0104-4036
ensaio@cesgranrio.org.br
Fundação Cesgranrio
Brasil

Santos Fontanive, Nilma
A Divulgação dos Resultados das Avaliações dos Sistemas Escolares: limitações e
Perspectivas
Ensaio: Avaliação e Políticas Públicas em Educação, vol. 21, núm. 78, enero-marzo,
2013, pp. 83-100
Fundação Cesgranrio
Rio de Janeiro, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=399538144005>

- Como citar este artigo
- Número completo
- Mais artigos
- Home da revista no Redalyc

A Divulgação dos Resultados das Avaliações dos Sistemas Escolares: limitações e Perspectivas

Nilma Santos Fontanive*

Resumo

O artigo vai discutir o impacto que os resultados das avaliações em larga escala podem ter na prática docente com vistas a promover ganhos de aprendizagem. Os sistemas de avaliação de desempenho dos alunos, introduzidos no Brasil há quase duas décadas, apesar da enorme quantidade de dados coletados e divulgados, parecem não ter sido capazes de provocar as mudanças esperadas no cotidiano da relação professor-aluno-aprendizagem. São levantadas hipóteses de que os professores, coordenadores pedagógicos e diretores das escolas não compreendam adequadamente as tecnologias empregadas na avaliação dos estudantes e nos mecanismos e materiais utilizados na divulgação dos resultados e, assim, não são capazes de orientar a prática docente para corrigir os erros e as dificuldades apresentadas pelos alunos.

O artigo também faz referência a algumas iniciativas realizadas por agentes públicos da educação para promover melhorias na aprendizagem, mediante a criação de materiais estruturados para alunos e professores, incluindo questões de avaliação e exercícios para os alunos. Tais iniciativas podem produzir, a curto e médio prazos, impactos positivos no desempenho dos alunos. Discutem-se ainda as condições para que a avaliação efetivamente possa, em um futuro próximo, promover a aprendizagem dos estudantes.

Palavras-chave: Avaliação em Larga Escala. Impactos dos Resultados na Sala de Aula. Apropriação de Resultados pelos Professores.

Presentation of Results from School Systems Assessments: limitations and perspectives

Abstract

This paper discusses the impact of large scale assessments in the teacher practice towards promoting learning gains. The assessment systems of

* Doutorado em Educação, PUC-Rio; Coordenadora de Avaliação em Larga Escala da Fundação Cesgranrio.
E-mail: nilmaf@cesgranrio.org.br

cognitive achievement introduced in Brazil about two decades ago apparently was not able to induce the expected changes in the day-to-day relationship teacher-student learning, in spite of the enormous quantity of collected and released data. The paper raises the hypothesis that teachers, pedagogical counselors and school principals do not understand adequately the technologies used in student assessments and the materials released with the result and, in this way, are not capable to guide teacher's practice to correct the mistakes and difficulties presented by the students. The paper also makes references to some initiatives taken by public agents of education to improve learning, by creating structured materials for students and teachers, including assessment items and exercises for students. Such initiatives can produce, in the short and medium term, positive impacts in student achievement. The paper discuss yet conditions so that the assessment can affectively, in the near future, promote student learning.

Keywords: *Large scale assessment. Impact of the results in the classroom. Teacher appropriation of the results.*

La Divulgación de los Resultados de las Evaluaciones de los Sistemas Escolares: Limitaciones y Perspectivas

Resumen

El artículo discute el impacto que los resultados de las evaluaciones en larga escala pueden tener en la práctica docente con el objeto de promover mejora en el aprendizaje. Parece que los sistemas de evaluación de desempeño de los alumnos, introducidos en Brasil hace casi dos décadas, a pesar de la enorme cantidad de datos recogidos y divulgados, no fueron capaces de originar los cambios esperados en la relación cotidiana profesor-alumno-aprendizaje. Se elaboran hipótesis de que los profesores coordinadores pedagógicos y directores de las escuelas quizás no entiendan adecuadamente las tecnologías empleadas en la evaluación de los estudiantes y en los mecanismos y materiales utilizados en la divulgación de los resultados y, por eso tal vez no sean capaces de orientar la práctica docente para corregir los errores y las dificultades que los alumnos presentan. El artículo también se refiere a algunas iniciativas que los agentes públicos de la educación realizaron para mejorar el aprendizaje, mediante la creación de materiales estructurados para alumnos y profesores, incluyendo cuestiones de evaluación y ejercicios para los alumnos. Tales iniciativas pueden producir, a corto y medio plazo, impactos positivos en el desempeño de los alumnos. Finalmente se discuten las condiciones para que la evaluación efectivamente pueda, en un futuro próximo, promover el aprendizaje de los estudiantes.

Palabras clave: *Evaluación en Larga Escala. Impactos de los Resultados en el Aula. Apropiación de Resultados por parte de los Profesores.*

Contextualização

Os sistemas de avaliação do desempenho de alunos desenvolvidos no mundo contemporâneo, em particular nas duas últimas décadas, frequentemente aplicados a um grande número de estudantes e, por essa razão, chamados de sistemas de avaliação em larga escala, têm assumido prioritariamente um caráter diagnóstico. Eles podem ser mais bem definidos como sistemas de informações consistentes, confiáveis, válidos e comparáveis, voltados, em uma primeira instância, para fornecer às autoridades educacionais e aos demais interessados respostas sobre a qualidade da educação oferecida aos alunos pelas nações, estados, municípios ou determinada rede escolar.

