

Nigel Brooke
Wagner Silveira Rezende

Os Dilemas da Gestão Escolar

FINO TRACO
FT
EDITORA

Fino Traço Editora Ltda.

© Nigel Brooke, Wagner Silveira Rezende

Este livro ou parte dele não pode ser reproduzido por qualquer meio sem a autorização da editora.

As ideias contidas neste livro são de responsabilidade de seu autor e não expressam necessariamente a posição da editora.

**CIP-BRASIL. CATALOGAÇÃO NA PUBLICAÇÃO SINDICATO
NACIONAL DOS EDITORES DE LIVROS, RJ**

B888d

Brooke, Nigel

Os dilemas da gestão escolar / Nigel Brooke, Wagner Silveira Rezende.
- 1. ed.- Belo Horizonte [MG]: Fino Traço, 2020.

304 p.; 21 cm.

Inclui bibliografia

ISBN: 978-65-9901-400-0

1. Escolas-Organização e administração. 2. Formação do Diretor Escolar.
I. Brooke, Nigel. II. Rezende, Wagner Silveira. III. Título.

20-64650

CDD:371.2

CDU: 37.091



FINO TRACO EDITORA LTDA

www.finotracoeditora.com.br

CAPÍTULO 6

Informação para a gestão escolar

Este capítulo tem como objetivo apresentar argumentos que possam auxiliar o diretor (em especial, mas não somente) a entender melhor as características, as virtudes e os limites das avaliações externas em larga escala¹, particularmente, no que diz respeito ao seu uso como um instrumento importante e útil para a gestão educacional. Para tanto, é preciso superar a resistência que a avaliação sofre e compreender como a interpretação correta dos dados é um aliado para o planejamento da escola e das redes de ensino.

Há alguns anos, no âmbito do Avalia BH, o programa de avaliação da rede municipal de Belo Horizonte, uma equipe de pesquisadores do CAEd/UFJF foi convidada a realizar uma pesquisa para melhorar os materiais de divulgação dos resultados. A proposta envolvia a pesquisa com uma escola: localizada em um bairro considerado “difícil” e com alunos oriundos de famílias de baixa renda, a escola havia alcançado um resultado “surpreendente” nas avaliações do Avalia BH, pois dela era esperado muito menos. Com o intuito de descobrir o que levava a instituição a alcançar resultados além do esperado, foi desenhada uma pesquisa na tradição das pesquisas em eficácia escolar de imersão na cultura da escola.

A conclusão da pesquisa foi, entre outras coisas, reconhecer que o papel da gestão havia sido o principal motivo para os bons resultados da escola: uma gestora com convicção sobre seu papel à frente da escola, capaz de organizar o trabalho dos professores em torno de objetivos comuns, exigente com

1. De agora em diante, quando nos referirmos à avaliação externa, estamos falando da avaliação em larga escala. Apesar de serem características distintas, no cenário brasileiro, a avaliação externa é, regra, em larga escala. Com isso, queremos fazer referência aos sistemas de avaliação, tanto o nacional (Saeb), quanto aqueles que existem em estados e municípios.

respeito aos horários e normas da escola, exigente com a aprendizagem dos alunos, capaz de solucionar conflitos entre os alunos e entre os professores, e disposta a dialogar com a secretaria de educação. Ao falar dos bons resultados da escola, a diretora narra uma história. Como professora de matemática, ao assumir a gestão, decidiu que enfrentaria o problema que conhecia no 8º ano (antiga 7ª série): muitos alunos apresentavam dificuldades com álgebra e eram incapazes de solucionar problemas envolvendo equações de primeiro grau. Como professora, ela chegou a acreditar que álgebra era mesmo difícil, que o problema estaria na própria natureza do conhecimento.

Para enfrentar um problema que ela conhecia tão de perto, decidiu promover uma ação de reforço no contraturno, focada no ensino de resolução de problemas com equações de primeiro grau. Ao final do primeiro ano em que a ação foi implementada, já se viam os resultados: a reprovação em matemática diminuiu muito. Entretanto, no começo do ano seguinte os alunos chegaram novamente com as mesmas dificuldades. Isso significava que o reforço, para continuar a ter efeitos, deveria se tornar uma ação perene e contínua. O problema não estava resolvido.

Depois de relutar em aceitar o sistema de avaliação de seu município como fonte de informações úteis para sua função, a diretora começou a responder pelos resultados de sua escola. Ao consultar e analisar os resultados, ela percebeu que os alunos do 8º ano não estavam com problemas só em equações de primeiro grau: a questão era muito mais grave, visto que eles se mostravam incapazes de resolver operações simples, habilidades que deveriam ter sido desenvolvidas ainda no primeiro ciclo do ensino fundamental. Com base nessa informação, o reforço no contraturno do 8º ano foi trocado por ações focados na alfabetização matemática.

A história dessa diretora ilustra bem o que se espera do uso de resultados dos sistemas de avaliação pela gestão da escola: a capacidade de, entendendo bem os resultados, usá-los como suporte para o desenho de ações destinadas a solucionar os problemas de aprendizagem de seus alunos, diagnosticados, inclusive, pela própria avaliação.

A resistência à avaliação

Talvez seja a avaliação externa um dos temas mais polêmicos e espinhosos em educação. Ela polariza opiniões: nos casos mais radicais, de um lado, há quem defenda a avaliação como a solução de todos os problemas educacionais, ao passo que, do outro, estão aqueles que a consideram o inimigo público número um da educação. Entre os polos, as opiniões divergem substantivamente, oscilando entre aqueles que acreditam que, mesmo com limites, as avaliações podem ser bem usadas e aqueles que, no máximo, a toleram, como um mal necessário.

A avaliação, nesse cenário dividido, é personificada (REZENDE, 2014). A ela são atribuídas características humanas. Nas críticas mais duras, a avaliação ganha volição, capacidade de pensar, decidir e agir. A avaliação torna-se uma pessoa, de caráter duvidoso. Ela faz comparações indevidas, hierarquiza escolas, rotula alunos, julga professores, contamina a escola com a perversa e desumana lógica do mercado. A avaliação transforma tudo em números, frios, superficiais, desconectados com a escola e sua lógica própria. A avaliação é uma ameaça.

Essa personalização da avaliação desconsidera sua natureza. A avaliação é um instrumento. Como tal, está sujeita aos desígnios de seus usuários. Suas consequências dependem de como é usada. Fazer considerações indevidas a partir das informações oriundas das avaliações é tão possível quanto usar as mesmas informações para diagnosticar problemas graves de aprendizagem e a iniquidade nas escolas (como desigualdades raciais, por exemplo, em BASSO *et al.*, 2012). Dificilmente, alguém seria contra esse último uso. O uso da avaliação, passível de crítica ou elogio, depende da ação de atores educacionais. E isso está diretamente vinculado a como esses atores compreendem a avaliação, conhecem seus limites e suas potencialidades e de como foram inseridos nesse processo. Eis aqui um grande problema.

A avaliação não é um tema central para o currículo das licenciaturas e do curso de Pedagogia. Contudo, são justamente os egressos desses cursos que assumirão posições na estrutura educacional relacionadas ao uso das avaliações. No âmbito da formação, a avaliação costuma ser apresentada como uma disciplina secundária, de menor importância, empurrada na grade para

o último período do curso de Pedagogia, nos horários que menos favorecem a participação dos alunos. Isso quando existe uma disciplina obrigatória para tratar desse assunto. No último período do curso de formação, o aluno já desenvolveu um argumento, pronto, copiado e irrefletido, de que a avaliação é um problema, um paladino a serviço do capitalismo e do neoliberalismo contemporâneos (como vemos, por exemplo, em FREITAS, 2012, 2013a, 2013b). Sua linguagem, ancorada em números, seria uma prova de sua vocação mercadológica. A própria lógica de organização e discurso de uma disciplina de avaliação educacional é um contraponto ao discurso que os alunos aprenderam a internalizar nas demais disciplinas do curso de Pedagogia. A avaliação é estrangeira, estranha, exógena. É como se não coubesse na formação do pedagogo e do professor. No caso das licenciaturas, em geral, sequer existe uma disciplina obrigatória sobre avaliação educacional.

Sem formação adequada sobre o tema, esses recém-formados ingressam na vida profissional e um dos primeiros desafios que lhes é apresentado na escola ou em algum cargo da rede é justamente o trabalho com os dados educacionais oriundos da avaliação. Geralmente, esse primeiro contato já vem em forma de cobrança: os resultados da escola precisam ser melhorados. A responsabilidade é de todos na escola. Algo precisa ser feito, justamente, por aqueles que desconhecem a avaliação.

Às lacunas de formação soma-se a linguagem mobilizada pela avaliação externa. Ancorada em métodos estatísticos e em fórmulas para a produção de medidas, a avaliação é entendida pelos profissionais da educação, muitas vezes, como uma espécie de caixa-preta, uma linguagem cifrada, compreendida apenas por poucos iniciados, normalmente, pertencentes à área da matemática e disciplinas conexas, e, portanto, sem compreensão adequada da educação como fenômeno humano e social. São dois mundos, duas linguagens, dois tipos de profissionais, mais, dois tipos de pessoas, antagônicos, contraditórios. A defesa por um lado ou outro, inclusive, é carregada por julgamentos morais. Seu caráter está sob suspeita se você acredita que a avaliação é um instrumento importante para a educação.

Contribui para esse cenário o histórico de implementação da avaliação como política pública no cenário nacional. Alçada a uma posição importante

no contexto de defesa das reformas da educação em fins dos anos 1980 e ao longo dos anos 1990, a avaliação é vista como um produto direto da influência de organismos internacionais a serviço do mercado mundial nos países em desenvolvimento. A defesa da qualidade da educação como preocupação central das políticas educacionais na América Latina (SCHWARTZMAN; COX, 2010) trouxe à cena a importância de mensurar e monitorar essa qualidade como a garantia de um direito. O vínculo da avaliação com a proposta de reforma educacional, ancorada na defesa da eficiência do Estado no que diz respeito à prestação desse serviço, atou o instrumento aos propósitos do gerencialismo. Assim, ato contínuo, os críticos do gerencialismo tornaram-se críticos da avaliação.

A forma como políticas públicas são implementadas no Brasil aguçou ainda mais a leitura pejorativa acerca da avaliação. O papel que o Estado sempre desempenhou, entre nós, para o desenho, implementação e condução de políticas é caracterizado por um diálogo escasso com a sociedade civil, em regra, agente passivo e expectador desse processo, sem, efetivamente, participar das decisões. A via prussiana, nossa tradição de implementar políticas de cima para baixo, contribui para a visão de que a avaliação externa é uma política imposta, conduzida sem a participação dos atores educacionais. A avaliação surge no cenário nacional de maneira abrupta, sem uma discussão adequada (e sem uma preparação adequada) com os atores que por ela seriam diretamente afetados.

Completa essa conjunção de forças a serviço da resistência à avaliação educacional a própria forma como a avaliação é tratada pelas escolas, e não estamos falando da avaliação externa, mas da avaliação interna, aquela conduzida pelos professores. Aprendemos, na escola, a não gostar de avaliação, a olhar para os instrumentos avaliativos com desconfiança. Seu uso como instrumento histórico de punição no ambiente escolar é um dos fatores centrais para o desenvolvimento desse olhar, mas é mais do que isso. Nós não gostamos de ser avaliados, mesmo fora da escola. Ninguém gosta. O próprio termo, avaliar, é carregado de um sentido pejorativo. A avaliação é tratada como sinônimo de julgamento. A pesquisa educacional seria beneficiada com a investigação dos fundamentos psicológicos da resistência à própria ideia de

avaliar. Professores e diretores não gostam de ser avaliados muito em virtude de terem sido alunos. Como alunos, aprendemos a não confiar na avaliação e a não ver nela nenhum fundamento pedagógico. Como professores, esses alunos tendem a reproduzir essa concepção, dessa vez, na outra posição da relação de ensino e aprendizagem.

O golpe final é dado pelas políticas de responsabilização, possivelmente, o principal alvo das críticas direcionadas à avaliação (como em FREITAS, 2013b). Atrelar os resultados da avaliação a bonificações para as escolas de melhores resultados e, como contrapartida, punir as de pior resultado (com cobranças excessivas, com a ausência de destinação de recursos, com a troca de gestores etc.), é visto como um dos maiores problemas da política de avaliação, aquele que mais ofende a escola. Ximenes (2012), por exemplo, afirma que a responsabilização a partir do fracasso nas avaliações tem como objetivo alterar a concepção de responsabilidade relacionada ao agente político e coletivo para uma responsabilidade do funcionário, ou seja, dos professores e diretores de escola.

