

ÍNDICE DE DESENVOLVIMENTO DA EDUCAÇÃO DE SÃO PAULO – IDESP

bases metodológicas

JOSÉ FRANCISCO SOARES

Resumo: Os sistemas educacionais de ensino básico brasileiros têm sido analisados pelos resultados do aprendizado de seus alunos. Este artigo apresenta as bases metodológicas, introduzidas pela Secretaria de Estado da Educação de São Paulo, para a produção do Índice de Desenvolvimento da Educação de São Paulo – Idesp. Este indicador, que será usado para monitoramento das escolas paulistas, além de ser uma síntese da qualidade da escola, capta também a equidade das estratégias educacionais usadas.

Palavras-chave: Idesp. Educação. Qualidade. Equidade.

IDESP – DEVELOPMENT OF EDUCATION INDEX OF SAO PAULO: METHODOLOGICAL BASES

Abstract: The systems of basic education in Brazil are now analyzed by the learning results of its students. This paper presents the methodological bases of Idesp – Development of Education Index of Sao Paulo. This index, developed as tool for monitoring the schools, is at the same time a school quality synthesis and a measure of the equity of its educational practices.

Key words: Idesp. Education. Quality. Equity.

Em 2008, a Secretaria de Estado da Educação de São Paulo tomou decisões de grande impacto para o sistema estadual paulista de ensino básico. Primeiramente, adotou a escala de medida usada nacionalmente para apresentar os resultados do Sistema de Avaliação da Aprendizagem de São Paulo – Saresp. Isso tornou os dados do sistema comparáveis com aqueles gerados no âmbito nacional e nos outros Estados, dando-lhes maior credibilidade e visibilidade. Em seguida, definiu nova forma de divulgação dos resultados, mais relevante pedagogicamente, e, finalmente, criou um novo indicador para o monitoramento de suas escolas.

O objetivo deste artigo é mostrar as bases metodológicas dessas decisões, fundamentadas na ideia de que a educação escolar – a dimensão da educação que ocorre no âmbito da escola – deve ser estudada, principalmente, pela sua contribuição para o atendimento dos direitos educacionais dos alunos. Entre estes, destacam-se o direito ao acesso à escola, à permanência nesta e ao aprendizado.

No Brasil, e especialmente em São Paulo, o acesso ao ensino fundamental está praticamente universalizado. No entanto, sua ausência em épocas anteriores teve como consequência o analfabetismo, tema há muito presente nas análises educacionais brasileiras, como mostra o clássico artigo de Lourenço Filho (1965). O fluxo pode ser caracterizado pela repetência, que, como mostrou Sérgio Costa Ribeiro ao cunhar o termo “pedagogia da repetência”, é um dos flagelos da educação pública brasileira. Outra dimensão do fluxo, captada pela permanência do aluno na escola e pela conclusão das etapas em que o ensino se organiza, é usualmente medida pelo número de anos completos de estudo. Como esse indicador pode ser calculado com dados coletados pelo Censo Demográfico e, anualmente, pela Pesquisa Nacional por Amostra de Domicílios – PNAD, existem séries históricas confiáveis para estudos sobre fluxo. Um exemplo é o artigo de Barros, Henriques e Mendonça (2002), que analisou tendências históricas do sistema educacional brasileiro usando informações relativas à escolaridade média da população do país.

No entanto, os indicadores de qualidade da educação elaborados com base nos anos de estudo são limitados. Duas pessoas com a mesma escolaridade, ao frequentarem escolas diferentes, podem adquirir capacidades cognitivas muito distintas e assim ter inserção social também distinta. A partir da organização do Sistema Nacional de Avaliação da Educação Básica – Saeb e da construção de escalas de medidas em 1995, o Brasil passou a dispor de dados que informam sobre o desempenho dos alunos. Com a introdução da Prova Brasil em 2005, foram testados os alunos das escolas públicas com mais de 20 alunos na série. Como consequência de todos esses fatos, hoje se pode estudar a educação básica brasileira por meio dos resultados de aprendizado de seus alunos.

Essa maneira de considerar a educação está completamente consolidada seja nos movimentos da sociedade civil que acompanham o desenvolvimento da educação, como o Compromisso Todos pela Educação, seja por atos do governo federal, que ao criar o Plano de Desenvolvimento da Educação – PDE, por meio do Decreto n. 6.094, de 24 de abril de 2007, afirmou que a iniciativa visa “estabelecer como foco

a aprendizagem, apontando resultados concretos a atingir”. Esta é também a forma de atuar de muitos outros países, principalmente aqueles que participam do *Program for International Student Assessment – Pisa*, ‘programa de avaliação estudantil internacional’, cujo objetivo é “monitorar os resultados dos sistemas educacionais em termos do aprendizado de seu alunado”.¹

A Secretaria de Estado da Educação de São Paulo, além de adotar o monitoramento de resultados do aprendizado, inovou ao incorporar políticas educacionais orientadas à equidade, assegurando que os avanços educacionais obtidos não ficassem restritos a alguns grupos de alunos.

Importante ressaltar que eleger os resultados da aprendizagem dos alunos como forma privilegiada de análise da realidade educacional não é postura consensual no debate educacional brasileiro. Os insumos e as condições de trabalho dos professores têm sido tomados como as questões centrais. Sem rejeitar essas dificuldades, este texto assume que as questões devem ser tratadas dentro do novo marco teórico: o direito de aprendizagem de crianças e jovens.