Os sistemas de avaliação, em larga escala, mesmo com feições próprias em cada local de aplicação, em geral adotam um programa de testes cognitivos, externos à escola, associados a outros instrumentos que buscam identificar os contextos intra e extraescolares nos quais a educação se realiza. As avaliações em larga escala são bastante diferentes das práticas avaliativas realizadas pelos professores no âmbito de uma sala de aula, pois, não só abrangem um grande número de alunos de diferentes séries ou anos escolares, como também, precisam apresentar uma grande quantidade de questões para garantir a validade curricular, ou seja, que a avaliação abranja os principais conteúdos ensinados. Esses sistemas contam ainda com novas teorias e práticas de medidas educacionais que conjugam métodos de coleta de dados e diferentes maneiras de julgar sua qualidade, de um lado, e se apoiam no uso de modelos matemáticos e estatísticos complexos, de outro. Tais especificidades trouxeram para a área da Avaliação Educacional o grande desafio de buscar formas eficazes de divulgação dos resultados encontrados, uma vez que as tecnologias empregadas não são do senso comum dos professores, dos demais membros da comunidade escolar e da sociedade em geral.

Considere-se que, no cotidiano da escola, o professor avalia os alunos da sua turma, alunos, que ele conhece, com provas ou outros instrumentos que ele elabora, tendo como referências o programa de ensino que ele cumpriu e os materiais didáticos que selecionou para criar as experiências de aprendizagem na sala de aula. Assim, quando o professor atribui um grau ou um conceito a um aluno particular, pode avaliar com relativa precisão o que aquele aluno sabe, que habilidades adquiriu, o que ele aprendeu ou deixou de aprender. Portanto, a nota ou conceito atribuído aos alunos tem, para esse professor, um significado próprio, particular que é fruto do domínio do seu processo de trabalho como um todo. Se, por hipótese, esse resultado, nota ou conceito, fosse apresentado a um professor de outra escola, essa nota ou conceito só poderia ser bem interpretado, caso esse professor tivesse acesso às questões da prova ou aos outros instrumentos aplicados, porque o significado dos resultados dependem do conhecimento do que foi utilizado para obtê-los.

Outro aspecto a considerar quanto ao conhecimento preciso dos resultados das avaliações de desempenho refere-se ao uso desses resultados para fornecer *feedbacks* aos alunos e propiciar ganhos de aprendizagem. No chão da escola, a

interação professor-aluno possibilita, não só, que ele selecione variados instrumentos de avaliação - provas, trabalhos, *portfólios*, apresentações orais, entre outros - como também orienta a seleção dos recursos didáticos mais adequados para superar as dificuldades e corrigir as eventuais lacunas de aprendizagem encontradas.

Embora com as limitações inerentes às suas características e aos seus propósitos, os sistemas de avaliação em larga escala popularizaram-se no mundo e também no Brasil, a ponto de poder-se afirmar que, no século XXI, poucos são os países, estados ou municípios que deixaram de desenvolver e aplicar avaliações abrangentes do desempenho dos seus estudantes.

Apresenta-se como exemplo dois grandes sistemas de avaliação em larga escala: o Sistema de Avaliação da Educação Básica- SAEB/Prova Brasil e o Programa Internacional de Avaliação de Estudantes - PISA.

O SAEB, aplicado pela primeira vez em 1995 pelo Instituto de Estudos e Pesquisas Educacionais Anísio Teixeira- INEP/MEC, adotando as modernas técnicas de obtenção e análise dos desempenhos dos alunos, produziu escalas de proficiências em Língua Portuguesa e em Matemática, comparáveis ao longo de todos esses anos, permitindo assim o acompanhamento da evolução do desempenho dos alunos brasileiros. (KLEIN; FONTANIVE, 1995; FONTANIVE, 1997). O SAEB utiliza amostras de alunos de escolas públicas e particulares do País e é complementado, desde 2005, pela Prova Brasil, aplicada censitariamente a estudantes de escolas públicas brasileiras (KLEIN; FONTANIVE, 2009)

Ambos os Sistemas de Avaliação utilizam modelos estatísticos da Teoria da Resposta ao Item (TRI) para obtenção de escalas de desempenho. Essas escalas são invariantes - exceto pela escolha da origem e do intervalo de medida - e ordenam os desempenhos dos alunos do menor para o maior em um *continuum*.

Em 1997, as distribuições de proficiência, tanto em Língua Portuguesa como em Matemática, da 8^a série do Ensino Fundamental de 8 anos (hoje 9º ano do Ensino Fundamental de 9 anos) tiveram suas médias arbitradas em 250 e seus desvios padrões em 50. Os pontos interpretados das escalas do SAEB/Prova Brasil, até hoje, têm valores numéricos compreendidos entre 125 e 425, com intervalos de 25 pontos, meio desvio padrão. Os pontos das escalas são interpretados em termos das habilidades e conhecimentos que os alunos, em geral, dominam quando suas proficiências estão situadas em torno desses pontos das escalas. A interpretação é feita analisando-se as respostas dadas pelos alunos aos itens de teste aplicados que discriminam os desempenhos em cada um desses pontos da escala. (KLEIN, 2005).

É importante assinalar que as escalas do SAEB, até sua última edição em 2011, apresentavam as seguintes propriedades:

- são comuns aos anos e séries avaliadas para cada componente curricular;
- a interpretação dos pontos das escalas são cumulativas, ou seja, os alunos posicionados em um ponto em geral dominam as habilidades descritas nos pontos anteriores. Assim, quanto maior o ponto da escala, melhor o desempenho dos alunos nele posicionados.