Não nos sentimos confortáveis com a ideia de avaliação; aprendemos na escola que a avaliação é um problema; recebemos uma política de avaliação externa da qual não participamos do processo de construção e implementação; essa política é vista como fruto da expansão do neoliberalismo que estende sua influência lesiva aos bancos da escola; é caracterizada por uma linguagem fria e cifrada da matemática; é personificada como de caráter duvidoso que ameaça a lógica virtuosa que a educação deveria ter; incentiva a competição e não a cooperação entre os pares, e é atrelada a uma política de responsabilização polêmica. Não é difícil compreender os motivos pelos quais a resistência à avaliação dominou os primeiros anos após a criação do sistema nacional, nos anos 1990, e ao longo da expansão dos sistemas próprios de avaliação, em meados dos anos 2000.

Essa resistência, disseminada entre sindicatos, professores universitários, professores da educação básica, diretores de escolas, coordenadores pedagógicos e mesmo entre os técnicos das secretarias (principalmente, no início do processo), não foi capaz de frear a expansão e o desenvolvimento dos sistemas de avaliação no país. Ao contrário da descontinuidade que

caracteriza boa parte das políticas públicas educacionais (DOURADO, 2007), a avaliação externa é uma política educacional de surpreendente longevidade, mesmo com todo o cenário de críticas narrado anteriormente. Ela foi capaz de atravessar governos de diferentes orientações ideológicas (FRANCO; ALVES; BONAMINO, 2007), não apenas resistindo às mudanças governamentais, mas, mais do que isso, reforçando sua centralidade diante de cada uma delas. Mesmo persistindo as críticas à avaliação, parte significativa dos atores educacionais alterou sua percepção sobre ela ao longo dos anos. Se ainda é possível encontrar, por exemplo, muitos professores da educação básica contrários aos sistemas de avaliação, não é difícil encontrar aqueles que não apenas compreendem a avaliação como um instrumento a serviço da qualidade educacional, mas fazem uso de suas informações como parte de suas rotinas de planejamento.

No caso da gestão, essa mudança tornou-se ainda mais perceptível. Diretores de escola dificilmente são indiferentes aos resultados das avaliações externas. Se a cobrança por melhores resultados, por vezes desmedida por parte das secretarias, ainda é um ponto de conflito para os gestores escolares, eles passaram a tratar a avaliação, em regra, como um aliado para a melhoria da escola. O mesmo pode ser dito para os profissionais que ocupam posições na gestão das redes de ensino. O foco principal da crítica permanece sendo a academia: a maior resistência em relação à avaliação educacional externa está nas universidades.

Mas, afinal, por que a avaliação foi capaz de se desenvolver em meio a tantas críticas, resistências e trocas de governo? Antes de tudo, é preciso chamar a atenção para um ponto. Todas essas críticas encontram motivos e razões para existir. Com isso, queremos dizer que, ao analisá-las, não estamos procurando culpados. O cenário em torno da avaliação, como vimos, não favoreceu sua compreensão e sua aceitação. Tendemos a rechaçar aquilo que não compreendemos. Reconhecer isso, contudo, não quer dizer que seja pertinente a crítica que nega a importância da avaliação para as escolas e para as redes. O fato de entendermos a origem das críticas não significa que concordamos com elas. Entendemos a avaliação como uma ferramenta fundamental para a garantia da qualidade da educação.

Natureza e fundamentos da avaliação educacional

Dizer que a avaliação é um instrumento significa reconhecer que ela não possui um fim si mesma. Muitas vezes, a crítica recai sobre esse aspecto. É como se o interesse por trás da avaliação fosse o de, pura e simplesmente, criar rótulos. A avaliação é um instrumento no sentido técnico (uma ferramenta, construída com base em metodologia específica, encarnada em testes e questionários), mas também no teleológico: ela serve a um fim diferente de sua própria existência. A avaliação é base para que decisões sejam tomadas. Ela fundamenta decisões, em todas as instâncias da educação, das secretarias às salas de aula. A avaliação dá suporte ao desenho de políticas educacionais e ações pedagógicas e de gestão nas escolas, orientando o processo decisório. Tomar decisões, em qualquer instância da vida, tem custos. Não é um processo simples, pois envolve escolhas, cálculos, caminhos possíveis, dos quais uns serão descartados em privilégio de outros, sem que haja certeza absoluta dos resultados. Em ambientes coletivos, tomar decisões implica ouvir opiniões diferentes e compreender que a escolha contará com o apoio de alguns e com a oposição de outros. Nesse processo, muita energia é necessária e os custos políticos são imprevisíveis.

A avaliação assume a posição de suporte a decisões por conta de suas características como instrumento técnico. Ela é construída com base no desenvolvimento das ciências a ela conectadas (psicometria, pedagogia, sociologia, estatística, matemática, metodologia de pesquisa etc.). Um processo de avaliação externa em larga escala não é pensado e construído de modo trivial. Há amplo conjunto de aspectos que devem ser observados. A produção da medida de proficiência, por exemplo, é informada pelo que há de mais sofisticado na teoria para a mensuração de traços latentes (a Teoria da Resposta ao Item – TRI), como é o caso da aprendizagem. A medida produzida, inclusive, vem acompanhada da estimativa do próprio erro (isso não quer dizer que a medida está errada, da maneira como tendemos a entender essa ideia de erro; toda estimativa carrega em si um erro; o que uma metodologia como a TRI permite fazer é calcular a dimensão desse erro). Os atores envolvidos com o desenvolvimento de um sistema de avaliação sabem que o instrumento possui limites e são capazes de estimar o quanto

esses limites influenciam o resultado. A avaliação é base para a tomada de decisões por oferecer informações qualificadas aos atores educacionais.

Em cenários como a escola, compostos pela heterogeneidade de atores e opiniões, tomar uma decisão que afeta a todos a partir de uma única visão pode levar a problemas de aceitação. No caso das escolas públicas, esse cenário torna-se ainda mais complexo em virtude da relação entre o número de problemas e os recursos (não apenas financeiros) disponíveis para solucioná-los. De maneira geral, há muito mais problemas do que recursos, de modo que a decisão é sempre trágica: a escolha pelo enfrentamento de um problema pressupõe o não enfrentamento de outros. Para tanto, é preciso estabelecer prioridades. Um gestor escolar pode decidir por investir no ensino médio, destinando esforços, energia e recursos para essa etapa de escolaridade, tomada por ele como prioritária. O ciclo de alfabetização, nesse exemplo, ficaria descoberto de ações mais efetivas. O grupo de professores alfabetizadores, provavelmente, discordará dessa escolha do gestor, afirmando que os alunos com problemas de alfabetização, possivelmente, sequer chegarão ao ensino médio, ao passo que os professores desta última etapa tenderão a apoiar a decisão da gestão. Para lidar com escolhas dessa natureza, rotineiras nas escolas públicas, é melhor que o gestor tenha bons motivos para justificar sua decisão. Se ele, por exemplo, apresentar evidências de que os alunos do ensino médio estão com problemas graves de aprendizagem e que tais problemas não são oriundos de uma formação lacunar no ciclo de alfabetização, sua decisão ganhará força diante de toda a escola, em especial, daqueles que, a princípio, eram contrários a ela. Essas evidências são propiciadas por informações qualificadas. A avaliação externa é uma importante fonte para isso. Decisões justificadas, com base em evidências, são mais convincentes, mais legítimas. A avaliação está a serviço de decisões legítimas.

Produzir informações para a tomada de decisão pode parecer um requisito óbvio. No entanto, muitas vezes, decisões na escola são tomadas com base em outros elementos, como a autoridade dada pela função (“estou tomando essa decisão por ser eu o gestor e cabe a mim escolher”) e a antiguidade na profissão (“sou professor há mais de 30 anos e diretor há mais de 20, sei exatamente o que estou fazendo”). A função não faz ninguém

um melhor tomador de decisão, ao passo que o tempo de profissão não é, necessariamente, sinônimo de capacidade decisória. As escolas ainda carecem de transformar o uso de evidências em um elemento central de suas rotinas de trabalho.

O fato é que as avaliações são fontes valiosas de informações qualificadas sobre as redes e as escolas. E informações qualificadas permitem a tomada de decisões mais legítimas. Produzir informações sobre a aprendizagem e sobre outros aspectos ligados ao fenômeno educacional é o fundamento pedagógico da avaliação. A partir dessas informações, o processo de ensino e aprendizagem pode ser melhor compreendido e tudo aquilo que se associa ao aprendizado dos alunos pode ser pesquisado, identificado, analisado e mapeado. Os instrumentos, nesse sentido, não avaliam apenas os alunos, mas todo o processo educacional escolar e os efeitos das políticas educacionais. Isso permite que a educação, representadas por todos os atores educacionais, pense e reflita sobre si mesma. Toda avaliação é capaz de desencadear um processo de autoavaliação.

Além do fundamento pedagógico, há, pelo menos, mais dois importantes fundamentos da avaliação educacional: o político e o jurídico. Os sistemas de avaliação, como dissemos, produzem e disponibilizam informações qualificadas para a tomada de decisões. Essas informações, os resultados da avaliação, estão relacionadas com a qualidade da educação. Por mais que seja tarefa árdua definir com precisão o que é qualidade educacional, em virtude de se tratar de um conjunto amplo de fatores, não há dúvidas de que a aprendizagem dos alunos é uma das dimensões que compõem o que chamamos de qualidade. O suporte para a tomada de decisão propiciado pela avaliação não diz respeito, contudo, somente aos atores escolares e educacionais, como secretários de educação, suas equipes, gestores escolares e professores. Esse é um processo que mobiliza, ou, pelo menos, num cenário de democracia ativa, deveria mobilizar, toda a sociedade.

A publicidade dos resultados da avaliação, com dados sendo divulgados por escola, caracterizam o que Bonamino e Sousa (2012) chamaram de avaliações de segunda geração, ou seja, aquelas que tornam públicos os resultados de todas as escolas participantes, levando, desta forma, a um grau brando de responsabilização: pais e responsáveis pelos alunos têm acesso

aos resultados das escolas, passam a selecionar escolas com base nessas informações e pressionam as escolas para que melhorem os seus próprios resultados. Mesmo que, na prática, a maioria dos pais e responsáveis não aja dessa maneira, o pressuposto é esse: a avaliação monitora a prestação de um serviço por parte do Estado e, através dos resultados, pressiona o poder público por melhorias. Com base no ideal democrático, isso significa que o Estado, ao avaliar as redes públicas de ensino e tornar públicas essas informações, presta contas à sociedade que, por meio dos impostos, custeia a prestação desse serviço. A avaliação é um dos mecanismos a serviço do funcionamento da democracia, por meio do qual a sociedade recebe informações sobre a qualidade de um serviço que o poder público tem o dever de prestar. Esse é o fundamento político da avaliação.

Entretanto, mais do que um serviço público, a educação é um direito. O artigo 205 da Constituição Federal de 1988 garante a educação como um direito de todos e como um dever do Estado (BRASIL, 1988). O constituinte não precisou adjetivar o termo educação: evidentemente, o que está sendo garantido não é qualquer tipo de educação ou uma educação de qualquer tipo; trata-se de uma educação de qualidade. Novamente, o termo qualidade carece de definição e a legislação infraconstitucional é que deveria ser capaz de definir, com clareza, o que é a qualidade da educação. Esse tema não é simples e está longe de estar resolvido. Se a educação é um direito, cabe ao Direito definir quais são os parâmetros de seu atendimento, para que possa assegurar sua observância e sua garantia. De todo modo, a despeito dessa discussão, ainda por ser aprofundada no cenário nacional, o artigo 205 é o fundamento constitucional da avaliação: se a educação de qualidade é um direito de todos e um dever do Estado, o monitoramento da garantia desse direito depende de instrumentos capazes de mensurar a referida qualidade, sob pena de não sabermos se o direito está sendo garantido ou não.

Soma-se a isso a previsão da legislação infraconstitucional. A Lei de Diretrizes e Bases da Educação (Lei 9394/96), em seu artigo 9º, possui dois incisos que dão fundamento à avaliação educacional: o inciso V prevê a necessidade da avaliação implicitamente, ao afirmar que cabe à União “coletar, analisar e disseminar informações sobre a educação”; e o inciso VI o faz de maneira explícita, ao “assegurar processo nacional de avaliação do rendimento

escolar no ensino fundamental, médio e superior, em colaboração com os sistemas de ensino, objetivando a definição de prioridades e a melhoria da qualidade do ensino". A norma do inciso VI não poderia ser mais clara: estabelece a avaliação como um dever da União e define seus objetivos – definição de prioridades e melhoria da qualidade (BRASIL, 1996).