MEDIDA DA COMPETÊNCIA

O uso de uma medida de resultados é necessário somente quando se pretende acompanhar todos os alunos de uma escola ou sistema e, portanto, o número de alunos envolvidos impede a consideração individual. Todo aluno tem o direito de ser avaliado por seus professores durante as atividades de ensino de forma que suas eventuais necessidades educacionais sejam conhecidas e possam ser atendidas. Ou seja, os sistemas de medida do aprendizado dos alunos por estruturas externas às escolas não substituem as avaliações feitas pelos professores como parte das estratégias de implementação de seu projeto pedagógico.

Ainda não é consensual qual aprendizado deve ser medido. O Saeb mede as competências de leitura e matemática. O ensino médio é também avaliado por meio do Exame Nacional de Ensino Médio – Enem e dos vestibulares, exames com especificações muito diferentes da utilizada no Saeb. Além desses exames nacionais, o Brasil participa regularmente do Pisa,

que testa alunos de 15 anos nas competências científica, leitora e matemática.

Para a produção das medidas das competências é preciso primeiramente definir a competência de interesse e a partir dela produzir itens que verifiquem o seu domínio nas suas várias expressões. Há duas escolas de pensamento a este respeito, conforme se opte mais ou menos por uma abordagem centrada em competências.

O Pisa, conforme seus relatórios técnicos (OECD, 2005), descreve em geral o domínio de interesse, de forma suficientemente clara para orientar a produção de itens. Depois de análise minuciosa dos itens considerados adequados, estes são reanalisados por um processo denominado *skill audit*, em que as habilidades necessárias para sua solução são explicitadas e registradas. Assim sendo, os documentos do Pisa que descrevem as competências não são organizados como detalhamento de habilidades. No entanto, os itens incluídos no teste são classificados em três grandes categorias: tópicos de conteúdo, situação de uso e processos cognitivos necessários para a sua solução.

A tradição do Saeb, adotada também por todos os sistemas estaduais, inclusive pelo Saesp, é exatamente a oposta. Refletindo a abordagem mais curricular destes exames, as habilidades e conteúdos que descrevem as competências a serem medidas são listadas *a priori*. A descrição das habilidades é também usada como comando para construção de itens, sendo usual a recomendação que cada item contemple apenas uma habilidade.

Construído e aplicado o teste, a medida da competência, que recebe o nome técnico de proficiência, é obtida com o uso de modelos da Teoria da Resposta ao Item – TRI. Uma função matemática, denominada curva de característica do item, mostra como a probabilidade de acerto do item se associa à proficiência do examinando e à dificuldade do item. Assim, a probabilidade de um acerto de um item aumenta com a proficiência do examinando e cai de acordo com a dificuldade do item.

Em provas de múltipla escolha, o modelo logístico de três parâmetros da TRI é usualmente preferido. Esse modelo tem um parâmetro (*b*) para representar a dificuldade, outro (*a*) que capta o poder de discrimi-

nação e um terceiro (*c*) que representa a taxa de acerto casual do item. O modelo de dois parâmetros é geralmente usado com itens de respostas construídas curtas, corrigidas apenas como acerto e erro, situação onde não há acerto casual. Finalmente, modelos de crédito parcial são usados para questões abertas e redações.

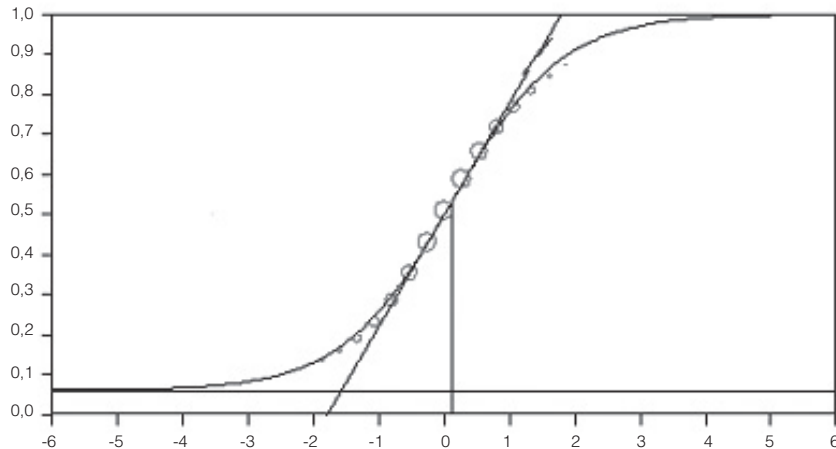
A Figura 1 mostra a curva do modelo de três parâmetros para um item que tem nível de dificuldade $b = 0,114$, discriminação $a = 0,796$, e acerto casual baixo de $c = 0,088$. Essa função matemática representa a relação entre os dados empíricos, i.e., as respostas ao item, e qualidades não observadas, as proficiências latentes. O parâmetro de dificuldade *b*, representado pela linha vertical da Figura 1, caminha na curva de característica para a direita ou esquerda, conforme o nível de dificuldade cresce ou diminui. O parâmetro da discriminação do item *a*, representado pela linha tangente à curva, altera a inclinação da curva característica, captando o poder de discriminação do item. O parâmetro do acerto casual do item *c*, representado pela linha horizontal na base do gráfico, representa a probabilidade de um acerto por pessoas sem nenhuma proficiência na matéria medida pela prova.

Os círculos na Figura representam os dados empíricos, e o tamanho de cada um deles representa a frequência relativa observada em cada ponto.