Os pontos ou níveis como o 125, 150, 250, por si sós não determinam o significado da aprendizagem demonstrada pelos alunos, da mesma forma que a nota 7 ou o conceito B só fazem sentido para o professor que elaborou as questões, aplicou-as e corrigiu as provas. Entretanto, em uma avaliação em larga escala como o SAEB/ Prova Brasil faz-se necessário utilizar uma quantidade considerável de itens em cada componente curricular, ano/ série e, por essa razão, é pouco prático apresentá-los um a um para explicar os resultados obtidos.

Visando a facilitar a compreensão desses resultados, desenvolveu-se, desde 1995, uma metodologia de interpretação de escala baseada em dois procedimentos:

1. Identificação dos níveis âncora da escala

Deve-se acrescentar que a adoção da Teoria da Resposta ao item permite ao avaliador colocar em uma mesma escala os itens de teste utilizados na avaliação de uma disciplina e o desempenho dos alunos que participaram daquela avaliação. Para a interpretação das escalas obtidas são selecionados os itens que se posicionam em um determinado nível (ou ponto) da escala. A seleção desses itens para um nível da escala é feita a partir de critérios previamente definidos (KLEIN, 2005)¹.

2. Apresentação dos itens âncora à especialistas das áreas curriculares avaliadas para que eles interpretem o que os alunos sabem, são capazes de fazer ou que habilidades demonstraram possuir para acertar aqueles itens.

Os especialistas também recebem os resultados estatísticos obtidos pelos itens para analisar seus parâmetros, tais como índices de dificuldade, de discriminação, os coeficientes bisseriais e o ajuste das curvas de informação do item obtidas pela TRI. Os especialistas podem então apreciar dificuldade de cada item, tipos de erros mais frequentemente cometidos pelos alunos, analisar a atração de cada uma das alternativas, entre outros comportamentos que os itens apresentam.

¹ Nota Técnica: Um item é Âncora em um nível

- se o número de alunos no ponto que respondeu ao item é maior que 50. Esse valor é arbitrário e seu uso evita conclusões com poucas observações.
- O percentual de acerto do item nos pontos anteriores é menor que 65%;
- O percentual de acerto do item no ponto considerado e nos pontos acima é maior que 65%;
- A curva tem bom ajuste.

O PISA avalia, desde 2000, o desempenho de jovens de 15 anos dos países membros da OCDE e de outros países convidados. São avaliadas as disciplinas de Matemática, Ciências e Leitura, a cada 3 anos, sendo que na sua última edição em 2009, contou com a participação de 65 países.

Para facilitar a interpretação dos resultados, o PISA estabeleceu em cada ciclo de avaliação vários níveis de desempenho baseados em alguns princípios estatísticos (OCDE: PISA Results, v. 1). A interpretação dos níveis permite posicionar o desempenho dos estudantes e descrever o que são capazes de fazer.

A ideia principal das avaliações em larga escala, como o SAEB e o PISA, é saber como está o sistema educacional de um país e, de outros estratos de interesse, e não avaliar, classificar, aprovar ou promover indivíduos em particular.

O PISA também apresenta os resultados em escalas de habilidades com a pontuação que delimita os níveis de proficiência definido em cada um dos três domínios avaliados – Leitura, Matemática e Ciências. Os níveis vão de 1 a 6, nas 3 áreas, Leitura, Matemática e Ciências.

Em avaliações de larga escala, como é o caso do PISA que envolve países em diferentes estágios de desenvolvimento social e educacional, é previsível que a distribuição das proficiências dos diferentes países varie muito. Em qualquer das três áreas, considera-se que os estudantes abaixo do nível 1 de proficiência não são capazes de executar as tarefas mais simples que a avaliação solicita.

O Brasil, até o ano de 2006, evitou definir os pontos de proficiências considerados adequados nas duas disciplinas nos anos e séries escolares avaliadas. A partir de então, surgem duas iniciativas para definir padrões de qualidade para a Educação Básica: o Movimento Todos pela Educação e o Índice de Desenvolvimento da Educação Básica – IDEB.

O Movimento Todos pela Educação, criado em 2006 é uma aliança de representantes da sociedade civil, da iniciativa privada, organizações sociais, educadores e gestores públicos de Educação que tem como objetivo garantir Educação Básica de qualidade para todos os brasileiros até 2022, bicentenário da independência do País. O Movimento Todos Pela Educação definiu os pontos das escalas recomendados para as três séries avaliadas pelo SAEB apresentados na tabela abaixo.

Tabela 1 – Pontos das Escalas Recomendados pelo Movimento Todos Pela Educação

Série	Língua Portuguesa	Matemática
4 ^a série	200	225
8 ^a série	275	300
3 ^a série EM	300	350

Fonte: Todos Pela Educação.

O Índice de Desenvolvimento da Educação Básica – IDEB tem o objetivo de medir qualidade de cada escola e de cada rede de ensino. O indicador é calculado com base no desempenho dos estudantes no SAEB/Prova Brasil e em taxas de aprovação. Assim, para que IDEB de uma escola ou rede pública cresça é preciso que o aluno aprenda, não repita o ano e frequente a sala de aula.

Para que pais e responsáveis acompanhem o desempenho da escola de seus filhos, basta verificar o IDEB da instituição pública que é apresentado numa escala de zero a dez. Da mesma forma, gestores acompanham o trabalho das secretarias municipais e estaduais pela melhoria da educação. O índice é medido a cada dois anos e o objetivo é que o País, a partir do alcance das metas, tenha IDEB 6 nos anos iniciais em 2022 – correspondente à qualidade do ensino à média da OCDE.