Construir o argumento de que a avaliação é um direito carrega a dificuldade de enfrentar o histórico narrado no primeiro tópico deste capítulo. A avaliação é vista como punição, como obrigação, como dever, mas não como direito. Contudo, os alunos têm direito à avaliação, assim como os professores, os gestores e todos aqueles envolvidos com a educação. Todos eles compõem a sociedade e, como tal, contribuem para o financiamento da prestação desse serviço. A educação de qualidade é um interesse de todos. Se a avaliação está a serviço da garantia da qualidade da educação, ela também deveria ser.

Acreditamos que a exposição dos fundamentos da avaliação sirva ao propósito de afastar algumas das razões geradoras da resistência contra ela. A avaliação não é uma imposição governamental pura e simples. Desde a Constituição de 1988, ela é um mecanismo necessário, a serviço da garantia de um direito. Deveríamos lutar pela avaliação educacional nesse sentido, se defendemos que a qualidade da educação é um direito de todos. A partir de 1996, ela adquire o sentido jurídico de obrigação, vinculando o poder público à sua execução. Isso não quer dizer que o processo de implementação dos sistemas de avaliação não tenha tido problemas. Conforme narramos anteriormente, a implementação, em regra, não foi discutida com os atores educacionais. O desconhecimento acerca dos objetivos, limites e possibilidades da política alimentou a resistência. Apesar disso, com base nos fundamentos apresentados, não é possível dizer que a avaliação é apenas uma imposição de governos mal-intencionados.

Os fundamentos também não afastam críticas de outra natureza. De fato, principalmente nos primeiros anos, as políticas de avaliação se valiam de uma linguagem de difícil compreensão. O uso de teorias estatísticas sofisticadas para a produção das medidas não encontrou, inicialmente, uma linguagem correspondente para os não especialistas. Com o tempo, à medida que a avaliação foi se consolidando como política pública e seus

resultados passaram a ser mais conhecidos pelos atores educacionais, a pesquisa em avaliação voltou sua atenção para outros aspectos, além do aprimoramento da produção da medida de desempenho: as formas de aplicação começaram a se modificar, com a capacitação dos aplicadores, uso de aplicadores externos à escola, capacitados e remunerados para tanto; os cuidados com a produção de itens e os critérios para sua validação se tornaram mais rígidos e exigentes, melhorando a qualidade dos testes; os mecanismos de divulgação de resultados foram ampliados e aprimorados, buscando levar as informações para os mais variados públicos, do secretário de educação ao professor em sala de aula.

Em relação a esse último ponto, a linguagem tornou-se um elemento particularmente importante. Não apenas a escolha dos termos, mas também a forma de apresentação da informação passou a ser cuidadosamente pensada. A estatística ganhou o auxílio da comunicação, do design instrucional e dos especialistas em linguagem para apresentar a informação da maneira mais fiel e compreensível possível, observando a diferença de interesse e de função de seus destinatários. A maneira como um resultado era comunicado para uma secretaria de educação passou a ser diferente dos materiais e caminhos escolhidos para divulgar dados para os diretores e também para os professores. Todo o aprimoramento dessa área de comunicação dos dados foi acompanhado pelo desenvolvimento de capacitações destinadas ao uso dos resultados, tema que retomaremos a seguir. O fato é que, reconhecendo problemas de leitura e comunicação dos dados, a pesquisa em avaliação tem se movimentado para enfrentá-los. Prova disso é o aumento do número de profissionais nas secretarias e nas escolas capazes de compreender e utilizar de forma produtiva os resultados em favor do planejamento educacional.

A introdução da Teoria da Resposta ao Item – TRI como fundamento metodológico para a produção de medidas educacionais permitiu que os resultados da avaliação pudessem ser comparáveis ao longo do tempo, desde que alguns requisitos fossem observados (como a existência de itens comuns entre os testes) (KLEIN; FONTANIVE, 1995; PASQUALI; PRIMI, 2003; SARTES; SOUZA-FORMIGONI, 2013). A TRI permitiu também a produção de medidas mais ajustadas, com capacidade de estimar o próprio erro, e a produção de escalas e padrões de desempenho. A ideia de comparação é

outro fator a contribuir para a resistência em relação à avaliação. Vinculando avaliação exclusivamente com ranqueamento, a crítica tende a menosprezar a importância da comparação como elemento de análise e conhecimento. É comparando um aluno com o padrão estabelecido para o que é ser alfabetizado que um professor pode alfabetizá-lo. A comparação, por si, não é um problema, mas sim a forma como ela é utilizada. A principal e mais importante comparação permitida pela avaliação educacional é a da escola com ela mesma, ao longo do tempo. A TRI é um exemplo da contínua busca da pesquisa em avaliação pela sofisticação e ajuste de seus instrumentos e preceitos. O fato de ser uma teoria abstrata e de difícil entendimento não retira sua importância para a produção de medidas. Isso é um problema a ser resolvido pela comunicação, conforme dissemos no parágrafo anterior. Coisas necessárias e importantes, não raras vezes, são de difícil entendimento. O desafio é realizar um bom trabalho de comunicação, não eliminar a dificuldade por meio da eliminação da teoria.

O risco de um professor ou uma escola usar a avaliação como base para o enxugamento do currículo, concentrando seus esforços em ensinar apenas “o que cai no teste”, é um problema relacionado ao mau uso da avaliação e não à ideia de avaliar (MADAUS; RUSSELL; HIGGINS, 2009). Sua importância permanece inalterada, mesmo diante desses casos, que são, inclusive, criticados pelos próprios especialistas em avaliação. Isso não significa que esse não seja um problema real. Ele é, mas seu combate não passa pela exclusão da avaliação do cenário educacional. A recorrência de problemas dessa natureza está mais relacionada à falta de informação e ao desconhecimento da avaliação do que ao instrumento em si.

O instrumento também tem seus limites, como qualquer outro instrumento. Ele é incapaz de abarcar tudo o que está previsto no currículo. Pelo formato adotado, os testes foram compostos, tradicionalmente, por questões de múltipla escolha. Nem tudo o que está previsto no currículo é capaz de ser avaliado por meio desse formato. Para lidar com isso, com o tempo, a pesquisa em avaliação desenvolveu testes para a escrita, com produção de medidas para questões abertas (com controle de leniência da correção). Agora, estamos vendo o desenvolvimento de avaliações da

fluência em leitura, outra dimensão de base curricular (a oralidade) que passa a contar com instrumentos para sua avaliação em larga escala.

Os avanços da pesquisa em avaliação não atestam a ausência de limites, desafios e problemas dos instrumentos e da política de avaliação, mas são capazes de nos mostrar que, diante dos problemas existentes, a área e seus especialistas estão se movimentando para tentar dar soluções a eles. Isso pode ajudar a explicar a expansão dos sistemas próprios de avaliação e a já ressaltada longevidade das políticas de avaliação. Secretarias de educação e escolas têm, cada vez mais, feito uso de resultados para o desenho de políticas e ações e como base do planejamento. A resistência de boa parte da academia, nos cursos de Pedagogia, à avaliação educacional confronta a importância que a ela tem sido atribuída pelos atores do que a própria academia gosta de chamar de “chão da escola”. Esse é um dos exemplos de afastamento entre a teoria e a prática educacionais. O vínculo da avaliação com as propostas de reformas gerenciais na educação fez cegar a academia em relação à importância de usar evidências (informações qualificadas) para a tomada de decisões tempestivas, eficazes e legítimas. É irônico observar que a academia se construiu e mantém-se, justamente, por conta de seu vínculo com a produção de evidências.

Nós nos afinamos com a leitura de Bauer, Alavarise e Oliveira (2015, p. 1369) a respeito das avaliações: as avaliações são instrumento importantes e úteis, mesmo que sejam questionáveis alguns de seus usos para a gestão. Utilizar a avaliação como critério único para a alocação de recursos e para a definição do que é qualidade em educação, estabelecer rankings que estimulam uma competição improdutiva entre as escolas não devem ser tratados como os únicos usos que definem a avaliação. Ele é um instrumento poderoso para a gestão educacional.

Avaliação e gestão educacional

Conforme resalta Calderón (2017), o uso de resultados educacionais tornou-se um tema tão importante para a pesquisa em avaliação que se transformou em área própria de pesquisa. O número de trabalhos dedicado ao uso de resultados responde por boa parte das dissertações de programas

específicos em avaliação e gestão da educação, como é o caso do PPGP/CAEd (Programa de Pós-Graduação Profissional em Gestão e Avaliação da Educação Pública/UFJF). Se a avaliação é um instrumento a serviço da tomada de decisão no âmbito educacional, ela serviu, inicialmente, ao trabalho das secretarias de educação, dando base para ações que alcançassem a rede com um todo (a ideia de uma avaliação externa em larga escala serve a esse propósito). Entretanto, o uso privilegiado que os resultados da avaliação podem ter para as secretarias não faz olvidar a importância desses resultados para as escolas. Na verdade, podemos defender que o processo de implementação de uma política de avaliação só é completo e eficaz quando os resultados são utilizados pelas escolas. Conforme já ressaltamos, a avaliação é instrumental, servindo ao propósito de informar e dar base à ação. Somente a aplicação de testes não esgota a política, que pretende muito mais do que isso. Da mesma forma, a divulgação dos resultados, embora seja uma etapa necessária, não é suficiente para que a política tenha sido, de fato, implementada. A política de avaliação funciona com completude quando os resultados são utilizados pelos atores escolares e educacionais com fonte para a ação. Pelo menos duas questões importantes podem ser desenvolvidas a partir disso: i) o protagonismo do gestor para que a política de avaliação tenha efeitos na escola; e ii) os desafios que se colocam para o uso efetivo dos resultados pelas escolas e redes.

Afirmar o protagonismo do diretor para o processo de uso dos dados na escola não significa diminuir a importância e o papel do professor nesse processo. No entanto, é inegável que a gestão é a porta de entrada para os dados educacionais na escola, responsável por sua organização e utilização. Cabe ao diretor a organização do processo pedagógico, a condução do processo de definição de objetivos comuns, o estabelecimento de funções, a distribuição de tarefas e responsabilidades, a identificação de problemas, a organização conjunta das soluções e do trabalho colaborativo. Não é à toa que a literatura sobre o tema vê na gestão o caminho para que os dados sejam de fato utilizados como parte das rotinas de planejamento e trabalho na escola. É o caso do Data Wise (BOUDET; CITY; MURNANE, 2013), obra que ganhou grande destaque no contexto norte-americano, propondo um uso

inteligente do enorme contingente de dados que os Estados Unidos possuem sobre educação. O Data Wise é um conjunto de orientações e sugestões para as escolas utilizarem dados educacionais de modo produtivo e eficaz, com proposta baseada, entre outras coisas, na organização do trabalho colaborativo na escola, na inserção dos indicadores educacionais como fonte rotineira de tomada de decisão, como base para o planejamento e para a organização de reuniões e atividades de trabalho específicas para o uso de dados. Toda a proposta é centrada na ação da gestão da escola, evidentemente, envolvendo a participação de professores e demais atores escolares. Voltaremos a falar sobre ela adiante.

Há aqui um ponto importante para o argumento que estamos desenvolvendo ao longo de todo o livro. Se o uso dos resultados se tornou um elemento central para toda a política de avaliação e se a responsabilidade por tal uso é atribuída, preponderantemente, à ação da gestão, a crítica direcionada à liderança do gestor por ser contrária à gestão democrática encontra aqui um paralelo. Conforme já ressaltamos em capítulo anterior, a ideia de liderança, no âmbito da gestão, é vista como sinônimo de centralização, e mesmo de autoritarismo, no contexto escolar. É dar destaque ao gestor como o único responsável pelas decisões da escola, sem que a participação dos demais atores escolares seja estimulada e observada. A avaliação, assim, seria mais um dos exemplos onde essa autoridade em vias de transformação para o autoritarismo é exercida. Da mesma forma que, naquele momento, rechaçamos a defesa da liderança como sinônimo de autoritarismo, também não faz sentido entender o protagonismo da gestão em relação ao uso de dados da avaliação como uma exclusão da participação dos demais atores.

A escola possui atores relacionados a diferentes objetivos e funções. Professores estão envolvidos mais detidamente com suas turmas e disciplinas. Além disso, em regra, não são exclusivos da escola, atuando em diferentes unidades e redes. O gestor é responsável por ter uma visão completa da escola, conduzindo o processo de elaboração compartilhada dos objetivos. Faz sentido que ele conduza o processo de utilização de dados relativos a diferentes etapas e disciplinas. O compartilhamento de informações e a distribuição de tarefas e responsabilidades é parte fundamental do processo

de uso de resultados, como vemos na proposta de Boudett, City e Murnane (2013), por exemplo. O protagonismo que o gestor exerce no âmbito de ações dessa natureza está longe de ser sinônimo de centralização.