O uso da TRI exige que o teste meça apenas uma competência, hipótese de unidimensionalidade, que precisa ser verificada. Quando isso ocorre pode-se estudar a competência por meio de uma única medida-síntese, sem necessidade de considerar qual foi a resposta dada a cada um dos itens do teste. Um ajuste dos dados ao modelo, adequado para as situações reais, requer que o princípio da unidimensionalidade se aplique pelo menos aproximadamente. Uma implementação prática dessa recomendação exige que a matriz de correlação entre os itens revele um fator dominante que responda por pelo menos 20% da variância total, e seja substancialmente maior que o segundo fator. Nessas circunstâncias é viável atribuir um único escore de proficiência a cada examinando. Esse fator dominante ou traço latente representa a proficiência medida pela prova.

Figura 1

Curva Característica de um Item no Modelo de Três Parâmetros da TRI



O pressuposto adicional da independência local implica que cada item pode ser tratado como um evento independente. Isso requer que as respostas aos itens do teste sejam independentes do ponto de vista estatístico, depois de considerada a proficiência de cada examinando. Ou seja, o uso da TRI assume que a resposta a qualquer item do teste não deve determinar ou influir na resposta a qualquer outro item. Nestas condições, denominadas independência local, a probabilidade de obter qualquer conjunto de respostas de item se dá pelo produto das probabilidades individuais dos itens. Isso permite a utilização do princípio da máxima verossimilhança para estimar a proficiência dos examinandos.

Quando pode ser demonstrado que as respostas dos examinandos se ajustam ao modelo de resposta, os parâmetros de cada item não dependem dos examinandos avaliados em determinada ocasião, e também que as estimativas das proficiências não dependem dos itens aplicados em determinada ocasião. Da mesma forma, os procedimentos da verossimilhança acomodam bem a substituição, a eliminação e o acréscimo dos itens, assim como também toleram bem a ausência de respostas em itens que não foram resolvidos pelos alunos ou que não lhes foram

apresentados. Finalmente, a TRI permite que testes diferentes aplicados a grupos distintos de alunos sejam organizados de forma que possam produzir proficiências comparáveis. Isso permite que os resultados de testes de anos diferentes possam ser confrontados. Todos estes fatos explicam o amplo uso da TRI em processos de avaliação educacional, apesar das fortes hipóteses exigidas. Outros detalhes da TRI podem ser encontrados em Hambleton, Swaminathan, e Rogers (1991).

INTERPRETAÇÃO DA ESCALA DE MEDIDA

Na situação hipotética em que um teste foi construído com tantos itens quanto necessário para captar toda a diversidade de tarefas requeridas para o domínio completo da competência considerada, e este foi aplicado a elevado número de alunos com todos os níveis na respectiva competência, diz-se que as proficiências desses alunos definem uma escala de medida da competência. Usualmente a escala é um *continuum*, mas por opções metodológicas varia usualmente entre -3 e 3, limites que podem ser transformados em quaisquer outros valores julgados mais convenientes.

Uma escala só é útil para finalidade de diagnóstico ou de ação pedagógica se os seus diferentes pontos tiverem uma interpretação pedagógica. A interpretação começa com a construção de um mapa de itens, que consiste em associar cada item do teste a um ponto da escala utilizada para medir as proficiências dos alunos. Como mostrado pela curva de característica, a relação entre um item e a escala é probabilística e, portanto, aquele pode ser localizado em pontos diversos desta. Frequentemente toma-se como ponto de locação a proficiência em que a probabilidade de acertar o item seja de 65%. A Figura 2 expressa graficamente de um mapa de itens, mostrando que tanto os alunos quanto os itens estão referenciados a pontos da escala.

Construído o mapa de itens, a interpretação pedagógica considera que alunos com proficiência em um dado valor são capazes de fazer tarefas implícitas nos itens localizados em valores menores. Existem várias nuances para a aplicação desse princípio. A referência clássica é o artigo de Beaton e Allen (1992). Os detalhes podem ser encontrados nas várias interpretações publicadas da escala do Saeb, principalmente em Fontanive, Elliot e Klein (2007). Os materiais de divulgação do Saresp também apresentam a interpretação da escala do Saeb com itens que foram utilizados nesse exame.

Essa interpretação, embora fundamental para orientar intervenções pedagógicas, não explicita se as proficiências obtidas pelo grupo de alunos analisados

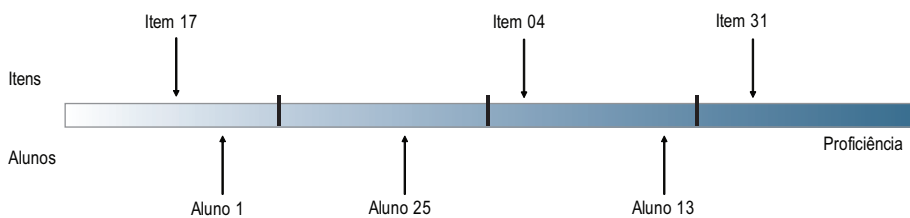
são ou não adequadas, considerando a trajetória escolar dos alunos testados. Isso exige a produção de outra interpretação, usualmente apresentada por meio da escolha de níveis, ou seja, da divisão da escala de medida em intervalos aos quais são atribuídos significados pedagógicos e normativos.

Há muita diversidade de opinião sobre o número e denominação dos níveis que devem ser usados. Neste texto, foram escolhidos quatro níveis com as seguintes denominações: Abaixo do Básico, Básico, Proficiente e Avançado. Essas etiquetas escolhidas para os níveis têm um claro significado normativo. O aluno classificado no nível Proficiente demonstra dominar os conteúdos e habilidades esperados para o seu estágio escolar. Os do nível Avançado dominam a competência de forma especialmente completa, ultrapassando o esperado para o seu estágio escolar. O nível Básico congrega os alunos que demonstram domínio apenas parcial da competência. Finalmente, os alunos do nível Abaixo do Básico mostram domínio rudimentar da competência medida. Na literatura, nomes alternativos, mas equivalentes, são frequentemente utilizados.