Na última década, o Brasil assistiu ao crescimento da aplicação de avaliações do desempenho dos alunos nas principais redes públicas e particulares de estados e municípios. Inicialmente, a maioria dessas avaliações possuía um caráter independente, ou seja, obtiveram e descreveram seus resultados sem referência à avaliação nacional – SAEB/Prova Brasil. Muito embora as avaliações aplicadas permitiram aos gestores das redes determinar a qualidade da aprendizagem dos seus alunos, faltou a elas a possibilidade de relacionar esses desempenhos com aqueles obtidos pelos alunos brasileiros, prejudicando, em uma certa medida, o alcance das análises realizadas. Acrescente-se ainda que as referidas avaliações, de uma maneira geral, incidiram sobre séries/anos escolares distintos daqueles avaliados pelo SAEB/ Prova Brasil e também não utilizavam as tecnologias de obtenção e descrição de escalas de proficiências, impedindo o acompanhamento da evolução do desempenho dos alunos e, também a avaliação do impacto da adoção de políticas e programas de melhoria da qualidade do ensino-aprendizagem ao longo dos anos.

Entretanto, todo o esforço que os gestores das redes fizeram para implementar seus sistemas de avaliação foram muito importantes uma vez que contribuíram para colocar a avaliação na agenda de discussão e preocupação no País e ajudaram a criar uma cultura e uma consciência avaliativas.

Como consequência da multiplicação dos sistemas de avaliação, associado à popularização do SAEB/Prova Brasil, assiste-se, nos dias atuais, um crescimento do número de estados e municípios que, graças aos acordos de cooperação técnica realizados com o INEP/MEC, obtêm itens calibrados nas escalas do SAEB e, desse modo, colocam os resultados dos seus alunos nessas escalas, ampliando o escopo das avaliações do desempenho escolar realizadas.

Analizando-se a evolução do desempenho dos alunos na série histórica do SAEB/Prova Brasil, não se constata grandes ganhos de aprendizagem nos anos / séries escolares e disciplinas avaliadas, apesar de estarem disponíveis e divulgados tantos resultados do desempenho de alunos em todo o País. Uma pequena melhoria de aprendizagem já se verifica atualmente no 5º ano do E.F. Permanece então a indagação do porqué do pouco impacto das avaliações nas salas de aula e coloca-se como a premente necessidade de fomentar pesquisas em duas direções. A primeira seria investigar as formas de apresentação dos resultados às escolas visando identificar como esses resultados estão chegando aos professores e demais membros da comunidade escolar, se, por exemplo, eles são efetivamente compreendidos e capazes de orientar processos mudanças na prática pedagógica. A outra direção da pesquisa seria verificar quais são os impactos que os resultados das avaliações em larga escala têm provocado na formação inicial e continuada dos professores, coordenadores pedagógicos e diretores das escolas.

Estratégias da Apresentação dos resultados das Avaliações de Sistemas Escolares – Algumas Limitações

Desde a divulgação dos resultados do primeiro SAEB, realizado em 1995, os pesquisadores tiveram que enfrentar o desafio de divulgá-los de maneira que eles pudessem ser compreendidos pela comunidade escolar, pais e a todos os interessados em conhecer os níveis de desempenho de uma determinada população de alunos.

As escalas de proficiências foram divulgadas e, os pontos descritos, foram acompanhados de exemplos de itens demonstrativos do desempenho dos alunos nesses diferentes pontos. Apesar do esforço em traduzir pedagogicamente os resultados encontrados, as escalas não foram bem compreendidas pela comunidade escolar. Acredita-se que os professores tiveram dificuldades em transpor as descrições das proficiências para sua prática cotidiana, associando as habilidades apresentadas nos pontos das escalas aos enunciados dos objetivos dos seus programas de ensino, tão familiares para eles.

Essa afirmação pode ser ilustrada com os exemplos apresentados a seguir, nos quais algumas descrições das habilidades do ponto 125 da escala de Língua Portuguesa e do 150 da escala de Matemática do SAEB foram selecionadas para subsidiar a discussão que está sendo proposta.

Quadro 1 – Escala do SAEB – Nível 125

125	<p>A partir de textos curtos, como contos infantis, histórias em quadrinhos e convites, os alunos do 5º e da 9ºanos do Ensino Fundamental:</p> <p>localizam informações explícitas que completam literalmente o enunciado da questão; inferem informações implícitas; reconhecem elementos como o personagem principal; interpretam o texto com auxílio de elementos não-verbais; identificam a finalidade do texto; estabelecem relação de causa e consequência, em textos verbais e não-verbais e conhecem expressões próprias da linguagem coloquial.</p>
-----	---

Fonte: Escola de língua portuguesa do SAEB.

As habilidades descritas informam o que os alunos demonstraram saber fazer quando apresentados aos gêneros literários especificados: contos infantis, histórias em quadrinhos e convites. Eles deveriam ser capazes de, por exemplo, de reconhecer o personagem principal de um conto infantil ou de uma história em quadrinhos. Essa é uma habilidade que o professor trabalha frequentemente nas suas aulas quando, utilizando os materiais ou livros didáticos, explora a história com perguntas do tipo: quem é o personagem principal da historia? Como ela/ele se chama? Igualmente, o professor costuma ter como objetivo do ensino das habilidades de leitura induzir o aluno a identificar a finalidade do texto (para que o texto serve? Para convidar as pessoas a uma festa de aniversário, por exemplo), ou ainda, em uma história infantil conhecida como a do chapeuzinho vermelho trabalhar as relações de causa e consequência, tais como: Porque o lobo vestiu-se com as roupas da vovó? O que ele pretendia? Porque a chapeuzinho desconfiou que não era a vovó? O que ela fez? E assim por diante.