Defender que os resultados devem ser utilizados pela escola, contudo, é mais desafiador do que parece e envolve a observância de uma série de questões. Podemos afirmar que há, pelo menos, quatro grupos de desafios apresentados à gestão da escola para o uso efetivo dos resultados das avaliações (e isso se aplica a qualquer indicador educacional): i) problemas de conhecimento dos resultados; ii) problemas de acesso aos resultados; iii) problemas de interpretação dos resultados; e iv) problemas de utilização propriamente dita dos resultados. Esses problemas estão relacionados uns aos outros, na medida em que só posso ter problemas de acesso aos resultados se sei da existência dos mesmos; da mesma maneira, só posso ter problemas de interpretação se, previamente, tive acesso aos dados.

Problemas de conhecimento estão relacionados com a informação sobre a existência dos resultados. Em casos mais graves, o ator educacional sequer sabe que existem resultados referentes à aprendizagem dos alunos e que dizem respeito à sua escola. Problemas desse tipo eram mais comuns no início da implementação dos sistemas de avaliação no Brasil. Para diretores escolares, tendo em vista o contexto atual de desenvolvimento dos sistemas de avaliação no Brasil, é difícil imaginar uma situação como essa. No entanto, há professores, em especial aqueles de etapas e disciplinas não avaliadas, que ainda hoje não sabem da existência de informações dessa natureza. E, de acordo com o que já argumentamos, o uso de resultados é uma questão de mobilização da escola como um todo. É preciso pensar ainda a que estamos nos referindo quando falamos em resultados de avaliação. O uso no plural não é gratuito, pois há diferentes resultados educacionais, tanto em natureza quando em relação ao nível de agregação, e os atores podem saber da existência de alguns e desconhecer a de outros (saber da existência da medida de proficiência é mais provável do que conhecer a distribuição dos alunos pelos padrões de desempenho ou os resultados de participação, por exemplo).

A segunda categoria de problemas é mais comum de ser observada. Os atores educacionais podem saber da existência dos resultados, mesmo que de forma superficial ("sei que existe um sistema de avaliação em meu estado e provas são aplicadas em minha escola"), mas não sabem onde eles estão disponíveis, onde são publicados para acesso público. A própria natureza pública da informação pode ser desconhecida, levando a crer que determinado resultado não é disponibilizado para os atores escolares. Isso estimula, inclusive, a crítica e a resistência à avaliação, tratada como um instrumento que produz informações por poucos e para poucos, aumentando a percepção da política como destinada apenas a seus elaboradores, sem serventia para a escola. Da mesma forma que ocorre com os problemas de conhecimento sobre a existência da informação, os atores educacionais podem saber onde encontrar determinados resultados e desconhecer o caminho para os demais.

Problemas de interpretação já estão relacionados a uma natureza analítica, diferente daquela que caracteriza os problemas de conhecimento e acesso. Nesse caso, o ator educacional sabe que os resultados existem e sabe onde encontrá-los. Entretanto, esbarra no significado do dado, em sua interpretação, naquilo que ele diz e no que não diz. A interpretação dos resultados é uma etapa muito delicada para a política de avaliação como um todo e para o uso dos resultados, em específico. Uma interpretação equivocada pode levar a conclusões equivocadas. Se os resultados forem efetivamente usados como base para o desenho de políticas públicas, planejamento e ações escolares, estando ancorados em uma leitura equivocada das informações, as consequências disso podem trazer grandes problemas. Considerando o histórico de resistência à avaliação que narramos anteriormente, baseado em lacunas de formação sobre o tema, não é difícil imaginar que problemas de interpretação ocorrem com uma frequência preocupante. O tipo de conhecimento mobilizado pelos dados e a linguagem com que são divulgados podem contribuir para aprofundar ainda mais o problema, assim como a maneira pela qual as cobranças por melhores resultados chegam aos atores. A questão da interpretação correta dos resultados é um ponto nevrálgico para a sobrevivência da avaliação como um instrumento importante para as políticas e ações educacionais. A contínua interpretação equivocada dos

dados pode levar a ações infundadas repetidamente, com resultados que advogam contra a avaliação, inflamando a crítica. Aliás, parte da resistência à avaliação, responsável por confundir instrumento com seu uso, relaciona-se com isso: resultados mal interpretados, levando a conclusões não autorizadas, sendo usados como base para a tomada de decisões.

Embora os problemas de interpretação representem ainda um grande perigo para a avaliação, o maior desafio atual da política de avaliação é o uso dos resultados. Segundo o que já afirmamos, o uso dos resultados é a etapa em que a política vê sua implementação se completar. É o momento em que os objetivos da política podem ser alcançados. Se a avaliação, conforme sua proposta, natureza e fundamentos (vale lembrar o inciso VI da LDB/96), está a serviço da melhoria da qualidade da educação, por meio de diagnósticos qualificados que darão base a ações informadas, o uso dos resultados é essencial para que isso aconteça. Nenhum efeito relacionado à qualidade da educação pode ser esperado de uma política que apenas mapeia e identifica problemas de aprendizagem, sem que essas informações sejam usadas para orientar a busca de soluções para seu enfrentamento. O problema do uso é o que aflige grande parte dos atores educacionais que já incorporaram a avaliação como parte das rotinas de trabalho da escola. Eles sabem da existência das informações, são capazes de acessá-las e mesmo de interpretá-las corretamente. O passo agora diz respeito ao que fazer com elas. Uma vez compreendidos os dados, e isso envolve a identificação dos seus limites, não é evidente, sempre, o que se pode e o que se deve fazer com eles. A interpretação correta dos resultados é necessária para seu uso adequado, mas não é suficiente. Obviamente, a defesa do uso dos resultados não envolve um tipo qualquer de uso. Os resultados podem ser utilizados, mas de maneira indevida, o que presta um desserviço ainda maior do que a ausência de utilização.

Os resultados da avaliação educacional

Quais são então, os resultados de um sistema de avaliação educacional? Normalmente, o aspecto mais conhecido de toda a avaliação é a “nota” que um aluno ou uma escola recebe. Na verdade, trata-se de uma medida de

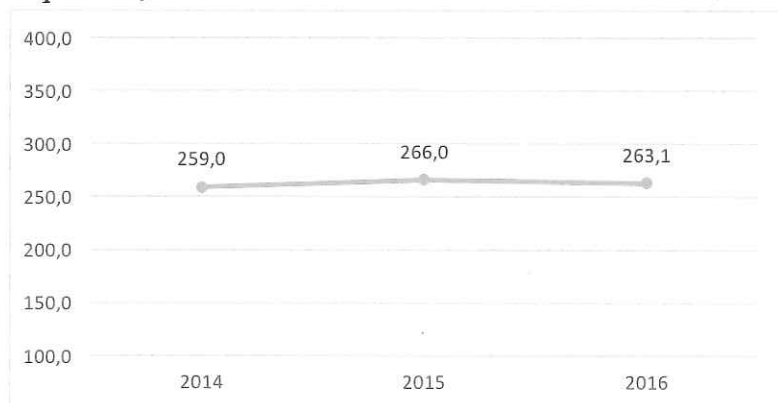
proficiência, calculada com base nos parâmetros da TRI. A aprendizagem é um traço latente, que não pode ser medido diretamente, necessitando, para sua mensuração, de instrumentos que sejam capazes de medi-lo indiretamente. A proficiência, portanto, é uma medida estimada de desempenho, com base nas respostas de um indivíduo às questões de um teste preparado para avaliar o desenvolvimento de determinadas habilidades a partir de itens construídos especificamente para tanto. Assim, não se trata de atribuir uma nota, com base na leitura e avaliação de alguém sobre a aprendizagem. Nos testes padronizados, ninguém dá a nota. A proficiência não depende do juízo de valor de um corretor, como ocorre na avaliação interna, própria da escola, realizada pelo professor, por exemplo. Esse é um dos pontos de vantagem da TRI. Além disso, conforme já dissemos, a metodologia utilizada é capaz de estimar o erro da medida, coisa que a avaliação interna também não faz. Um dos limites da teoria, contudo, está no número de respondentes ao teste necessário para que a medida seja produzida com a incidência de um erro menor. Quanto maior o número de respondentes, menor o erro da medida. Por isso, as avaliações, como a Prova Brasil, por exemplo, preveem um número mínimo de alunos participantes por turma para produzir uma medida nesse nível.

A TRI se baseia, em regra, em três parâmetros, com funções específicas: i) ela controla o nível de discriminação dos itens, de modo a garantir que o aluno que acertou determinado item o fez por ter desenvolvido efetivamente a habilidade que está sendo avaliada; ii) a metodologia é capaz de estimar a dificuldade do item, calculada com base em um pré-teste e definida pelo número de acertos e erros que aquele item apresentou; iii) a TRI estima a chance do aluno ter respondido aleatoriamente a uma questão, com base em seus acertos e erros em itens com dificuldades diferentes.

A medida de proficiência é o que aparece com mais destaque nos resultados da avaliação. É a informação da mídia, dos jornais. É a informação utilizada pelas secretarias para anunciar melhorias na educação do estado ou do município. A proficiência também é o que mobiliza a crítica para dizer que as políticas educacionais não estão dando resultados e para relatar piores. Mas o que significa, efetivamente, melhoras ou piores quando observamos

os resultados de proficiência? Observe o gráfico 1, a seguir. Ele informa os resultados de proficiência para o 3º ano do ensino médio, em língua portuguesa, da rede estadual de Pernambuco, avaliada pelo Saepe (Sistema de Avaliação Educacional de Pernambuco), entre os anos de 2014 e 2016.

Gráfico 1: Proficiência média – língua portuguesa – 3º EM – rede estadual – Saepe – 2014-2016



Fonte: Universidade Federal de Juiz de Fora, (SAEPE), 2016. Tabulação própria.

O que podemos dizer sobre esses resultados? Analisando os números absolutos, é possível observar que a proficiência média aumentou entre 2014 e 2015 (7 pontos) e diminuiu entre 2015 e 2016 (2,9 pontos). Mas o que querem dizer essas diferenças? Esse aumento foi significativo? E a redução, o que significa? Antes de discutir sobre essas questões, é necessário fazer uma observação. A utilização da TRI como base para a construção dos testes e para a produção das medidas de proficiência no Brasil (a partir de meados dos anos 1990) permitiu a construção de escalas de proficiência, nas quais os resultados são distribuídos e podem ser interpretados. Permitiu, igualmente, a construção de padrões de desempenho estudantil, dos quais falaremos adiante. Em geral, as escalas de proficiência das avaliações no Brasil, sejam nacionais ou estaduais, apresentam dois tipos de variação: de 0 a 500 pontos, como é o caso do Saepe, em Pernambuco, no exemplo dado; e de 0 a 1000 pontos, que tem como o exemplo mais conhecido o Enem. Se a escala do Saepe varia de 0 a 500, os resultados apresentados pelo gráfico 1

revelam um resultado próximo ao meio da escala (250 pontos). Isso pode ser considerado ruim, se o leitor acredita que estar próximo ao meio é algo ruim, ou pode ser considerado razoável. Contudo, os resultados de proficiência não operam por meio dessa lógica.

A escala de proficiência é dividida em níveis, em geral, em intervalos de 25 pontos. Cada nível representa um grau de desenvolvimento de determinadas habilidades. Os níveis, por sua vez, são reunidos em padrões de desempenho (em regra, em número de 4), que apresentam informações pedagógicas importantes: eles dizem o que os alunos já consolidaram ou não em termos de habilidades, dando fundamento para o planejamento de intervenções específicas para atender ao necessário para cada padrão. A definição dos intervalos na escala para os padrões é dada pelas secretarias de educação: elas dizem o que é considerado adequado, em termos de desempenho, para cada etapa e disciplina avaliadas. E o estabelecimento do que é adequado pela secretaria não guarda uma relação direta e necessária com o ponto central da escala de proficiência.

Diante dessas informações, voltemos agora à análise do gráfico. O que os resultados dizem? Se o que é adequado em termos de desempenho para o 3º ano do ensino médio, em língua portuguesa, é definido pela secretaria estadual de educação de Pernambuco, não somos capazes, por meio das informações disponibilizadas pelo gráfico, de dizer se esses resultados atendem ao que a secretaria estabeleceu como adequado ou não. Sem a informação sobre os padrões, não conseguimos dizer se os resultados são bons ou ruins, considerando as definições da secretaria. Entretanto, normalmente, quando os resultados de proficiência são divulgados, eles aparecem desse modo, números absolutos, sem maiores explicações (em alguns casos, sem sequer a informação sobre a variação da escala, de 0 a 500 pontos). Para quem não os conhece a fundo, esses números, divulgados dessa forma, não querem dizer nada.