A opção por quatro níveis reflete a necessidade de orientar a organização de atividades a serem oferecidas aos alunos após a análise dos dados obtidos em uma avaliação. O uso de níveis permite dar às medidas das proficiências dos alunos uma utilidade pedagógica e não apenas gerencial. Os alunos que estão

Figura 2

Mapa de Itens



Fonte: Pesquisa do autor.

no nível Abaixo do Básico precisam de acompanhamento imediato para que a situação em que se encontram não se cristalice. Aos alunos no nível Proficiente devem ser oferecidas atividades de desafio como olimpíadas, que favorecem a criação de um grupo com desempenho avançado. As atividades adequadas aos alunos do nível Básico exigem análise mais detalhada de sua situação que considere o nível de ensino e as preferências individuais dos alunos. Alguns estão neste nível por entenderem que, em relação à competência, não precisam de desempenho melhor. No entanto, na maioria das vezes, precisam aprimorar seu desempenho, merecendo atividades de consolidação.

O problema que se coloca, entretanto, é a escolha de pontos de corte na escala de proficiência que definem exatamente a localização dos níveis. A forma escolhida para o Saresp é apresentada em seguida.

CONSTRUÇÃO DOS NÍVEIS

Há expressiva produção a partir da proposta de Angoff (1971) para a escolha de níveis de aprendizado em avaliações educacionais de larga escala. Todos os métodos recomendam que a escolha dos pontos de corte utilize, concomitantemente, o mapa de itens e considerações substantivas sobre o significado pedagógico de cada item e a organização do ensino. Embora os dados para este tipo de pesquisa estejam disponíveis, ainda não se organizou no Brasil um esforço de interpretação da escala do Saeb como recomendado por essa literatura, sintetizada por Cizek (2001).

Diante disso, os níveis do Saresp foram estabelecidos de forma empírica e comparativa. Escolheu-se uma distribuição de referência para as proficiências dos alunos e, a partir dessa distribuição, fizeram-se os cortes que definiram os níveis. A adoção de uma distribuição de referência traz implicitamente o reconhecimento de que as proficiências dos alunos de um sistema educacional assumirão valores diversos. No entanto, a variação das proficiências deve se limitar a valores compatíveis com as necessidades de participação social dos alunos. Ou seja, não se pretende, quando se monitora a competência cognitiva de alunos, produzir a igualdade, mas dar a todos a possibilidade de inserção em diferentes formatos. Pessoas distintas

terão competências diversas em níveis mais ou menos altos; todos terão, entretanto, domínio das competências que permitam sua atuação como cidadãos.

DISTRIBUIÇÃO DE REFERÊNCIA²

A distribuição de referência a ser utilizada na definição dos níveis foi escolhida por meio da comparação do desempenho dos alunos brasileiros com o dos alunos de um grupo de países participantes do Pisa.

O Pisa é organizado pela Organização para Cooperação e Desenvolvimento Econômico – OCDE e dele participam os países-membros dessa organização e também países convidados, como o Brasil. Assim, a escolha de países cujos alunos comporiam o grupo de referência para os estudantes brasileiros comporta várias soluções. A comparação dos resultados dos brasileiros apenas com os dos alunos de países pequenos, ricos e homogêneos socioculturalmente não é útil. Em contrapartida, considerar todos os países da OCDE implica incluir aqueles com desempenho baixo. Diante disso, neste texto optou-se por usar, como referência, a média do comportamento dos alunos dos seguintes países: Austrália, Áustria, Bélgica, Canadá, Suíça, Alemanha, Dinamarca, Espanha, Finlândia, França, Inglaterra, Irlanda, Islândia, Itália, Japão, Coreia, Luxemburgo, Holanda, Noruega, Nova Zelândia, Polônia, Portugal, Suécia e Estados Unidos. Esse grupo de alunos será denominado grupo OCDE de alunos. O uso da média de vários países é solução similar à adotada nos relatórios oficiais sobre os resultados do Pisa.

Para a construção da distribuição ideal utilizou-se o Pisa de 2000 para leitura e o de 2003 para matemática. Considerando cada uma dessas situações, calculou-se, para cada percentil das proficiências, a distância em desvios-padrão dos alunos brasileiros e o mesmo percentil para os alunos do grupo OCDE. A seguir, aumentaram-se os percentis da distribuição dos resultados do Saeb no mesmo número de desvios-padrão correspondentes à diferença anteriormente observada. Esse processo equivale a definir como ideal a distribuição de proficiências que seria obtida caso os alunos dos países da OCDE fizessem o teste do Saeb.

A Tabela 1 mostra a organização dos cálculos necessários e explicita os detalhes dessa metodologia. A primeira coluna apresenta simplesmente uma lista de percentis. Para a construção da distribuição de referência foram utilizados todos os percentis e não apenas os apresentados na tabela. A segunda e terceira colunas mostram o valor em pontos do Pisa de cada percentil da distribuição de notas dos alunos brasileiros e dos alunos OCDE. Por exemplo, o percentil de ordem 30 para os brasileiros é 307. Isso significa que 30% dos estudantes avaliados pelo Pisa tiveram nota abaixo de 307. Enquanto isso, o mesmo percentil para os alunos OCDE é 465. Os percentis dos brasileiros são inferiores aos dos alunos OCDE, fato bastante conhecido. Na quarta e quinta colunas registra-se essa diferença, primeiro em termos absolutos e depois em termos de desvios-padrão da distribuição dos alunos brasileiros no Pisa. Ou seja, a última coluna exibe o número de desvios-padrão a aumentar em cada percentil da nota dos alunos brasileiros para que sua distribuição seja igual à dos estudantes do grupo OCDE.