Na disciplina matemática o ponto 150 da escala apresenta a descrição das seguintes habilidades:

Quadro 2 – Escala do SAEB – Nível 150

150	<p>Os alunos do 5º e do 9ºanos do Ensino Fundamental:</p> <p>resolvem problemas envolvendo adição ou subtração, estabelecendo relação entre diferentes unidades monetárias (representando um mesmo valor ou numa situação de troca, incluindo a representação dos valores por numerais decimais); reconhecem o valor posicional dos algarismos em números naturais e localizam números naturais (informados) na reta numérica; leem informações em tabela de coluna única e identificam quadriláteros.</p>
-----	---

Fonte: Escala de matemática do SAEB.

A análise do trecho apresentado conduz a interpretações bastante claras dos conceitos e operações matemáticas, em geral, dominadas pelos alunos cujos desempenhos posicionam-

se no nível 150 da escala. Eles sabem resolver problemas com adição e subtração em situações que apresentem unidades monetárias incluindo as operações com centavos (numerais decimais), conhecem o valor de posição de números naturais e localizam esses algarismos em uma reta numérica. Dada uma tabela de coluna única, os alunos são capazes de extrair as informações solicitadas. Na área de geometria os alunos são capazes de identificar o quadrilátero entre formas geométricas apresentadas na questão.

A escala de matemática foi descrita em 13 níveis do 125 a 425 e a de Língua Portuguesa apresenta 11 níveis do 125 a 375. As descrições das habilidades de cada um desses níveis podem ser encontradas em www.inep.gov.br.²

Como os dois exemplos mostram, as habilidades descritas nos níveis das duas escalas utilizam linguagem clara e cujos significados são bem próximos dos enunciados dos objetivos instrucionais encontrados nos programas de ensino, materiais e livros didáticos.

A maioria dos materiais de divulgação dos resultados dirigidas à comunidade escolar, divulgam alguns exemplos de itens representativos dos pontos das escalas. Esses itens, em geral acompanhados das suas estatísticas, procuram mostrar para os interessados a habilidade requerida do aluno para que ele acerte o item. Além das estatísticas, é frequente a apresentação de comentários pedagógicos, explicações da solução do item e ainda análise das respostas erradas e correção dos erros.

Figura 1 - Exemplo de item de Ciências – SIADE 2009

As imagens abaixo mostram pés de três aves que vivem em ambientes diferentes.



1



2



3

Fonte: STEFANELLI, G.; VALENTINI, G. Scienze 1. Paraíba. Italy, 1988, p. 315.

Com base no tipo de pés dessas aves, podemos afirmar que, provavelmente, a ave

A) 1 sobe bem em árvore e a 3 é nadadora.
 B) 1 é nadadora e a 2 sobe bem em árvores.
 C) 1 só anda no chão e a 2 é nadadora.
 D) 3 só anda no chão e a 1 sobe bem em árvores.

		PERCENTUAIS DE RESPOSTAS			
GAB	A	B	C	D	
B	0.06	0.75	0.08	0.11	

Esperava-se que os alunos associassem o tipo de pés das aves aos seus hábitos. Como se vê na figura 1, aves com este tipo de pés nunca poderiam subir em árvores; no entanto, elas são bem adaptadas à natação; já aves com os pés da figura 2 são adequados para subir em árvores.

Foram 75% os alunos que conseguiram fazer essas associações corretamente; para os demais, esta habilidade precisa ser ainda fortalecida, uma vez que, a despeito do aspecto dos pés do tipo 1, as demais respostas indicam que os alunos entendem que a ave com este tipo de pé sobe bem em árvores (*distratores* A e D) e só anda no chão (*distrator* C).

Fonte: Sistema de Avaliação do Desempenho das Instituições Educacionais do Distrito Federal (SIADE) 2009.

² <http://portal.inep.gov.br/web/prova-brasil-e-saeb/prova-brasil-e-saeb>

Nesse outro exemplo, apresentado a seguir, os autores do material de divulgação optaram por apresentar as estatísticas completas obtidas pelo item com a adoção da Teoria Clássica dos Testes (TCT) e da Teoria da Resposta ao Item (TRI), incluindo também comentários e análises da resposta correta e dos erros. Em um primeiro momento, o público-alvo dos materiais estranharam a complexidade dessas estatísticas e tiveram dificuldade em compreendê-las e aceitá-las.

No entanto, apesar da resistência inicial, os responsáveis por essas avaliações continuaram a apresentá-las, acreditando que com essa decisão estavam contribuindo para formar uma cultura de avaliação nos professores, coordenadores pedagógicos e gestores de redes. Houve também o esforço de promover inúmeros seminários e jornadas de capacitação, análise e discussão dos resultados.

Figura 2 - Exemplo de item de Matemática – Rio Grande do Sul 2009

DONA MÁRCIA ESCREVEU NA LOUSA A SEGUINTE OPERAÇÃO:

$524 + 47 =$

O RESULTADO DESSA OPERAÇÃO É:

561.
 571.
 994.