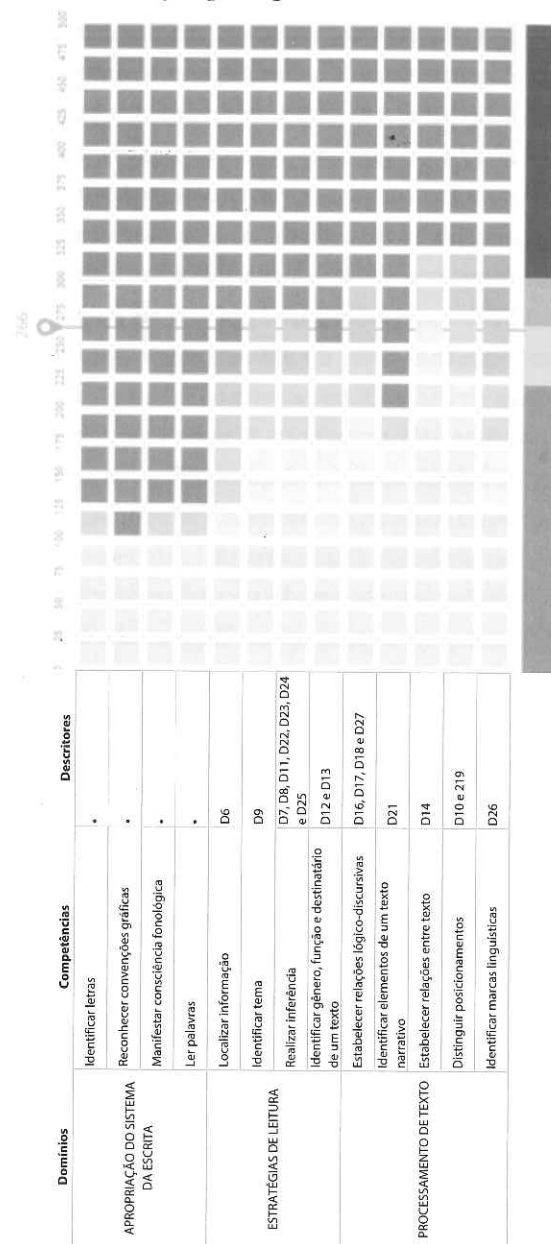
Mas os problemas de interpretação não param por aqui. O aumento de 7 pontos na média de proficiência geral do estado para o 3º ano, entre 2014 e 2015, foi certamente comemorado pela secretaria de educação. Em termos absolutos, o aumento, de fato, existiu. Contudo, o que ele quer dizer? Mais

uma vez, sem informações adicionais, a análise não avança. Já dissemos que a escala é dividida em intervalos, níveis de proficiência de 25 pontos. Os níveis, vale lembrar, representam graus de desenvolvimento de habilidades e oferecem informações pedagógicas ao sinalizar quais das habilidades estão consolidadas em cada nível.

Sendo assim, mudanças de pontuação na escala só são pedagogicamente relevantes quando representam mudanças de níveis. Aumentar 7 pontos de proficiência na escala é indício de uma melhora nos resultados, contudo, pode significar um aumento sem que esse resultado represente uma troca de nível, ou seja, sem alterações em termos de consolidação e aumento de habilidades. Hipoteticamente, se as proficiências fossem sempre no valor exato dos intervalos, melhorias nos resultados só seriam significativas se ultrapassassem 25 pontos. Diante disso, os 7 pontos de aumento que observamos no gráfico⁰¹, entre os anos de 2014 e 2015, são muito menos significativos do que podem parecer.

Para uma melhor compreensão dos padrões de desempenho, observe a Figura 1, a seguir. Ela apresenta a escala de proficiência de língua portuguesa para o 3º ano do ensino médio, no Saepe. Na parte de baixo da figura, estão representados os padrões de desempenho (nas cores vermelha, amarela, laranja e verde). O nível Elementar II é representado pela cor amarela do padrão. Os quadradinhos na escala (em cinza, amarelo claro, amarelo escuro, laranja e vermelho) representam o grau de desenvolvimento das habilidades da matriz (o quadradinho vermelho informa que a habilidade já está desenvolvida a partir daquele ponto da escala). Cada linha da escala é representada por um descritor (nem todos os descritores são avaliados no 3º ano do ensino médio, visto que a escala é a mesma para 5º e 9º anos do ensino fundamental e para o 3º ano do ensino médio) e cada coluna representa o intervalo de 25 pontos de cada nível. Na figura, o ponto representado pela proficiência 266 está marcado.

Figura 1: Escala de proficiência e padrões de desempenho – Saepe – 3º ano do ensino médio – língua portuguesa



O padrão de desempenho Elementar II, definido pela secretaria de educação, responde pelo intervalo na escala de 225 a 270 pontos (pontuação ainda distante dos 305 pontos, que marcam o início do intervalo do padrão Desejável, representando o desempenho esperado para os alunos do 3º ano do ensino médio em língua portuguesa no estado). Isso quer dizer que as médias de proficiência 259 e 266 têm a mesma interpretação pedagógica, representando o mesmo grau de desenvolvimento de habilidades. Evidentemente, apresentar uma melhoria é sempre um bom indício, mas é preciso ter cuidado ao analisar esses dados de modo aligeirado, para não estabelecer conclusões não autorizadas por eles.

O mesmo raciocínio pode ser aplicado à situação entre 2015 e 2016, apresentada pelo gráfico. O que representa uma queda de 3 pontos? Considerando o intervalo do padrão Elementar II, observamos que, do ponto de vista pedagógico, essa redução no valor da proficiência média entre um ano e outra nada representou. A variação de 3 pontos na escala, inclusive, pode ser atribuída a uma eventual diferença na dificuldade dos testes entre as edições, ou seja, pode ser atribuída a diferenças entre os instrumentos, comuns em avaliações em larga escala. Vale lembrar que os testes são comparáveis, graças à metodologia utilizada (TRI), mas nunca são exatamente iguais, de modo que podem apresentar diferenças em termos de dificuldade por conta na variação dos itens entre um ano e outro. Essas informações nunca são divulgadas, porém. O fetiche com o resultado de proficiência em si, com a observação e a comparação dos valores brutos, sem realizar uma interpretação adequada sobre eles, pode levar a superestimar o resultado, com propagandas de que a educação no estado está, de fato, melhorando, ou a subestimá-lo, levando a críticas descabidas e injustas de que a educação no estado está piorando, quando, na verdade, a variação de pontos é insignificante do ponto de vista pedagógico. Para pensar em melhorias ou pioras de resultados significativas, é melhor se orientar pelos intervalos dos níveis. Não há dúvidas de que uma variação de 25 pontos é substantiva e tem significado pedagógico, seja indicando uma melhoria ou uma piora nos resultados. Abaixo dessa variação, há um intervalo que pode ser significativo, observando atentamente os intervalos dos níveis,

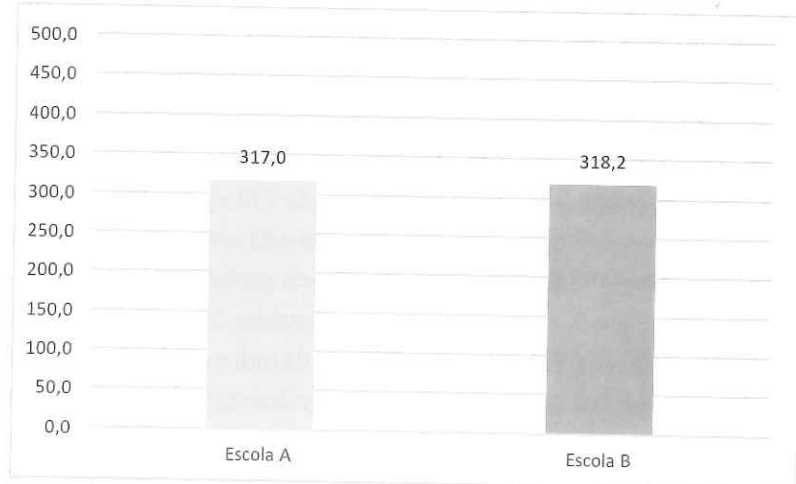
mas, de maneira geral, mudanças de até 7 a 8 pontos, não deveriam causar o estardalhaço que geram. Lembre-se de que se trata de uma avaliação em larga escala e os resultados apresentados pelo gráfico 1 representam todo o estado. Escolas, por representarem um nível de agregação inferior ao de toda a rede, podem apresentar variações muito maiores em seus resultados, abrindo espaço para uma interpretação pedagógica acerca do que significam, de fato, tais variações.

A proficiência apresenta ainda uma limitação. Ela é uma média, o que significa que seu resultado, puro e simples, não carrega consigo uma medida de dispersão. Estamos falando de uma avaliação em larga escala, em que milhares de alunos são avaliados para que uma média de proficiência seja estabelecida. Já ressaltamos que a capacidade da TRI de produzir medidas de proficiência reduzindo ao máximo o erro está associada ao número de respondentes do teste. Por isso, a medida de proficiência é tanto mais precisa quanto mais alunos forem avaliados. A média de proficiência de um estado é formada pelas médias de proficiência de todos os alunos avaliados. Acabamos de apresentar a ideia de padrões de desempenho, intervalos na escala de proficiência com significado pedagógico, representando o nível de desenvolvimento das habilidades da matriz de referência alcançado pelos alunos. Assim, há alunos ocupando os diferentes padrões de desempenho da escala. Eles não possuem desempenhos iguais e, em muitos casos, nem mesmo semelhantes. Pode haver, é claro, uma concentração maior de alunos em um intervalo da escala, o que pesará mais na média de proficiência de todo o grupo. Mas há alunos dispersos ao longo da escala, uns com desempenhos altos e outros com desempenhos baixos.

Um dos objetivos da avaliação educacional é que seus resultados sirvam como indicadores de qualidade da educação. O tema da qualidade está associado ao da equidade: o intuito é potencializar, de modo generalizado, a aprendizagem de todos os alunos, não gerar mais aprendizagem para aqueles que já sabem mais. Uma média de proficiência, contudo, esconde os polos: ela não nos informa sobre o número, e quanto ele representa (o percentual), de alunos localizados em cada intervalo da escala. E isso pode nos levar a cenários bastante distintos. Observe os gráficos a seguir. O

gráfico 2 apresenta a proficiência média de duas escolas (escola A e escola B), no Paebs 2017 (Programa de Avaliação da Educação Básica do Espírito Santo), em biologia, para o 3º ano do ensino médio. O gráfico 3 informa a distribuição dos alunos por padrões de desempenho dessas duas escolas.

Gráfico 2: Proficiência média das escolas A e B - Paebs 2017 - 3º ano do ensino médio - biologia

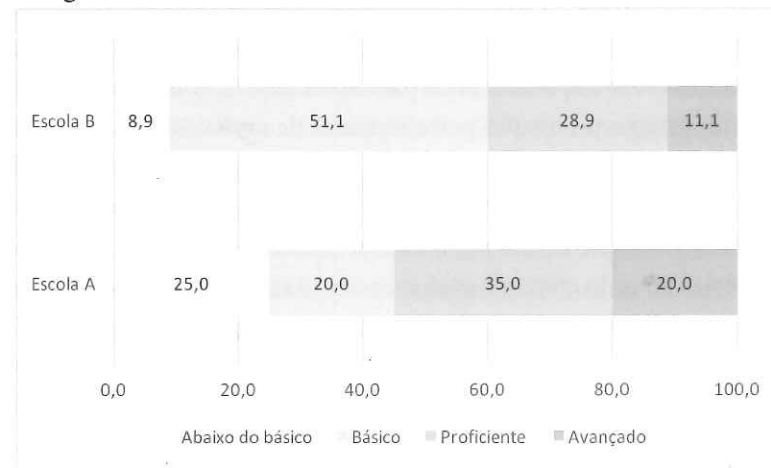


Fonte: Universidade Federal de Juiz de Fora, (PAEBES), 2017. Tabulação própria.