Finalmente, para obter a distribuição de referência, aplicaram-se os valores definidos na última coluna da Tabela 1 às notas dos alunos no Saeb. A Tabela 2 apresenta os resultados dessa operação para os alunos da oitava série no teste de matemática. Nessa tabela, os valores da última coluna são os percentis da distribuição ideal, ou seja, a distribuição que seria observada caso os alunos do grupo de referência fizessem o teste do Saeb.

O Gráfico 1 mostra os resultados dessa operação. É evidente o deslocamento em direção a valores maiores arbitrados no processo de escolha da distribuição ideal.

Obtida a distribuição de referência, os níveis para o Saresp foram estabelecidos conforme descrito a seguir. O ponto de corte que define o nível Proficiente foi fixado no mesmo valor escolhido pelo movimento Todos pela Educação para seu indicador, ponto acima do qual deveriam estar 70% dos alunos no momento em que a educação brasileira pudesse ser considerada de qualidade. Embora a escolha tenha utilizado a experiência dos membros da comissão e as interpretações já feitas pela escala do Saeb, os pon-

Tabela 1
Diferenças entre o Desempenho dos Alunos Brasileiros e Alunos OCDE no Pisa 2003 em Matemática
Brasil e Países Selecionados – 2003

Percentil	Pontuação no Pisa		Diferença entre Alunos OCDE e Brasileiros	
	Alunos Brasileiros	Alunos OCDE	N ^{os} Absolutos	Desvios-Padrão
5	223	361	138	1,35
15	266	417	151	1,47
30	307	465	158	1,54
50	354	514	160	1,57
75	419	576	157	1,53
90	481	626	145	1,41
95	523	654	131	1,29

Fonte: OECD (2005).

Tabela 2
Construção da Distribuição de Referência na Métrica do Saeb para Alunos da 8ª Série do Ensino Fundamental – Matemática
Brasil – 2003

Percentil	Valor dos Percentis	
	Distribuição Atual	Distribuição Ideal
5	170	238
15	192	266
30	223	300
50	242	320
75	278	355
90	316	387
95	341	405

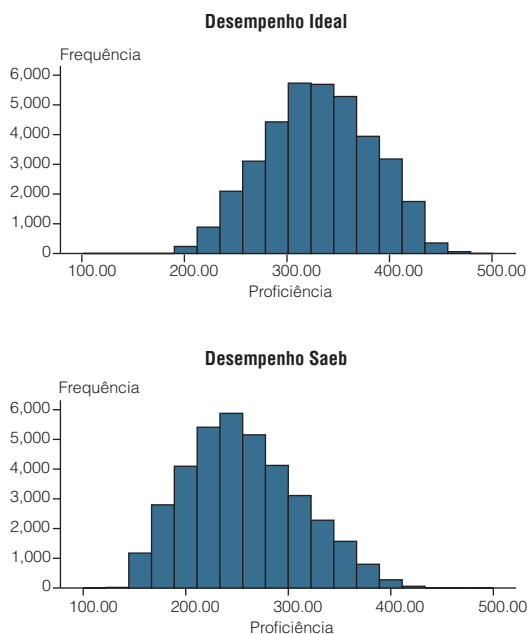
Fonte: MEC. Inep.

tos obtidos são muito próximos do percentil 70 da distribuição de referência.

A definição de apenas um ponto de corte pode ensejar atitudes pouco razoáveis. Por exemplo, os alunos que estiverem muito abaixo desse ponto não receberão atenção nas suas escolas, pois dificilmente alcançarão o nível desejado em prazos curtos. Em compensação, todas as escolas devem ter excelência acadêmica, o que implica que alguns de seus alunos

Gráfico 1

Comparação entre o Desempenho Ideal e o Desempenho dos Alunos da 8ª Série do Ensino Fundamental que Fizeram o Teste de Matemática – Saeb Brasil – 2003



Fonte: MEC.Inep.

tenham desempenho além do esperado. Diante dessas considerações, usuais em muitos outros sistemas de avaliação no mundo, optou-se por criar para o Saresp mais dois pontos de corte. Estes foram escolhidos de forma *ad hoc*, de modo a colocar 25% dos alunos no nível avançado e 25% nível básico. Assim, os pontos de corte correspondem àqueles que definem, na distribuição de referência, intervalos com frequência de alunos de 5%, 25%, 45% e 25%.

Depois de todas estas considerações, os pontos de corte adotados em São Paulo foram definidos conforme o Quadro 1.

O IDESP

O conhecimento do percentual de alunos em cada um dos quatro níveis é suficiente para a análise do desempenho dos estudantes de uma escola e para identificar a necessidade de intervenções pedagógicas de recuperação ou de desafio intelectual. No entanto, para a visão geral de todas as escolas de um sistema é necessária a introdução de um indicador-síntese.