ÍNDICES												
ITEM	GAB	TIPO	DIFI	A	B	C	D	PPOLIS	PBISA	PBISB	PBISC	PBISD
5	B	1	0,46	0,19	0,46	0,22	0,12	0,48	-0,14	0,48	-0,20	-0,32

RS08 Mat02 It 5 Ibg 16 a=0.02 b= 187.73
 aorig=1.18 borig= 0.69

Tema: Números e operações
Habilidade: Determinar o resultado de uma adição com ou sem reagrupamento (reserva).

Continuação

Comentários pedagógicos:

Para resolver adequadamente este item, os alunos precisariam conhecer as regras do sistema de numeração decimal e o algoritmo da adição, fazendo reagrupamento da ordem das unidades para a ordem das dezenas. Os alunos que acertaram esta questão marcaram como resposta o número 571.

Este item foi relativamente difícil, pois menos da metade dos alunos avaliados 5 2 4
o acertaram (46%). Além da necessidade de reagrupar as unidades, o fato + 4 7
de as parcelas terem quantidade diferente de ordens pode ter aumentado o 9 9 4
grau de dificuldade.

Os alunos que marcaram o número 994 como resposta devem ter armado a conta de forma equivocada, como mostrado ao lado. Essa alternativa incorreta foi escolhida por 22% dos alunos. 19% dos alunos testados marcaram o número 561 como resultado da operação e devem ter armado corretamente a conta, mas quando reagruparam as unidades ($4u + 7u = 11u = 1d + 1u$), encontraram uma unidade solta, mas esqueceram de somar a dezena formada às já existentes nas parcelas.

Houve uma boa correlação entre o desempenho no teste e o acerto no item.

A curva de informação do item mostrada na figura revela que o item posiciona-se no nível 225 da escala. Aproximadamente 15% dos alunos estão acima do nível 225 nessa escala.

Fonte: Projeto de Avaliação de Leitura e Escrita e Matemática do Ensino Fundamental no Rio Grande do Sul.

Entretanto, mesmo depois de quase duas décadas de divulgação das escalas e, apesar da grande quantidade de material produzido, tanto pelos técnicos do INEP/MEC, secretarias de educação e equipes das Instituições responsáveis pelas avaliações nos diferentes níveis de abrangência, constata-se que os resultados das avaliações parecem não ter sido adequadamente incorporados à prática docente e capazes de promover a melhoria da aprendizagem dos alunos da Educação Básica, no Brasil.

Em outras palavras, as tentativas de simplificar informações complexas, tanto nos materiais de divulgação de variados formatos, como, manuais, livretos, boletins, entre outros, quanto na realização de numerosos seminários, videoconferências, jornadas de capacitação de professores, coordenadores pedagógicos e gestores, parecem não ter sido bem sucedidas, pois não provocaram os impactos esperados na qualidade e equidade da educação oferecida às crianças e aos jovens brasileiros.

É claro também que a melhoria da aprendizagem dos alunos está condicionada a outros fatores, e que a divulgação dos resultados das avaliações tem um papel relativo no conjunto das práticas educacionais, no entanto, seria razoável esperar que o desempenho dos alunos brasileiros apresentasse uma evolução mais positiva em quase duas décadas de avaliação sistemática.

É provável que o pouco impacto verificado seja explicado, em uma primeira análise, pela distância entre resultados obtidos nacionalmente, como os do SAEB, ou por secretarias estaduais e municipais de educação e a sala de aula ou o chão da escola, a despeito das inúmeras iniciativas fazer os resultados chegarem aos professores e demais membros da comunidade escolar, como apontado nos parágrafos anteriores. Em resumo, tem-se a impressão de que o professor não se identifica com esses resultados ou não vê sua prática pedagógica neles representada.

O Uso das avaliações de desempenho para promover ganhos na aprendizagem dos alunos

Nos meados do anos 60, introduziu-se no cenário pedagógico o conceito de avaliação formativa desenvolvido por Scriven (1967) e com a função de constatar o que está sendo aprendido mediante coleta de informações feita pelo professor de forma contínua, com diversos procedimentos metodológicos e julgar o grau de aprendizagem, ora em relação a todo grupo-classe, ora em relação a um determinado aluno em particular.

A avaliação formativa, na sua concepção original, não tem como objetivo classificar, selecionar, aprovar ou reprovar o aluno, mas adequar o processo de ensino ao grupo e ou àqueles alunos que apresentam dificuldades, tendo em vista o alcance dos objetivos propostos.

Para que o aluno seja imediatamente informado dos seus erros e acertos, o professor, após a aplicação dos testes, de exercícios, de questões construídas pelos alunos (abertas) deve corrigir as questões com a classe, utilizando variadas formas de estimular a participação de todos os alunos. Os alunos podiam ser convidados a apresentar oralmente como chegaram à resposta ou irem ao quadro para demonstrar como resolveram a questão.

O professor deve aproveitar os erros mais frequentes para promover o reensino dos aspectos da matéria não dominados pela maioria dos alunos. explicando o significado dos erros.

Há bastantes evidências na literatura sobre a importância do professor corrigir os erros no momento em que eles são cometidos, pois a motivação é ativada quando o aluno corrige suas respostas logo após ter realizado o esforço para resolver as questões ou realizar o trabalho, Pesquisa realizada por Rosenshine e Stevens, em 1986, citada por Bressoux (2003 p.35) aponta que os professores eficazes corrigem imediatamente os erros dos alunos a fim de evitar que eles se tornem sistemáticos.

Num estudo mais recente feito por Frome, Lasater e Cooney (2005), os autores usaram informações sobre características dos professores de escolas secundárias e as ligaram com o desempenho dos alunos da 8^a série em avaliação

na Geórgia. Analisando dados da experiência e titulação e da prática docente eles encontraram características dos professores e das salas de aula de Matemática que se correlacionavam com melhores desempenhos.