Como podemos observar, a proficiência média em biologia das duas escolas é praticamente a mesma. Conforme o que já falamos, uma diferença de 1 ponto não tem significância estatística alguma. Em termos de padrões de desempenho, as escolas encontram-se na mesma situação (as médias de ambas correspondem ao padrão Básico no Paebs). Pedagogicamente, as médias são iguais, revelando situações iguais. Entretanto, tendo acesso a outros dados, podemos observar que há diferenças significativas entre as duas escolas. O desvio padrão, uma medida de dispersão (indicando quanto um conjunto de dados é ou não uniforme), da escola A difere significativamente da escola B: 71,4 pontos e 47,9 pontos. Isso significa que a escola A apresenta resultados menos homogêneos, com maior dispersão. Mas o que isso significa em termos pedagógicos? Significa que a escola A possui maior concentração

de alunos nos polos (nos padrões mais baixo e mais alto), apresentando maior diferença de desempenho entre os seus alunos. Podemos observar a diferença entre as duas escolas por meio do que é apresentado pelo gráfico 3.

Gráfico 3: Distribuição percentual dos alunos pelos padrões de desempenho das escolas A e B - Paebs 2017 - 3º ano do ensino médio - biologia



Fonte: Universidade Federal de Juiz de Fora, (PAEBES), 2017. Tabulação própria.

Note a diferença de concentração dos alunos nos padrões Abaixo do básico e Avançado nas duas escolas. Na escola B, 8,9% dos alunos encontravam-se no Abaixo do básico em biologia, no 3º ano do ensino médio. No Avançado, o percentual era de 11%. Já a escola A apresentou 25% de alunos no Abaixo do básico e 20% no Avançado. O Abaixo do básico representa, a partir do que é avaliado pelo teste, o pior cenário pedagógico, ao passo que o Avançado responde pela melhor situação em termos de aprendizagem. O que podemos notar é que a escola A é mais desigual do que a B: há mais alunos nos polos, ou seja, há mais alunos com maior diferença de aprendizagem no mesmo ano de escolaridade do que na escola B. Esta última apresenta maior uniformidade em termos de resultado do que a escola A. Essa uniformidade é revelada pelo desvio padrão, que é menor na escola B. O resultado disso manifesta-se na distribuição dos alunos pelos padrões

de desempenho. Se nos concentrarmos apenas na medida de proficiência, chegaremos à conclusão de que não há diferença de desempenho entre os alunos das escolas, que elas são iguais no que diz respeito à aprendizagem em biologia. Entretanto, a partir do acesso a outras informações, como o desvio padrão e a distribuição dos alunos pelos padrões de desempenho, observaremos outro cenário, bem diferente: a escola B possui a mesma média de proficiência do que a escola A, mas sem permitir a mesma concentração de alunos no padrão de desempenho mais baixo.

A distribuição dos alunos pelos padrões de desempenho é um resultado muito importante produzido pelos sistemas de avaliação, por uma série de motivos. Ele apresenta uma informação pedagógica que traduz em desenvolvimento de habilidades os números apresentados pelas médias de proficiência. Isso aproxima, sobremaneira, essa informação dos professores, em especial, fazendo com que as chances de sua utilização como base para o desenho de ações pedagógicas aumentem. No exemplo dado para a escola A, 25% dos seus alunos estarem no Abaixo do básico em biologia significa dizer que eles desenvolveram, ao fim da educação básica, apenas habilidades elementares e de baixa complexidade, como identificar organismos produtores e decompositores em um esquema de cadeia alimentar, identificar o oxigênio como produto da fotossíntese, identificar o modo de prevenção da dengue e reconhecer a função do DNA no comando da célula, por exemplo. Esse não é uma informação importante para os professores e para a gestão da escola? Os padrões de desempenho traduzem, em termos pedagógicos, o que os números querem dizer e apresentam uma visão geral do desempenho dos alunos, que não é informada pela média de proficiência.

Além disso, a partir de uma interpretação bem-feita dos resultados, podemos entender a distribuição dos alunos pelos padrões de desempenho como um indicador não apenas de desempenho, mas que também diz respeito ao fluxo escolar. Voltemos aos resultados apresentados pela tabela 3. Para efeitos de exemplo, vamos imaginar que eles se referem ao 1º ano do ensino fundamental. Ao observar que 25% dos alunos estão no Abaixo do básico, o gestor da escola, sabendo o que significa pedagogicamente esse padrão (que os alunos não desenvolveram, em grau minimamente satisfatório, as habilidades previstas pela matriz, que é de base curricular), já pode imaginar

que a reprovação no 1º ano será grande (pelo menos, 1/4 dos alunos vai reprovar). Isso sem contar nas reprovações dos alunos que se encontram no padrão Básico, também insuficiente em termos de desenvolvimento de habilidades (a rigor, deveriam ser aprovados apenas os alunos que obtiveram desempenho representado pelos padrões Proficiente e Avançado). No caso da escola A, o percentual de reprovados chegaria a impressionantes 45% de alunos, que não têm condições de avançar para a próxima série. Os resultados das avaliações externas, em regra, não são utilizados para fins de aprovação e reprovação na escola (o que é feito por meio das avaliações internas, produzidas e conduzidas pelo professor), mas o exemplo não deixa de apresentar uma questão importante sobre o tema.

Esses 25% de alunos da escola A, assim como os 8,9% da escola B, estão pedindo socorro. São alunos que apenas começaram a desenvolver habilidades básicas previstas para sua etapa de escolaridade. Eles não desenvolveram as habilidades mínimas que fariam com que eles estivessem aptos para avançar para a próxima série. São alunos que carecem de uma intervenção pedagógica substantiva. Essa é outra vantagem da distribuição dos alunos pelos padrões de desempenho: ela permite identificar diferentes grupos de alunos no que diz respeito a seu grau de aprendizagem, dando base à elaboração de soluções adequadas para as dificuldades enfrentadas por cada um deles. Se os alunos no Abaixo do básico precisam de uma recuperação em caráter emergencial, os alunos no Básico ainda não consolidaram todas as habilidades previstas para sua série, o que significa que exigirão do professor um trabalho de reforço para que essa consolidação aconteça. Os alunos no Proficiente atingiram o nível previsto para sua série, mas precisam ainda consolidar habilidades que estão na matriz e representam mais do que apenas o necessário para avançar para a próxima etapa. Os alunos do Avançado, por sua vez, precisam ser desafiados em termos de aprendizagem, de modo que possam se sentir estimulados com o processo de aprendizagem. Não é por estarem no Avançado que não estão mais sujeitos a nenhuma aprendizagem.

Se os resultados relativos aos padrões de desempenho são tão importantes, por que, de maneira geral, não são eles os apresentados com maior destaque para o público? Informações como o desvio padrão, que revelam a distância de aprendizagem entre os alunos da mesma série, também não costumam

ser divulgadas com o mesmo apelo da proficiência. Isso não quer dizer que a média de proficiência não tenha importância. Ela tem o mérito de reunir em um número uma série de informações, afinal, a proficiência é representativa do grau de desenvolvimento de habilidades curriculares. Contudo, ela é uma boa informação, por ser sintética, para quem já conhece a avaliação e os demais resultados que oferece. A proficiência não é o ponto de chegada dos resultados da avaliação, mas o ponto de partida para uma interpretação mais elaborada e com caráter mais pedagógico. O conhecimento acerca dos padrões de desempenho e da distribuição dos alunos por cada um deles oferece um mapeamento mais útil para o professor e a gestão fazerem com que a avaliação exerça sua vocação como suporte para a tomada de decisões no âmbito da escola. A busca pela melhoria da proficiência é substituída pela busca por melhorar a distribuição pelos padrões: o que importa é reduzir, cada vez mais, o percentual de alunos nos padrões mais baixos.

Exceção feita ao estado do Ceará, no qual, com o Spaece (Sistema Permanente de Avaliação da Educação Básica do Ceará), há um trabalho muito ligado aos padrões de desempenho, a maioria dos programas de avaliação dá mais ênfase à medida de proficiência. Quando a avaliação é abordada pela mídia e pelos críticos, é a proficiência que aparece. No Ceará, o trabalho com os padrões levou a um conhecimento, por toda a rede, do que significa, pedagogicamente, cada um deles. As escolas, em especial no ciclo de alfabetização, avaliado pelo Spaece Alfa, são identificadas pela cor de seu padrão: o objetivo é ser verde escuro, o padrão que representa a alfabetização completa. O estado chegou a divulgar uma linha histórica dos municípios, através de mapas², mostrando como eles foram ficando “verdes” ao longo dos anos posteriores à implementação do Paic (Programa de Alfabetização na Idade Certa).

Esses exemplos mostram como o conhecimento acerca dos resultados e do que eles querem dizer é aliado importante da gestão, e, mais do que isso, da própria avaliação, em sua busca contínua por legitimação como um instrumento a serviço da melhoria da qualidade da educação e do desenho

2. De 2007 a 2010, disponíveis em <https://www.seduc.ce.gov.br/wp-content/uploads/sites/37/2011/07/mapa_spaece_alfa_todos_anos_com_variacao_coro2.pdf>.

de políticas educacionais mais bem elaboradas e eficazes. Entender bem o resultado, saber da existência de outros resultados e informações e ser capaz de interpretá-los corretamente são fundamentais para que eles possam ser utilizados de modo realmente produtivo. Uma interpretação equivocada pode levar a decisões igualmente equivocadas em termos de ações e políticas. Vejamos mais um exemplo, dessa vez, com o Ideb.

O Ideb (Índice de Desenvolvimento da Educação Básica) é produzido e divulgado, bienalmente, pelo Inep desde 2007. Combinando fluxo e desempenho escolar, o Ideb é o indicador ligado à qualidade educacional mais relevante e conhecido no cenário nacional. A partir dele, metas foram pactuadas até o ano de 2021. Ele representa a primeira iniciativa nacional de reunir, em um único índice, duas dimensões importantes para a qualidade da educação. As notas do Ideb variam em uma escala de 0 a 10. Vamos observar os resultados do Ideb de duas escolas hipotéticas, C e D, para os anos iniciais, apresentados pelo gráfico 4.

Gráfico 4: Resultados do Ideb das escolas C e D – Anos iniciais



Fonte: Dados hipotéticos. Tabulação própria.

Como podemos ver, as duas escolas possuem exatamente o mesmo Ideb. A essa altura, você já deve estar se perguntando, com base no que vimos anteriormente, o que isso significa, se, de fato, elas são iguais. Lembre-se de que o Ideb é composto por duas dimensões, representadas por dois indicadores diferentes: o fluxo escolar (representado pela taxa de aprovação) e o desempenho (representado pelas proficiências combinadas de língua

portuguesa e matemática produzidas a partir da Prova Brasil). Sendo assim, há caminhos diferentes para que esses resultados possam ser obtidos.

Indo além dos resultados do Ideb, de sua “nota” na escala de 0 a 10, podemos observar como cada um de seus indicadores, fluxo e desempenho, contribui para o resultado final. Será que, necessariamente, a escola C e a escola D chegaram ao resultado de 4,5 pelo mesmo caminho? Vejamos a tabela 1, a seguir. Ela apresenta o desempenho e o fluxo escolar das escolas C e D, além de seu Ideb.

Tabela 1: Proficiência padronizada, taxa de aprovação e Ideb das escolas C e D

Escolas	N	P	Ideb
Escola C	5	0,9	4,5
Escola D	4,5	1	4,5

Fonte: Dados hipotéticos. Tabulação própria.

O Ideb é calculado através da fórmula $N \times P$, em que N representa as médias de proficiência de língua portuguesa e matemática combinadas, padronizadas para um indicador variando entre 0 e 10, e P representa a taxa de aprovação da etapa de ensino (no caso, os anos iniciais), variando de 0 a 1 (1 representa 100% de aprovação). Diante disso, o que podemos observar é que, embora possuam o mesmo resultado no Ideb, a forma como esse resultado foi obtido difere entre as duas escolas. A escola C obteve uma média de 5 no desempenho, com uma taxa de aprovação de 90%. Isso significa que a escola ainda poderá melhorar seu Ideb por meio de ações que reduzam a reprovação dos alunos (uma taxa de 10%), mesmo que seu desempenho não seja alterado (permanecendo com a média 5). Assim, se a escola C conseguir desenvolver ações para aumentar a aprovação, mantendo seu desempenho no mesmo nível, seu Ideb chegará a 5 (caso aprove 100% de seus alunos nos anos iniciais). A situação da escola D é diferente. Seu resultado no Ideb já foi estabelecido com base em uma aprovação de 100%. Não há nada que a escola possa fazer em relação a esse indicador para melhorar seu Ideb. A única saída para ela é melhorar o indicador de desempenho, ou seja,

propiciar uma melhoria da aprendizagem que permite a seus alunos um melhor desempenho na próxima edição da Prova Brasil. Ações de estímulo à aprovação não são uma saída para escola, que terá que se concentrar na aprendizagem.