O Índice de Desenvolvimento da Educação Básica – Ideb, introduzido pelo MEC/Inep, cumpriu essa função e, rapidamente, tornou-se a principal forma de análise das políticas educacionais brasileiras para

Quadro 1

Níveis do Sistema de Avaliação da Aprendizagem de São Paulo – Saresp de Desempenho para Leitura e para Matemática, por Série Estado de São Paulo – 2008

Nível do Saresp	Desempenho para Leitura			Desempenho para Matemática		
	4ª Série EF	8ª Série EF	3º Ano EM	4ª Série EF	8ª Série EF	3º Ano EM
Abaixo do básico	< 150	< 200	< 250	< 175	< 225	< 275
Básico	Entre 150 e 200	Entre 200 e 275	Entre 250 e 300	Entre 175 e 225	Entre 225 e 300	Entre 275 e 350
Adequado	Entre 200 e 250	Entre 275 e 325	Entre 300 e 375	Entre 225 e 275	Entre 300 e 350	Entre 350 e 400
Avançado	Acima de 250	Acima de 325	Acima de 375	Acima de 275	Acima de 350	Acima de 400

Fonte: Secretaria de Estado da Educação.

o ensino básico. A divulgação de metas educacionais para cada sistema e cada escola pública do país baseadas no Ideb mudou a maneira como os dados das avaliações nacionais passaram a ser usados. A partir desta iniciativa, os resultados das avaliações tiveram nova visibilidade e importância ampliada.

O Ideb é o produto de dois indicadores. O primeiro reflete o desempenho médio dos alunos na Prova Brasil em um número entre 0 e 10 e o segundo, expresso por um número entre 0 e 1, capta a eficiência com que um sistema ou uma escola consegue diplomar seus alunos em cada uma das etapas do ensino fundamental. Assim sendo, o Ideb é calculado separadamente para os anos iniciais e anos finais do ensino fundamental. Ao compor um indicador através do produto de duas quantidades, o Ideb assume que, em um bom sistema educacional ou escola, uma alta porcentagem dos alunos termina as diferentes etapas do ensino fundamental apresentando, no momento do término, um desempenho cognitivo considerado adequado.

O valor do Ideb aumenta quando um ou os dois indicadores que o compõem aumentam. A melhoria do primeiro termo, que descreve o desempenho, ocorre sempre que há aumento da nota média dos alunos de um sistema ou de uma escola. Uma mesma média, entretanto, pode ser obtida com valores muito diferentes para os percentuais nos diferentes níveis de desempenho. Muitos desses perfis são claramente pouco equitativos pois, para obter médias de desempenho maiores, deixam para trás alunos com desempenho baixo.

A Tabela 3 mostra esta fragilidade do Ideb. Foram selecionadas escolas com nota média em leitura variando de 199 a 201. Como se pode observar, esse valor médio é compatível com perfis pedagógicamente muito diferentes nos níveis.

O Idesp foi construído, por um lado, para manter a ideia básica de agregação dos indicadores de fluxo e desempenho e, por outro lado, para incorporar a dimensão de equidade no indicador de desempenho, ausente no Ideb. Isso foi feito mudando-se a maneira de cálculo do primeiro fator, passando-se a utilizar um indicador baseado nos percentuais de alunos nos diferentes níveis.

Tabela 3

**Distribuição dos Alunos das Escolas com a Mesma Média de Proficiência, por Nível
Brasil – 2003**

Escola	Proficiência Média	Nível (%)			
		Abaixo do Básico	Básico	Proficiente	Avançado
1	199,90	9,1	45,5	45,5	0,0
2	199,28	18,8	30,4	39,1	11,6
3	199,26	16,7	32,3	35,4	15,6
4	200,04	4,3	48,6	37,1	10,0
5	200,39	2,9	44,1	47,1	5,9
6	200,64	13,3	30,0	38,9	17,8
7	199,62	9,4	34,4	40,6	15,6

Fonte: MEC, Inep,

O indicador de desempenho usado no Idesp parte da defasagem média dos alunos da escola em relação ao nível avançado. Para isso, os alunos classificados nos níveis Abaixo do Básico, Básico, Proficiente e Avançado recebem, respectivamente, três, duas, uma e nenhuma defasagem. A defasagem do sistema ou escola é definida como a média ponderada das defasagens de cada nível, tomando-se como peso a porcentagem de alunos em cada nível. Para que a defasagem, como definida anteriormente, fosse similar ao indicador de desempenho usado no Ideb foi necessário transformá-la em um indicador que assume valores entre 0 e 10, sendo atribuído à pior escola o valor 0 e à melhor, escola o valor 10. Isso é obtido pela aplicação da seguinte fórmula:

$$\text{Indicador} = (1 - (\text{Defasagem}/3)) * 10$$

Alguns exemplos apresentados na Tabela 4 ilustram melhor essas definições.

A primeira linha representa uma escola hipotética que tem todos os seus alunos com proficiência abaixo do básico. Sua defasagem é igual a 3 e seu indicador de desempenho, igual a 0. No outro extremo, a quarta linha representa uma escola hipotética que tem todos os seus alunos com proficiência avançada. Sua defasagem é igual a 0 e seu indicador, igual a 10.

Tabela 4

Construção do Indicador Síntese do Desempenho de cada Escola
Brasil – 2003

Distribuição	Nível (%)				Defasagem	Indicador
	Abaixo do Básico (n*3)	Básico (n*2)	Proficiente (n*1)	Avançado (n*0)		
Limite inferior	100	0	0	0	3	0
Escolas públicas Brasil	56	33	10	1	0,81	1,9
Referência	5	25	45	25	1,1	6,3
Limite Superior	0	0	0	100	0	10,0

Fonte: MEC. Inep.

A segunda linha representa a escola pública brasileira típica em matemática na quarta série. Sua defasagem e indicador de qualidade são definidos por:

$$D = (56*3 + 33*2 + 10*1 + 1*0) / 100 = 2,44.$$

$$I = (1 - (2,44/3)) * 10 = 1,90$$

A terceira linha da tabela representa uma escola hipotética com os percentuais de alunos nos diferentes níveis equivalentes à distribuição de referência usada para a construção dos níveis. Sua defasagem e indicador de qualidade são definidos por:

$$D = (5*3 + 25*2 + 45*1 + 25*0) / 100 = 1,10.$$

$$I = (1 - (1,10/3)) * 10 = 6,30$$

O indicador de desempenho será maior que esse valor apenas se os alunos se concentrarem mais nos níveis Proficiente e Avançado do que na distribuição de referência. Isso é possível, mas ocorre em poucas escolas. Assim sendo, atingir o valor 6 neste indicador pode ser considerado como uma meta adequada para um sistema educacional.