- motivação do professor e expectativas positivas de sucesso dos alunos;
- prática instrucional tal como trabalho em grupo em situações desafiadoras, apresentação oral e relatórios escritos em projetos de Matemática, explicações da solução de problemas para a classe estavam correlacionados com altos desempenhos;
- participação em atividades de monitoramento, capacitação e indução de novas experiências de aprendizagem nas escolas estão correlacionadas com mais altos escores alcançados pelos alunos em Matemática.

Nos últimos anos, algumas secretarias de educação vêm produzindo materiais didáticos para professores e alunos, estruturados sob a forma de livros para o professor e cadernos/ folhas de exercícios para os alunos. Tais materiais não só unificam e sequenciam os programas de ensino, como também propõem questões/ sugestões de atividades de avaliação da aprendizagem.

Tais iniciativas, se bem aceitas e usadas pelos professores de uma rede de ensino, têm grande potencial de provocar mudanças positivas no desempenho dos alunos, desde que o professor saiba usar os resultados das avaliações e resolução de exercícios de maneira formativa para promover a aprendizagem.

Essas iniciativas são relativamente recentes para que possam ser observados aumentos significativos na aprendizagem dos alunos. Entretanto é de suma importância que essas iniciativas continuem e se multipliquem no País e que os professores estejam capacitados – em todos os níveis – para usarem os resultados das avaliações, tanto as em larga escala quanto as do dia a dia, para melhorarem o desempenho dos alunos fazendo efetivamente um uso formativo dos seus resultados,

Van Haneghan (2009) chama a atenção para que as informações sobre as avaliações sirvam para facilitar a aprendizagem. O conjunto de escores numéricos sem referência aos conceitos ou procedimentos sobre o que os alunos aprenderam não provê informação adequada para relacionar a avaliação com a aprendizagem.

Se o professor não tem o conhecimento conceitual para interpretar a avaliação, ele é incapaz de determinar como aquela avaliação pode ajudar o aluno a aprender.

As iniciativas de fornecer *feedbacks* aos alunos sobre os resultados da avaliação e também disponibilizar materiais didáticos para alunos e professores são bastante recentes no Brasil e, ainda não há evidências conclusivas sobre o impacto desses programas na melhoria do desempenho dos alunos.

Seria importante direcionar pesquisas para estudar o efeito do uso das avaliações externas na evolução do desempenho dos alunos.

No Brasil são poucas as pesquisas que procuram ligar à prática docente a melhoria do desempenho dos alunos, medido por avaliações externas. Essa não é uma área de fácil investigação, pois exige o uso de instrumentos variados que requerem um grande investimento de tempo para desenvolvê-los e validá-los, treinar pessoas para aplicá-los, analisar e consistir os bancos de dados.

Essas pesquisas que procuram relacionar a qualidade da prática docente e o desempenho dos alunos são também difíceis de conduzir, pois de acordo com Braun (2004) é difícil isolar o efeito da instrução do professor dos outros efeitos no nível da sala de aula, como clima, influência dos pares, disponibilidade de livros e outros materiais didáticos e outros fatores extraescolares que estão fora do alcance do professor.

Por outro lado investigar a relação entre o efeito do professor na aprendizagem de um grupo de alunos é bastante complexo do ponto de vista do desenho metodológico da pesquisa. Precisa-se contar com uma medida de desempenho de entrada, os fatores extraescolares, como as características das famílias, ou a pobreza, devem ser controlados, os alunos precisam permanecer um bom tempo com o mesmo professor e é necessário contar dados longitudinais de desempenho dos alunos nos anos focalizados no estudo.

Conclusões

O primeiro aspecto a ser discutido é o que diz respeito à ênfase que deve ser dada pelos programas de capacitação docente ao domínio da matéria, investindo fortemente no ensino dos conteúdos dos programas das disciplinas da Educação Básica e garantindo que os professores efetivamente os aprendam. Para tal, os programas devem ser relativamente longos, sistemáticos e apresentar mecanismos de avaliação frequentes dos professores, aplicando testes, exercícios e ou trabalhos sobre os conteúdos ensinados. É importante que os programas de capacitação criem também as condições de reforço para os professores que não alcancem os níveis adequados de aprendizagem.

Da mesma forma, é importante ensinar o professor a transpor os conteúdos aprendidos para seu dia a dia na sala de aula, criando situações dinâmicas de aprendizagem e, para isso, os programas de melhoria da qualidade docente devem desenvolver materiais de formatos variados, como roteiros, cadernos de exercícios para o aluno, Dvds com simulações de situações concretas e como o professor deve atuar na sala e na escola. Esses Dvds, além de demonstrarem o ensino de aspectos críticos do ensino dos conteúdos, podem também abordar questões sobre indisciplina, falta de motivação e violência, dinâmica do trabalho em grupos cooperativos, entre outros. É possível também, nos dias atuais, usar o computador para manter o professor em uma rede de comunicação entre os consultores e coordenadores das capacitações e com os demais colegas do curso, com plataformas de educação a distância.