Dois pontos importantes merecem destaque aqui. O Ideb, como todo índice, possui limitações, a começar pelo fato de que ele não é capaz de traduzir todas as dimensões educacionais importantes para serem avaliadas como parte da ideia de qualidade da educação. Contudo, ele apresenta vantagens. Trata-se de um indicador que combina fluxo e desempenho, em busca de um equilíbrio: desejamos que os alunos desenvolvam determinadas habilidades específicas de cada etapa de escolaridade, que podem ser avaliadas por meio de testes padronizados, ou seja, almejamos que os alunos aprendam o que é previsto no currículo; e desejamos também que eles aprendam em um intervalo de tempo específico, uma série ou um ciclo (como no caso da alfabetização). O conteúdo da aprendizagem é combinado com o tempo adequado para que essa aprendizagem ocorra. O outro ponto importante diz respeito à melhoria dos resultados. O que está em jogo quando observamos um indicador ao longo do tempo, como no caso do Ideb, não é a melhora do indicador a qualquer custo. O que se busca é a aprendizagem, sendo os resultados apenas a consequência dessa melhoria. O que os resultados permitem é uma análise acerca da aprendizagem, se ela está de fato acontecendo e em que medida. Ações que buscam melhorar o resultado por si alimentam as críticas que a avaliação recebe, como, por exemplo, deixar de ensinar todo o currículo para se concentrar apenas nas habilidades avaliadas pelos testes. Com isso, o que se busca não é melhorar o Ideb, mas, antes, potencializar a aprendizagem dos alunos e buscar soluções para a ampliação da aprovação. A melhoria nessas duas frentes terá como consequência a melhoria do Ideb.

Voltemos aos problemas de interpretação de resultados do Ideb, com os exemplos das escolas hipotéticas E e F. A tabela 2 apresenta a proficiência padronizada, a taxa de aprovação e o Ideb final das escolas para os anos iniciais.

Tabela 2: Proficiência padronizada, taxa de aprovação e Ideb das escolas C e D

Escolas	N	P	Ideb
Escola E	5	0,7	3,5
Escola F	5	0,7	3,5

Fonte: Dados hipotéticos. Tabulação própria.

Observando os resultados do Ideb das escolas E e F, chegamos à conclusão de que elas apresentam, além do mesmo Ideb, os mesmos indicadores para fluxo e desempenho. Ambas apresentam uma taxa de aprovação não satisfatória (70%), de modo que há espaço para a atuação da gestão nessa frente. Trata-se de um cenário diferente daquele que observamos para as escolas C e D. Aqui, as escolas E e F podem agir no intuito de aumentar suas taxas de aprovação, ou seja, elas têm o mesmo caminho a percorrer, correto? Antes de chegarmos a essa conclusão, é preciso pensar sobre o indicador *P*. Conforme já dissemos, *P* representa a taxa de aprovação da etapa. As escolas E e F, no exemplo dado, aprovaram 70% de seus alunos nos anos iniciais. Isso significa que essas escolas precisam de soluções para lidar com a reprovação, não é mesmo? Vejamos a tabela 3, a seguir.

Tabela 3: Percentuais de reprovação, abandono e aprovação das escolas E e F

Escolas	Reprovação	Abandono	Aprovação
Escola E	27%	3%	70%
Escola F	6%	24%	70%

Fonte: Dados hipotéticos. Tabulação própria.

A tabela 3 informa os percentuais de reprovação, abandono e aprovação das escolas E e F. Embora elas apresentem o mesmo percentual de alunos não aprovados (30%), os motivos pelos quais eles não foram aprovados diferem significativamente entre as escolas. Na escola E, a taxa de reprovação é realmente alta: 27% nos anos iniciais, ao passo que o abandono responde por 3% do total. Podemos afirmar que, no caso da escola E, emergencialmente,

devem ser procuradas soluções para uma reprovação tão alta. O abandono existe, mas não é o problema central a ser resolvido. Já na escola F, a situação se inverte. Ela apresenta uma taxa de reprovação, de 6%, que pode ser melhorada, mas a maior parte de seus não aprovados abandona a escola. Se a gestão da escola observa apenas a taxa de aprovação, o *P* da fórmula do Ideb, e desenha ações para o combate à reprovação com base nisso, ela corre o risco de não ver os efeitos dessas ações. Isso pode ser muito frustrante para a gestão. Por mais bem-intencionada que esteja, a ausência de uma leitura mais profunda e cuidadosa dos dados pode levar ao desenho de ações ineficazes. No caso da escola E isso até poderia funcionar, mas não ocorreria o mesmo na escola F. O problema que a escola F precisa enfrentar de modo mais emergencial é o abandono. E as ações para se combater o abandono não são necessariamente as mesmas do que aquelas destinadas a lidar com problemas de reprovação.

Diante da exposição desses exemplos, parece claro que uma interpretação bem-feita dos resultados das avaliações pode contribuir decisivamente para uma tomada de decisão bem informada e com mais chance de gerar os efeitos esperados em relação ao combate dos problemas identificados. Uma análise apressada e descuidada tem, pelo menos, duas consequências ruins: i) dá base a soluções com poucas chances de serem eficazes; se não sou capaz de identificar com precisão o problema, como serei capaz de desenhar ações ajustadas para combatê-lo?; ii) ações desajustadas fazem com que a própria avaliação seja julgada a partir dos efeitos inexistentes ou perniciosos dessas ações, ampliando a crítica de que ela não serve para nada ou, pior, leva a decisões ruins.

Por isso, para que a avaliação possa exercer sua vocação de fornecer informações qualificadas para decisões mais eficazes, precisas, ajustadas e legítimas, é necessário que os atores educacionais sejam capazes de fazer um uso adequado desse instrumento, compreendendo suas virtudes e seus limites, e extraíndo dessas informações conclusões autorizadas. Porém, conforme já ressaltamos anteriormente, onde esses atores seriam capacitados para tanto? Os cursos de formação de profissionais para a educação não concedem à avaliação um espaço qualificado em seus currículos. A resistência à avaliação é disseminada, em grande parte, justamente nesses contextos. Já no âmbito profissional, os gestores costumam receber os resultados das avaliações, e

serem cobrados por seu uso, sem que tenham tido uma formação adequada para lidar com os desafios que, por exemplo, a interpretação dos dados oferece a eles.

As questões que se colocam são as seguintes: esses resultados não são importantes? Eles não apresentam um mapeamento das escolas e das redes de ensino central para que boas decisões sejam tomadas? Não interessa ao gestor ter uma visão geral de sua escola no que diz respeito ao aprendizado dos seus alunos? Isso não contribui para o planejamento? Identificar os fatores associados ao desempenho, como faz a pesquisa em eficácia escolar, apontando os elementos que, na escola, são capazes de potencializar a aprendizagem dos alunos, não é algo importante para gestores, professores e atores educacionais das secretarias de educação (como vimos no capítulo 4, com o clima escolar)? Pesquisas dessa natureza são baseadas, muitas vezes, nos resultados produzidos pelos sistemas de avaliação.

Quanto aos resultados, não há uma receita sobre como eles devem ser utilizados. Não há uma forma única e óbvia de como essas informações devem ser melhor usadas pelas escolas e secretarias. Isso dependerá de cada contexto, da organização dessas instituições, do tipo de profissional que possuem, do grau de resistência observado, do perfil da gestão, do tipo de resultados disponíveis etc. A despeito disso, há um conjunto significativo de pesquisas sobre o uso de resultados na escola. No contexto norte-americano, por exemplo, o Data Wise é paradigmático. Os pesquisadores responsáveis pela proposta apresentam um conjunto de ações para propiciar o uso de informações educacionais para promover a melhoria da aprendizagem na escola. Alguns pressupostos independem do contexto: i) informações educacionais qualificadas, como os resultados das avaliações, são evidências que podem ser utilizadas pela escola para promover melhorias na aprendizagem; ii) a gestão exerce papel crucial para que isso aconteça, pois é responsável por organizar esse uso; ela é a porta de entrada dos resultados na escola; iii) é necessário um trabalho colaborativo para que a utilização dos resultados possa ter efeitos, de modo que todos devem ser mobilizados pela gestão para que isso aconteça; iv) os atores escolares precisam conhecer os limites dos resultados, bem como sua natureza e as conclusões que dali podem ser extraídas; v) o trabalho com os resultados envolve a definição de metas coletivas, pactuadas com todos os

atores escolares e pelas quais todos são responsáveis; e vi) os dados relativos à escola, bem como a linguagem e o tipo de raciocínio que mobilizam devem ser integrados à rotina de trabalho das escolas, como parte de suas ações cotidianas. Seguindo a tradição do pragmatismo norte-americano, a proposta do Data Wise chega ao nível do detalhe, dando sugestões, inclusive, de como organizar reuniões sobre como usar os dados como base do planejamento, quanto tempo cada ator tem para falar, o tipo de pauta da reunião, que tipo de argumento deve presidir as discussões etc.

Se a avaliação externa é uma política pública tão longa e central para a educação no Brasil, se os atores das secretarias e escolas têm, cada vez mais, feito uso desse instrumento, por que os cursos de formação de profissionais para a educação insistem em negligenciar a formação nessa área? A resistência à avaliação possui raízes profundas, conforme vimos anteriormente, mas, certamente, seu vínculo com uma perspectiva gerencial, de uso de evidências como base para a ação, e a centralidade que assumiu no contexto de reforma da educação no país contribuíram, sobremaneira, para que essa resistência tenha se ampliado. A despeito disso, em meio às críticas, a avaliação desenvolveu-se e tornou-se um instrumento central para informar a ação de atores posicionados em todos os níveis da estrutura educacional. Seu desafio fundamental agora é ser usada de modo consciente e adequado por esses atores, permitindo, de fato, mudanças na aprendizagem dos alunos e nas formas de organização e condução do processo de ensino, mostrando aos críticos que a ojeriza pura e simples à avaliação não encerra sentido. Nesse processo, a esperança é que a própria universidade reconheça a importância desse tema para a formação de professores e gestores.

Leituras complementares

SCHWARTZMAN, Simon; COX, Cristián. Introdução: Coesão social e políticas educacionais na América Latina. In: SCHWARTZMAN, Simon; COX, Cristián (org.). *Políticas educacionais e coesão social: uma agenda latino-americana*. Rio de Janeiro: Elsevier, 2010. p. 1-25. Disponível em: <https://ia600209.us.archive.org/4/items/CoesaoSocialEPolíticasEducativas/2009coesao_port.pdf>. Acesso em: 07 abr. 2020.

O texto apresenta o contexto de reforma da educação nos países da América Latina nos anos 2000, mostrando a mudança de foco das políticas educacionais: do acesso à qualidade. Neste contexto, a avaliação educacional em larga escala ganha centralidade na agenda educacional. A indicação do texto tem esse propósito: mostrar ao leitor as características do cenário no qual as políticas de avaliação ganham maior projeção no Brasil.

FRANCO, Creso; ALVES, Fátima; BONAMINO, Alicia. Qualidade do ensino fundamental: políticas, suas possibilidades, seus limites. *Educação e Sociedade*, Campinas, v. 28, n. 100 - Edição Especial, p. 989-1014, out. 2007. Disponível em: <<http://www.scielo.br/pdf/es/v28n100/a1728100.pdf>>. Acesso em: 04 mai. 2020.

Os autores apresentam as características das políticas educacionais no Brasil dos anos 1970 até o início do governo Lula, destacando o grau de continuidade de algumas delas. Eles mostram como a avaliação educacional manteve-se como política relevante para a discussão da qualidade da educação no país, atravessando governos de diferentes orientações ideológicas, como os do PSDB e do PT. O intuito do texto é mostrar a continuidade e longevidade da avaliação, que permaneceu no centro das ações educacionais em virtude do papel que desempenha para a busca da qualidade na educação.

BONAMINO, Alicia; SOUSA, Sandra Zákia. Três gerações de avaliação da educação básica no Brasil: interfaces com o currículo da/na escola. *Educação e Pesquisa*, São Paulo, v. 38, n. 2, p. 373-388, abr./jun. 2012. Disponível em: <<https://www.scielo.br/pdf/ep/v38n2/aopep633.pdf>>. Acesso em: 04 mai. 2020.

O texto propõe uma tipologia para as avaliações educacionais com base no uso que pode ser feito de seus resultados. As gerações da avaliação não representam, necessariamente, sucessões cronológicas, mas dizem respeito às consequências de seus resultados. O artigo já é canônico em estudos sobre avaliação educacional. Ele está sendo indicado para mostrar ao leitor como a avaliação, em pouco tempo, foi capaz de se desenvolver e se transformar, a despeito de como isso possa ser avaliado (positiva ou negativamente) e lança luz sobre a importância dos usos dos resultados para a política de avaliação,

um argumento apresentado ao longo do capítulo. Além disso, a tipologia proposta pelas autoras permanece sendo um instrumento analítico poderoso para estudos que têm a avaliação como objeto.