O uso da defasagem da escola como medida de seu desempenho assume que os desempenhos dos alunos pertencentes a um mesmo nível são equivalentes. Para aqueles no nível Abaixo do Básico interessa uma melhoria que os leve a outro patamar. Em contrapartida, aprimoramentos para os que já estão no nível Avançado não são captados. Melhorias que apenas mudam o desempenho de um aluno, mantendo-o dentro do mesmo nível, não são consideradas relevantes. Esta característica impede a exclusão de alunos mais fracos como estratégia para o aumento do indicador de desem-

penho e também a consideração de melhorias que não têm impacto significativo, já que os alunos se mantêm no mesmo patamar de aprendizagem. Assim sendo, o Idesp considera que há melhoria no desempenho da escola apenas quando pelo menos um de seus alunos muda de nível.

O indicador de fluxo usado no Idesp é o mesmo do Ideb. Essa medida é obtida para cada etapa do ensino fundamental pela combinação das taxas de aprovação apuradas no Censo Escolar em cada uma das séries. Importante ressaltar que, ao fim de cada ano letivo, os alunos da escola são classificados em três grupos: os aprovados para o ano seguinte, os reprovados e os que abandonaram a escola. Assim, a taxa de aprovação reflete também o abandono. Ao se priorizar a taxa de aprovação para a construção da medida de fluxo assume-se que a boa escola consegue também manter seus alunos.

Por motivos técnicos que ultrapassam o escopo deste texto, usa-se, como indicador do fluxo, a média harmônica das taxas de aprovação, i.e., o inverso da média aritmética dos inversos. Assim, uma escola com as taxas de aprovação nas diferentes séries conforme a Tabela 5 teria como medida de seu fluxo o valor de $1 / (1/1,12) = 0,88$.

Ou seja, se essa escola tivesse o indicador de desempenho igual a 5, por exemplo, seu Idesp seria dado por $5*0,88 = 4,4$.

O Idesp de todas as escolas estaduais paulistas que oferecem o ensino fundamental foi calculado com os dados referentes a 2006 e os resultados para a quarta

série matemática são apresentados o Gráfico 2. Pode-se observar o reduzido número de escolas que superaram o valor 6,0, o qual seria obtido se a distribuição das proficiências dos alunos da escola fosse a mesma da de referência e o indicador de fluxo igual a 0,96, novamente as indicações usadas no Ideb.

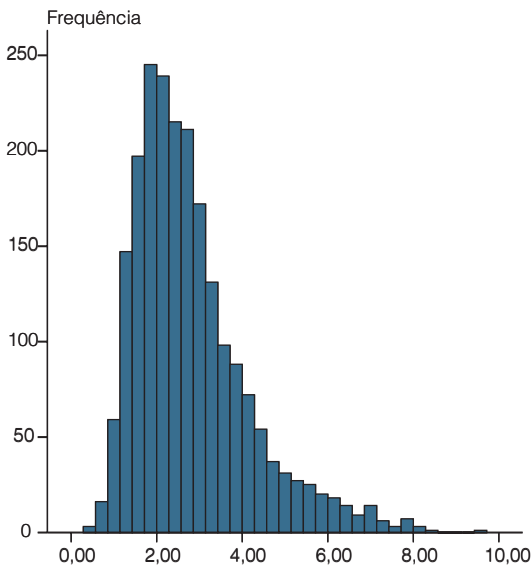
Tabela 5
Construção da Medida de Fluxo

Série	Taxa de Aprovação (%)	Inverso
1	91	1,10
2	93	1,08
3	87	1,15
4	85	1,18
Média		1,12

Fonte: Elaboração do autor.

Gráfico 2

Idesp das Escolas Estaduais Paulistas da 4ª Série
do Ensino Fundamental
Estado de São Paulo – 2006



Fonte: Secretaria de Estado da Educação.

DISCUSSÃO

A Secretaria de Estado da Educação de São Paulo, ao decidir usar a escala do Saeb para expressar o aprendizado dos alunos das escolas estaduais, trouxe para a discussão de suas políticas educacionais vários temas. Este artigo trata primeiramente da solução encontrada para a divulgação dos resultados. Optou-se por usar níveis de desempenho em vez de médias. Esta decisão simples tem, entretanto, uma consequência pedagógica importante ao enfatizar que a avaliação é feita para servir aos alunos. Este formato coloca à disposição de cada escola uma categorização de seus estudantes, convidando-a a promover as atividades que possam atendê-los de maneira mais focada. Usualmente duas atividades são necessárias. Para os alunos classificados no nível Abaixo do Básico correspondem à recuperação, já que não aprenderam minimamente o que deles se esperava. Em contrapartida, toda escola deve ter também alunos com nível excelente, o que só pode ser obtido com o oferecimento de atividades escolares com algum grau de desafio cognitivo para aqueles que já dominam a competência em um nível adequado.

Além disso, considerando os nomes escolhidos para a designação dos níveis produziu-se uma interpretação normativa dos resultados. Claramente os alunos classificados no nível Abaixo do Básico estão em situação de risco, necessitando de atenção imediata.