O acompanhamento do desempenho dos alunos deve ser realizado frequentemente e deve ser fornecido ao professor um banco de questões de avaliação formativa para serem aplicadas nos alunos. Há experiências no Brasil, criada pelo INEP que disponibilizou no seu Portal questões de alfabetização que os professores e escolas podem aplicar para acompanhar a aprendizagem dos alunos. Não seria impossível, portanto, criar e atualizar um banco de questões das disciplinas do Ensino Fundamental e Médio calibrados nas escalas do SAEB e colocá-las à disposição das escolas e dos professores. É importante por outro lado, ajudar ao professor a interpretar os resultados obtidos e os programas de capacitação têm que ensinar aos professores a dominar a tecnologia de avaliação e de interpretação de escalas de proficiências, empregada com sucesso no Brasil desde 1995.

O acompanhamento do desempenho dos alunos não pode, por outro lado, dispensar medidas de aprendizagem externas e, assim, os programas de capacitação precisam criar ou aproveitar mecanismos de avaliação externos às escolas.

Por fim, sugere-se o debate sobre dois aspectos importantes e que o Brasil precisa desenvolver e divulgar tecnologias apropriadas.

O primeiro trata dos mecanismos de monitoramento dos programas de capacitação dos professores na sala de aula e nas escolas, que tenham características de controle da efetividade docente e de apoio para as dificuldades encontradas. A formação, a supervisão e o acompanhamento desse profissional altamente especializado é um desafio que precisa ser enfrentado em conjunto com as Universidades, o Ministério de Educação e Secretarias de Educação.

A última questão levantada para o debate diz respeito à carreira docente sugerindo que, os estudos que estão hoje na agenda do Ministério de Educação e do Conselho Nacional de Secretários de Educação incluam, na avaliação docente, para admissão ou promoção na carreira, uma parcela variável da remuneração, além da titulação, anos de serviço, aprovação em exames, o desempenho dos alunos.

Para concluir retomam-se as recomendações do Relatório McKinsey, que estudou os melhores sistemas educacionais do mundo e aqueles que estavam situados nas primeiras posições apontaram que para garantir altos desempenhos dos alunos deve-se:

- atrair as pessoas certas para a carreira docente;
- transformá-las em instrutores efetivos;
- garantir que a escola e o sistema sejam capazes de fornecer a melhor instrução possível para cada aluno.

Referências

- BRESSOUX, P. As pesquisas sobre o efeito escola e o efeito professor. *Educação em Revista*, Belo Horizonte, n. 38, dez. 2003.
- BRAUN, H. *Value-added modeling: What does diligence require?* Princeton, NJ: Educational Testing Service, 2004.
- FONTANIVE, N.S. Avaliação em larga escala no Brasil: Sistema Nacional de Avaliação da Educação Básica (SAEB/95). In: BOMENY, HELENA. (Org.). *Avaliação e determinação de padrões na educação latinoamericana*. Rio de Janeiro: Fundação Getúlio Vargas/Preal, 1997.
- FROME, P.; LASATER, B; COONEY, S. *Well-qualified teacher and high-quality teaching: are they the same?* Atlanta, GA: Southern Regional Educational Board, Research Brief, 2005.
- FUNDAÇÃO CESGRANRIO. *Relatório Pedagógico do SIADE, Ciências e Ciências da Natureza*. Rio de Janeiro: Cesgranrio, 2009.
- _____. *Avaliação de Leitura, Escrita e Matemática do Ensino Fundamental do Rio Grande do Sul. Relatório Pedagógico*. Rio de Janeiro: Cesgranrio, 2009.
- INEP. Instituto Nacional de Estudos e Pesquisas Educacionais Anisio Teixeira. Brasilia, c. 2011, Disponível em: <www.inep.gov.br>. Acesso em: 19 nov. 2012.
- KLEIN, R.; FONTANIVE, N. S. Avaliação em larga escala. *Em Aberto*, Brasília, DF, v. 15, n. 66, p. 29-34, 1995.
- KLEIN, R. Utilização da Teoria de Resposta ao Item no Sistema Nacional de Avaliação da Educação Básica. *Ensaio – Aval. e pol. pub. Edu.*, Rio de Janeiro, v. 11, n. 40, p. 283-296, jul./set, 2003.
- KLEIN, R. Testes de Rendimento Escolar. In: SOUZA, A. M. (Org). *Dimensões da avaliação Educacional*. Petrópolis, RJ: Vozes, 2005. p. 110-138.
- KLEIN, R.; FONTANIVE, N. Alguns indicadores educacionais de qualidade no Brasil de hoje. *São Paulo em Perspectiva, São Paulo*, v. 23, n. 1, p. 19-28, jan./jun, 2009.
- McKINSEY REPORT: How the World's Best-performing School Systems Come Out on Top. London: McKinsey&Company, 2007.

OCDE: PISA 2009 Results: What Students Know and Can Do? [S.I.]: OECD, 2010. v.1.

ROSENSHINE, B; STEVENS, R. Teaching functions. In: WITROCK, M. C. (Ed.). *Handbook of Research on Teaching*. 3nd ed. New York: Macmillan, 1986. p. 376-391.

SCRIVEN, M. The methodology of evaluation . In: TYLER, R. W.; GAGNÉ, R. M.; SCRIVEN, M. (Ed.). *Perspectives of Curriculum Evaluation*. Chicago: Rand McNally, p. 39-83, 1967.

TODOS pela Educação. São Paulo,[s.d.]. Disponível em: <www.todospelaeducacao.org.br>. Acesso em: 19 nov. 2012.

VAN HANEGHAN, J. P. Validiting assessment for learning: Consequential systems approaches. *Journal of MultiDisciplinary Evaluation*, v. 6, n. 12, june 2009.

Recebido em: 19/12/2012

Aceito para publicação em: 18/04/2013