REZENDE, Wagner Silveira. Alguns problemas de interpretação sobre a natureza da avaliação em larga escala. *Pesquisa e Debate em Educação*, v. 4, n. 2, 2014. Disponível em: <<http://www.revistappgp.caeduff.net/index.php/revista1/article/view/104>>. Acesso em: 04 mai. 2020.

Neste texto está esboçado um argumento que aparece na primeira parte do capítulo, qual seja, a forma como a avaliação externa é encarada por boa parte dos atores educacionais: de maneira personificada, à avaliação são atribuídas características de um sujeito, confundindo o instrumento com seu uso. Essa confusão alimenta as críticas e a resistência à avaliação, responsáveis, em grande medida, pelo não uso ou pelo uso inadequado do instrumento, impedindo que a política avaliativa exerça sua vocação de dar base à tomada de decisões. Por isso, o texto é indicado: ele busca esclarecer os problemas de tomar o uso dos resultados feito por alguns atores educacionais como o único uso possível de resultados da avaliação.

BAUER, Adriana; ALAVARSE, Ocimar Munhoz; OLIVEIRA, Romualdo Portela de. Avaliações em larga escala: uma sistematização do debate. *Educação e Pesquisa*, São Paulo, v. 41, p. 1367-84, n. especial, dez. 2015. Disponível em: <<https://www.scielo.br/pdf/ep/v41nspe/1517-9702-ep-41-spe-1367.pdf>>. Acesso em: 04 mai. 2020.

Os autores fazem um mapeamento amplo da literatura sobre avaliação em larga escala, apontando as principais críticas que recebe, como o estreitamento do currículo e o vínculo com políticas de responsabilização (em especial, de transferência de recursos com base em seus resultados), e as principais virtudes que apresentam, como o monitoramento da qualidade da educação e a discussão que pode suscitar sobre a responsabilidade dos atores escolares pela qualidade da educação. O texto foi indicado por apresentar ao leitor uma visão sintética, embora de amplo alcance, sobre o debate atual em torno da avaliação em larga escala. Como todo instrumento, ela apresenta limites e virtudes e o texto expõe isso de maneira clara.

CALDERÓN, Adolfo Ignacio. Usos e apropriações das avaliações em larga escala. In: *Diálogos e proposições: Planos de Ação para a rede estadual de ensino de Minas Gerais*. Editora CAEd/UFJF: Juiz de Fora, 2017. v. 2. p. 31-45. Disponível em: <<http://www.mestrado.caedufjf.net/wp-content/uploads/2018/01/DI%C3%81LOGOS-E-PROPOSI%C3%87%C3%95ES-vol-II-WEB.pdf>>. Acesso em: 04 mai. 2020.

No texto, o autor mostra como o uso dos resultados da avaliação tornou-se tão central para a política de avaliação que assumiu a posição de um dos objetos privilegiados da pesquisa acadêmica em avaliação. O texto apresenta ao leitor um argumento fundamental para todo o capítulo 7 do livro: apenas com o uso dos resultados a política de avaliação pode ser considerada verdadeiramente implementada, tendo em vista que seu objetivo é precipuamente pragmático, qual seja, fazer com que informações qualificadas sobre as escolas e sobre as redes de ensino sirvam como suporte para o desenho de ações e políticas educacionais.

FARIAS, Maria Adalgiza de; MAGALHÃES JUNIOR, Antonio Germano. Gestão da escola e os resultados do IDEB: apropriações e usos de dados educacionais. *Polêm!ca*, Rio de Janeiro, v. 18, n. 2, p. 34-53, abr./jun. 2018. p. 35-77, 44-51. Disponível em: <<https://www.e-publicacoes.uerj.br/index.php/polemica/article/view/37788>>. Acesso em: 04 mai. 2020.

Referências

BASSO, Andreza Cristina Moreira da Silva *et al.* Desigualdade de desempenho e raça: uma análise a partir do Paebs 2009. *Estudos em Avaliação Educacional*, São Paulo, v. 23, n. 51, 2012. Disponível em: <<http://publicacoes.fcc.org.br/ojs/index.php/eae/article/view/1946/1927>>. Acesso em: 15 mai. 2020.

BAUER, Adriana; ALAVARSE, Ocimar Munhoz; OLIVEIRA, Romualdo Portela de. Avaliações em larga escala: uma sistematização do debate. *Educação e Pesquisa*, São Paulo, v. 41, p. 1367-84, n. especial, dez. 2015.

Disponível em: <<https://www.scielo.br/pdf/ep/v41nspe/1517-9702-ep-41-spe-1367.pdf>>. Acesso em: 08 mai. 2020.

BONAMINO, Alicia; SOUSA, Sandra Zákia. Três gerações de avaliação da educação básica no Brasil: interfaces com o currículo da/na escola. *Educação e Pesquisa*, São Paulo, v. 38, n. 2, p. 373-388, abr./jun. 2012. Disponível em: <<https://www.scielo.br/pdf/ep/v38n2/aopep633.pdf>>. Acesso em: 08 mai. 2020.

BOUDET, Kathryn Parker; CITY, Elizabeth A.; MURNANE, Richard J. (ed.). *Data wise: a step-by-step guide to using assessment results to improve teaching and learning*. Cambridge, Massachusetts: Harvard Education Press, 2013.

BRASIL. Congresso Nacional. Constituição da República Federativa do Brasil, de 05 de outubro de 1988. *Diário Oficial da União*, Brasília, DF, 05 out. 1988. Seção 191-A, p. 1. Disponível em: <https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm>. Acesso em: 08 mai. 2020.

BRASIL. Ministério da Educação e do Desporto (MEC). Lei n. 9.394, de 20 de dezembro de 1996. Estabelece as diretrizes e bases da educação nacional. *Diário Oficial da União*, Brasília, DF, 23 dez. 1996. p. 27833. Disponível em: <https://www.planalto.gov.br/ccivil_03/Leis/L9394.htm>. Acesso em: 08 mai. 2020.

CALDERÓN, Adolfo Ignacio. Usos e apropriações das avaliações em larga escala. In: *Diálogos e proposições: Planos de Ação para a rede estadual de ensino de Minas Gerais*. Editora CAEd/UFJF: Juiz de Fora, 2017. v. 2. p. 31-45. Disponível em: <<http://www.mestrado.caedufjf.net/wp-content/uploads/2018/01/DI%C3%81LOGOS-E-PROPOSI%C3%87%C3%95ES-vol-II-WEB.pdf>>. Acesso em: 08 mai. 2020.

DOURADO, Luiz Fernando. Políticas de Gestão da Educação Básica no Brasil: limites e perspectivas. *Educação e Sociedade*, Campinas, v. 28, n. 100 - Edição Especial, p. 921-946, out. 2007. Disponível em: <<http://www.scielo.br/pdf/es/v28n100/a1428100.pdf>>. Acesso em: 21 mar. 2020.

FRANCO, Creso; ALVES, Fátima; BONAMINO, Alicia. Qualidade do ensino fundamental: políticas, suas possibilidades, seus limites. *Educação e Sociedade*, Campinas, v. 28, n. 100 - Edição Especial, p. 989-1014, out.

2007. Disponível em: <<http://www.scielo.br/pdf/es/v28n100/a1728100.pdf>>. Acesso em: 21 mar. 2020.

FREITAS, Luiz C. Caminhos da avaliação de sistemas educacionais no Brasil: o embate entre a cultura da auditoria e a cultura da avaliação. In: BAUER, Adriana; GATTI, Bernardete A. (org.). *Vinte e cinco anos de avaliação de sistemas educacionais no Brasil: origens e pressupostos*. Florianópolis: Insular, 2013a. v. 2. Disponível em: <<https://www.fcc.org.br/fcc/fcc-publicacoes/vinte-e-cinco-anos-de-avaliacao-de-sistemas-educacionais-no-brasil-volume-2>>. Acesso em: 15 mai. 2020.

FREITAS, Luiz C. Os reformadores empresariais da educação: da desmoralização do magistério à destruição do sistema público de educação. *Educação e Sociedade*, Campinas, v. 33, n. 119, p. 379-404, abr./jun. 2012. Disponível em: <<https://www.scielo.br/pdf/es/v33n119/a04v33n119.pdf>>. Acesso em: 15 mai. 2020.

FREITAS, Luiz C. Políticas de responsabilização: entre a falta de evidência e a ética. *Cadernos de Pesquisa*, São Paulo, v. 43, n. 148, p. 348-65, jan./abr. 2013b. Disponível em: <<https://www.scielo.br/pdf/cp/v43n148/18.pdf>>. Acesso em: 11 mai. 2020.

KLEIN, Ruben; FONTANIVE, Nilma Santos. Avaliação em larga escala: uma proposta inovadora. *Em Aberto*, Brasília, ano 15, n. 66, abr./jun. 1995. Disponível em: <<https://pdfs.semanticscholar.org/3294/ce1b99dbe3c49eb75a701b375e80d1414559.pdf>>. Acesso em: 11 mai. 2020.

MADAUS, George; RUSSELL, Michael; HIGGINS, Jennifer. *The paradoxes of high stakes testing: how they affect students, their parents, teachers, principals, schools, and society*. Charlotte, N.C.: Information Age Publishing Inc., 2009.

PASQUALI, Luiz; PRIMI, Ricardo. Fundamentos da Teoria da Resposta ao Item – TRI. *Avaliação psicológica*, Porto Alegre, v. 2, n. 2, p. 99-110, dez. 2003. Disponível em: <<http://pepsic.bvsalud.org/pdf/avp/v2n2/v2n2a02.pdf>>. Acesso em: 11 mai. 2020.

REZENDE, Wagner Silveira. Alguns problemas de interpretação sobre a natureza da avaliação em larga escala. *Pesquisa e Debate em Educação*, v. 4,

n. 2, 2014. Disponível em: <<http://www.revistappgp.caeduff.net/index.php/revista1/article/view/104>>. Acesso em: 21 mar. 2020.

SARTES, Laisa Marcorela Andreoli; SOUZA-FORMIGONI, Maria Lucia Oliveira de. Avanços na Psicometria: da Teoria Clássica dos Testes à Teoria da Resposta ao Item. *Psicologia: Reflexão e Crítica*, Porto Alegre, v. 26, n. 2, p. 241-50, 2013. Disponível em: <<https://www.scielo.br/pdf/prc/v26n2/04.pdf>>. Acesso em: 11 mai. 2020.

SCHWARTZMAN, Simon; COX, Cristián. Introdução: Coesão social e políticas educacionais na América Latina. In: SCHWARTZMAN, Simon; COX, Cristián (org.). *Políticas educacionais e coesão social: uma agenda latino-americana*. Rio de Janeiro: Elsevier, 2010. p. 1-25. Disponível em: <https://ia600209.us.archive.org/4/items/CoesaoSocialEPoliticasEducacionais/2009coesao_port.pdf>. Acesso em: 07 abr. 2020.

UNIVERSIDADE FEDERAL DE JUIZ DE FORA. Faculdade de Educação. Centro de Políticas Públicas e Avaliação da Educação (CAEd). Resultados do Sistema de Avaliação Educacional de Pernambuco (SAEPE). *Língua Portuguesa – 3º ano do Ensino Médio*. 2018. Disponível em: <http://saepe.caeduff.net/escalas/?page_id=197>. Acesso em: 21 mar. 2020.

UNIVERSIDADE FEDERAL DE JUIZ DE FORA. Centro de Políticas Públicas e Avaliação da Educação (CAEd). *Sistema de Avaliação Educacional de Pernambuco* (SAEPE). 2016. Disponível em: <<http://www.saepe.caeduff.net/colecoes/anos-anteriores/>>. Acesso em: 22 maio 2020.

UNIVERSIDADE FEDERAL DE JUIZ DE FORA. Centro de Políticas Públicas e Avaliação da Educação (CAEd). *Programa de Avaliação da Educação Básica do Espírito Santo* (PAEBES). [2017]. Disponível em: <<http://www.paebes.caeduff.net/resultados/>>. Acesso em: 22 mai. 2020.

XIMENES, Salomão Barros. Responsabilidade Educacional: Concepções Diferentes e Riscos Iminentes ao Direito à Educação. *Educação e Sociedade*, Campinas, v. 33, n. 119, p. 353-77, abr./jun. 2012. Disponível em: <<https://www.scielo.br/pdf/es/v33n119/a03v33n119.pdf>>. Acesso em: 11 mai. 2020.