A introdução de níveis possibilitou ainda a incorporação da equidade no indicador de nível dos resultados. Isso levou à proposição de um indicador global de desempenho que é sensível a estratégias pouco equitativas, como deixar muitos alunos para trás, concentrando os esforços pedagógicos em alguns poucos, o que, se por um lado aumenta a média da escola, por outro constitui uma postura indesejável. A solução encontrada concilia a necessidade de aumento do desempenho sem permitir que as diferenças aumentem.

Definido um indicador de desempenho global das escolas do sistema educacional, vários outros estudos e políticas tornaram-se necessários. Embora ultrapasse o escopo deste texto, deve-se mencionar que

foram criadas metas para o Idesp de cada escola e um sistema de bônus financeiro associado ao cumprimento destas metas.

Finalmente, importante mencionar que a opção pela divulgação de resultados por meio de níveis, definidos por pontos de corte criteriosamente escolhidos e pedagogicamente desafiantes e substantivamente relevantes, permite o acompanhamento da qualidade da educação paulista e brasileira de forma simples e

sólida. Este é um passo fundamental para que o sistema melhore, atendendo aos direitos educacionais de seus alunos.

Isso só será obtido se os resultados escolares, registrados pelo Idesp, forem objeto de estudo detalhado, de forma a identificar, nas condições reais do Estado de São Paulo, como é possível melhorar o desempenho dos alunos. Ou seja, a solução virá na lenta incorporação das melhores práticas por cada unidade escolar.

Notas

Este artigo resulta principalmente de trabalho realizado com Naercio Menezes Filho e Rafael Terra para a Secretaria de Estado da Educação de São Paulo. O autor agradece ainda a Priscilla de Albuquerque Tavares pelas discussões nas etapas finais de definição do Idesp. Finalmente registra e agradece o apoio da Professora Maria Helena Guimarães Castro pelo desafio colocado e a sempre presente exigência por clareza

za na comunicação, garantia da qualidade e introdução da equidade na proposição do indicador-síntese para as políticas educacionais do Estado de São Paulo.

1. No original: “To monitor the outcomes of education systems in terms of student achievement, within a common international framework”.

2. Esta seção é baseada em texto produzido pelo autor a ser publicado como capítulo de livro organizado por Fernando Veloso et al. (2009).

Referências Bibliográficas

ANGOFF, W.H. Scales, norms, and equivalent scores. In: THORNDIKE, R.L. (Ed.). *Educational measurement*. Washington, DC: American Council on Education, 1971. p. 508-600.

BARROS, R.P.; HENRIQUES, R.; MENDONÇA, R. *Pelo fim das décadas perdidas: Educação e desenvolvimento sustentado no Brasil*. Rio de Janeiro: Ipea, 2002. Disponível em: <<http://www.jubamaravilha.hpg.com.br/abril/educDesenvolvBR2002.pdf>>. Acesso em: 12 mar. 2009.

BEATON, A.; ALLEN, N. Interpreting Scales Through Scale Anchoring. *Journal of Educational and Behavioral Statistics*, v. 17, n. 2, p. 191-204, 1992.

CIZEK, G.J. *Setting performance standards: Concepts, methods, and perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates, 2001.

FONTANIVE, N.S.; ELLIOT, L.G.; KLEIN, R. Os desafios de apresentação dos resultados da avaliação de sistemas escolares a diferentes públicos. *REICE – Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, v. 5, n. 2e., 2007. Disponível em: <<http://www.rinace.org/reice>>. Acesso em: 12 mar. 2009.

HAMBLETON, R.K.; SWAMINATHAN, H.; ROGERS, H.J. *Fundamentals of item response theory*. North Carolina: Sage, 1991.

LOURENÇO FILHO, M.B. Redução da taxa de analfabetismo no Brasil de 1900 a 1960, descrição e análise. *Revista Brasileira de Estudos Pedagógicos*, Rio de Janeiro, v. 44, n. 100, p. 250-272, out./dez. 1965.

MENEZES FILHO, N.; SOARES, J.F.; TERRA, R. *Definição de metas de desempenho para as escolas estaduais paulistas de ensino básico*. São Paulo, 2008. Nota técnica submetida à Secretaria de Estado da Educação de São Paulo.

OECD – ORGANIZATION FOR ECONOMIC COOPERATION AND DEVELOPMENT. *Pisa 2003 Technical Report*. Paris: OECD, 2005.

RIBEIRO, S.C. A pedagogia da repetência. *Estudos Avançados*, São Paulo, v. 5, n. 12, ago. 1991.

SOARES, J.F.; MAROTTA, L. Desigualdades no sistema de ensino fundamental brasileiro. In: VELOSO, F. et al. (Org.). *Educação básica no Brasil*. Rio de Janeiro: Campus/Elsevier, 2009. p. 73-91. No prelo.

VELOSO, F. et al. (Org.). *Educação básica no Brasil*. Rio de Janeiro: Campus/Elsevier, 2009. No prelo.

JOSÉ FRANCISCO SOARES

Doutor em Estatística, com pós-doutorado em Educação. Professor titular do Departamento de Ciências Aplicadas à Educação da FAE – UFMG. Coordenador do Grupo de Avaliação e Medidas Educacionais, Belo Horizonte-MG, Brasil.

(francisco-soares@ufmg.br)

Artigo recebido em 16 de junho de 2009.

Aprovado em 10 de agosto de 2009.

Como citar o artigo:

SOARES, J.F. Índice de desenvolvimento da Educação de São Paulo – Idesp: bases metodológicas. *São Paulo em Perspectiva*, São Paulo, Fundação Seade, v. 23, n. 1, p. 29-41, jan./jun. 2009. Disponível em: <<http://www.seade.gov.br>>; <www.scielo.br>.

Acesso em: