GT CONSED

AVALIAÇÃO DA EDUCAÇÃO BÁSICA

RELATÓRIO

Organização:

Parceiro realizador:

Parceiros:

EQUIPE DO RELATÓRIO DO GRUPO DE TRABALHO DE AVALIAÇÃO DO CONSED

PRODUÇÃO EXECUTIVA:
Laura Muller Machado
Emilio Veludo Lopes
Emília Marinho

EDIÇÃO:
Diana Coutinho

ELABORAÇÃO DO CONTEÚDO:

Cap. 1: A EXPERIÊNCIA RECENTE DAS UNIDADES DA FEDERAÇÃO EM AVALIAÇÃO DA EDUCAÇÃO BÁSICA
Rosane Mendonça
Diana Coutinho
Beatriz Garcia

Cap. 2: PROPOSTAS PARA O SISTEMA DE AVALIAÇÃO DA EDUCAÇÃO BÁSICA
Frederico da Costa Amancio
Severino José de Andrade Júnior

Cap. 3: CONTEÚDOS POSSÍVEIS PARA UM SISTEMA DE AVALIAÇÃO DA EDUCAÇÃO BÁSICA
Ricardo Paes de Barros
Diana Coutinho
Rosane Mendonça

Cap. 4: PROPÓSITOS, USOS, UTILIDADES E POTENCIAIS CONSEQUÊNCIAS NEGATIVAS DE UM SISTEMA DE AVALIAÇÃO DA EDUCAÇÃO BÁSICA
Ricardo Paes de Barros
Diana Coutinho
Rosane Mendonça
Laura Muller Machado

Cap. 5: ARQUITETURA DO SISTEMA DE AVALIAÇÃO DA EDUCAÇÃO BÁSICA
Ricardo Paes de Barros
Diana Coutinho
Rosane Mendonça

Cap. 6: ECONOMIAS DE ESCALA E OPORTUNIDADES PARA A REDUÇÃO NOS CUSTOS DA AVALIAÇÃO DA EDUCAÇÃO BÁSICA
Ricardo Paes de Barros
Diana Coutinho
Juliana Frizzoni Candian
Beatriz Garcia

Cap. 7: RESUMO DAS REUNIÕES TÉCNICAS DO GRUPO DE TRABALHO DE AVALIAÇÃO DO CONSED
Patricia Oliveira

REVISÃO:
Camila Mata Machado Soares

PRODUÇÃO ESTATÍSTICA:
OPESociais
Beatriz Garcia

DIAGRAMAÇÃO:
Otavio Barros Falcão Junior

COMPOSIÇÃO DO GRUPO DE TRABALHO DE AVALIAÇÃO DO CONSED

COORDENAÇÃO GERAL:
Frederico da Costa Amancio, coordenador
Idilvan Alencar, coordenador adjunto

REPRESENTANTES DAS SECRETARIAS ESTADUAIS DE EDUCAÇÃO:
Afonso Gomes Ferreira Filho (RN), coordenador adjunto do subgrupo de Usos e Arquitetura
Amanda Midori Amano (DF), coordenadora do subgrupo de Conteúdo
Aparecida Meireles de Souza e Souza (RO)
Augusto Cezar Barbosa Brito (AP)
Carla Bertânia Souza (RJ)
Carmilva Souza Flores (CE)
Cheila F. B. S. Vasconcelos (AL)
Cosme Rocha (PI), coordenador adjunto do subgrupo de Custos
Cyntia Lemes da Silva G. de Fonseca (SP)
Eduardo Malini (ES), coordenador do subgrupo de Custos
Emerson Azevedo Soares (TO)
Evandro dos Santos Paiva (PA)
Fatima Cristina D. Medeiros (BA), coordenadora do subgrupo de Usos e Arquitetura
Fredson da Costa (AP)
Geniana Faria (MG)
Hilda Maria Freire Montysuma (RR)
Irene de Souza Costa (MT)
Jonyele Cheyenne Moura Cruz (SE), coordenadora adjunta do subgrupo de Conteúdo
Katya Prust (PR)
Luis Alex Lopes Correia (MA)
Maria Cristina Guilherme (AC), *in memoriam*
Marinaldo Alves (PE)
Marta de Medeiros Correia (PB)
Salete Doria Albuquerque (RS)
Sandra Mara Cardoso (SC)
Shirlene Noronha Guimarães (AM)
Soraya Regina Cruz (MS)
Vania Machado Oliveira (RJ)
Weyne Maria Magalhães Carneiro (GO)

ASSISTÊNCIA AO GRUPO DE TRABALHO DE AVALIAÇÃO DO CONSED

COORDENAÇÃO:
Patricia Oliveira

FACILITAÇÃO:
Carolina Miotto Serain, subgrupo de Usos e Arquitetura
Diana Coutinho, subgrupo de Conteúdo
Gabriela Pluciennik, geral
Juliana Frizzoni Candian, subgrupo de Custos
Laura Muller Machado, subgrupo de Usos e Arquitetura

RELATORIA:
Maíra Moraes, subgrupo de Custos
Patricia Oliveira, subgrupo de Conteúdo
Raquel Tsukada Lehmann, subgrupo de Usos e Arquitetura

APRESENTAÇÃO

É comumente atribuída a Einstein a frase “nem tudo o que conta pode ser contado”, ou seja, nem tudo aquilo que é importante pode ser mensurado. No entanto, não é possível saber se algo que é importante melhorou se não sabemos como mensurar. Os seres humanos têm criatividade suficiente para aprenderem a mensurar aquilo que lhes é valioso. E se existe algo que é realmente valioso e que, por essa razão, precisamos aprender a mensurar é a garantia do direito à educação e, em particular, do direito ao aprendizado. Como consta na Declaração Universal dos Direitos Humanos, a educação é um direito fundamental da pessoa humana e um meio indispensável para realizar todos os outros direitos.

Medir não é uma tarefa fácil; medir de forma confiável e precisa é ainda mais difícil. Mas medir é preciso quando queremos assegurar o direito à educação a cada indivíduo. Ao longo das últimas décadas, desenvolvemos no Brasil e no mundo um impressionante sistema para a avaliação da Educação Básica (por exemplo, PISA, SAEB, Prova Brasil, ANA, ENEM, entre outros). Nossa capacidade de identificar e diagnosticar desafios educacionais e, consequentemente, nossa capacidade de formular metas e planejar foi impulsionada de forma espetacular por essas avaliações.

É com base nessas avaliações que hoje temos a capacidade de formular políticas públicas com base em evidência e são elas que permitem que essas ações sejam criteriosamente e cientificamente avaliadas quanto à sua eficácia, eficiência, pertinência dentre outros aspectos. Mais importante, porém, tem sido o crescente uso pedagógico dessas avaliações na busca por adequar as práticas e metodologias às necessidades de cada escola e de cada aluno, assim como seu uso para direcionar os programas de formação do corpo docente, identificando lacunas e apontando rumos.

Além disso, a visão presente em muitos sistemas educacionais e na nossa Base Nacional Curricular Comum em construção – de que a educação deve ir além da transmissão de conhecimento como fim em si mesmo, devendo promover o desenvolvimento de competências do estudante numa perspectiva de formação integral, – amplia o leque de conteúdos a serem avaliados. De fato, assim como a educação que tem como propósito o desenvolvimento integral precisa ultrapassar a fronteira do conhecimento e dos aspectos puramente cognitivos, a avaliação do desenvolvimento do estudante requer também ir além da mensuração desses aspectos. Por essa razão, vem ganhando cada vez mais presença nas avaliações dentro e fora do Brasil a mensuração de habilidades socioemocionais, de abordagens dos alunos para a aprendizagem, solução criativa de problemas etc.

O desenvolvimento de medidas e a sua implementação como parte de um sistema de avaliação do desempenho educacional não é nem uma tarefa fácil nem uma tarefa que possa ser alcançada sem a necessidade de recursos financeiros. Em cenário em que se decida avaliar todos os alunos dos anos que encerram ciclos escolares (5º e 9º anos do Ensino Fundamental e 3ª série do Médio), o que corresponderia a avaliar cerca de 1/4 dos 35 milhões de alunos matriculados nos Ensinos Fundamental e Médio, seria preciso investir um importante volume de recursos.

Dada a natureza descentralizada do sistema educacional brasileiro, o desenvolvimento da avaliação educacional vem ocorrendo de forma simultânea tanto em nível federal, graças aos fantásticos avanços tanto promovidos ou liderados em nível federal pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – INEP, como em nível estadual. Há municípios que aderem a avaliações estaduais e uma minoria, sobretudo capitais, que desenvolve sua própria. De fato, hoje no Brasil apenas quatro unidades da federação não realizaram em 2016, ou em algum ano recente, alguma avaliação própria do desempenho educacional.

O desenvolvimento desse rico sistema de avaliações, no entanto, vem ocorrendo de forma independente entre os estados e ainda carente de maior articulação entre os sistemas de avaliação estaduais e o sistema federal. Ademais, as avaliações têm sido utilizadas para os mais diversos fins e, a despeito desse uso intensivo, graças à sua imensa riqueza informacional, permanecem, em certa medida, ainda subutilizadas. Os ganhos de aprimorar a conectividade e a cooperação entre os sistemas estaduais – incluindo o envolvimento dos municípios – e entre os sistemas estaduais e o sistema federal, são enormes.

Para tanto, o Artigo 11º do Plano Nacional de Educação estabelece que o País deve desenhar e implantar um Sistema Nacional de Avaliação da Educação Básica – SINAEB: “O Sistema Nacional de Avaliação da Educação Básica, coordenado pela União, em colaboração com os estados, o Distrito Federal e os municípios, constituirá fonte de informação para a avaliação da qualidade da educação básica e para a orientação das políticas públicas desse nível de ensino”. Assim, uma estreita cooperação e coordenação entre os sistemas estaduais, e entre esses e o sistema federal, irá permitir que experiências e melhores práticas sejam mais prontamente difundidas, que as informações geradas pelos diversos sistemas sejam comparáveis, e que sinergias e economias de escala possam ser aproveitadas, permitindo que os mesmos resultados possam ser alcançados a um menor custo.

Por todas essas razões, foi com grande satisfação e entusiasmo que nossas organizações receberam o convite para apoiar técnica e financeiramente o Grupo de Trabalho (GT) criado pelo Consed para repensar os sistemas estaduais de avaliação, incluir novos conteúdos avaliativos na perspectiva do desenvolvimento de competências, trocar experiências e procurar um desenho que, embora flexível, acomodasse as especificidades de cada unidade da federação e permitisse que as sinergias e possibilidades de economias de escala fossem aproveitadas,

reduzindo esforço e custo para todos. Desde a constituição do GT de avaliação procuramos apoiar ao máximo, técnica e financeiramente, os trabalhos realizados.

Reconhecemos e parabenizamos o CONSED pelo amplo sucesso alcançado pelo GT na direção das metas e objetivos estabelecidos. Este relatório síntese das atividades do GT apresenta um valioso diagnóstico da situação atual dos sistemas estaduais de avaliação da educação, assim como uma proposta para um sistema de avaliação que alinhe os conteúdos da avaliação aos objetivos de desenvolvimento de competências dos estudantes, promova uma maior cooperação entre as unidades da federação e que permita maior comparabilidade entre os resultados e reduções de custo pelo aproveitamento de sinergias e economias de escala.

Esperamos que o sucesso dos trabalhos do GT sirva para ressaltar, por um lado, a importância de uma estreita cooperação entre as unidades da federação e, por outro, a viabilidade técnica e financeira de desenvolvermos um amplo sistema de avaliação dos resultados educacionais de forma que aquilo que realmente importa (a garantia do direito à educação para todos) possa ser efetivamente mensurado.

Viviane Senna, Presidente do Instituto Ayrton Senna
Ricardo Henriques, Superintendente Executivo do Instituto Unibanco
Angela Dannemann, Superintendente da Fundação Itaú Social

SUMÁRIO

Introdução: Repensando os Sistemas Estaduais de Avaliação da Educação Básica 11

Cap. 1: A experiência recente das unidades da federação em Avaliação da Educação Básica 17

Cap. 2: Propostas para o Sistema de Avaliação da Educação Básica 27

Cap. 3: Conteúdos possíveis para um Sistema de Avaliação da Educação Básica 39

Cap. 4: Propósitos, usos, utilidades e potenciais consequências negativas de um Sistema de Avaliação da Educação Básica 69

Cap. 5: Arquitetura do Sistema de Avaliação da Educação Básica 97

Cap. 6: Economias de escala e oportunidades para a redução nos custos da Avaliação da Educação Básica 111

Cap. 7: Resumo das reuniões técnicas do Grupo de Trabalho de Avaliação do Consed 123

INTRODUÇÃO

REPENSANDO OS SISTEMAS ESTADUAIS DE AVALIAÇÃO DA EDUCAÇÃO BÁSICA

REPENSANDO OS SISTEMAS ESTADUAIS DE AVALIAÇÃO DA EDUCAÇÃO BÁSICA

Nas últimas décadas, o governo federal, quase todos os governos estaduais e alguns governos municipais implantaram um vigoroso e moderno sistema de avaliação da educação básica. O desenho, o desenvolvimento e a implantação de um sistema dessa natureza é requisito legal: de acordo com o Inciso VI do Artigo 9º da Lei de Diretrizes e Bases da Educação Nacional – LDB, "*A União incumbir-se-á de 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".* Além disso, a possibilidade e a importância da parceria entre as unidades federadas nessa questão ficam claras no § 5º do Artigo 11º do Plano Nacional de Educação: "*A avaliação de desempenho dos(as) estudantes em exames, ... poderá ser diretamente realizada pela União ou, mediante acordo de cooperação, pelos estados e pelo Distrito Federal, nos respectivos sistemas de ensino e de seus municípios, caso mantenham sistemas próprios de avaliação do rendimento escolar, assegurada a compatibilidade".*

Esses preceitos legais certamente renderam os resultados esperados. Graças ao desenvolvimento de um sólido sistema de avaliação do desempenho escolar, hoje podemos monitorar com grande detalhe o desempenho educacional do país, dos estados, e também, em boa medida, dos municípios. O uso desse sistema tem se intensificado e diversificado enormemente ao longo da última década. As informações do sistema têm sido imprescindíveis à definição de metas e ao seu acompanhamento e, consequentemente, são utilizadas como uma forma de prestação de contas. Para além disso, essas informações têm sido indispensáveis à realização de diagnósticos e ao planejamento, com destaque especial à formulação dos Planos Nacional, Estaduais e Municipais de Educação. Seu uso pedagógico também vem se expandindo, permitindo aprimorar e melhor adequar as metodologias e as práticas pedagógicas às necessidades locais.

No entanto, como a construção desse sistema nem sempre pode se basear numa articulação perfeita entre os diversos entes federados envolvidos, persistem sobreposições e lacunas que requerem uma revisão dos desenhos adotados em cada unidade da federação e pelo governo federal. Além disso, a própria forma de financiamento e contratação dos serviços de avaliação, por ter resultado de ações independentes dos diferentes entes federados, deixou de beneficiar de potencias sinergias e economias de escala. Dessa forma, uma melhor coordenação entre as ações avaliativas dos governos estaduais e federal permitiria reduzir custos e compartilhar experiências e melhores práticas, relacionadas à implementação dessas avaliações ao seu uso, em particular, na função pedagógica.

Por conta desse cenário, é preciso desenvolver um novo desenho tanto para os sistemas de avaliação de cada unidade da federação como para a cooperação, coordenação e troca de

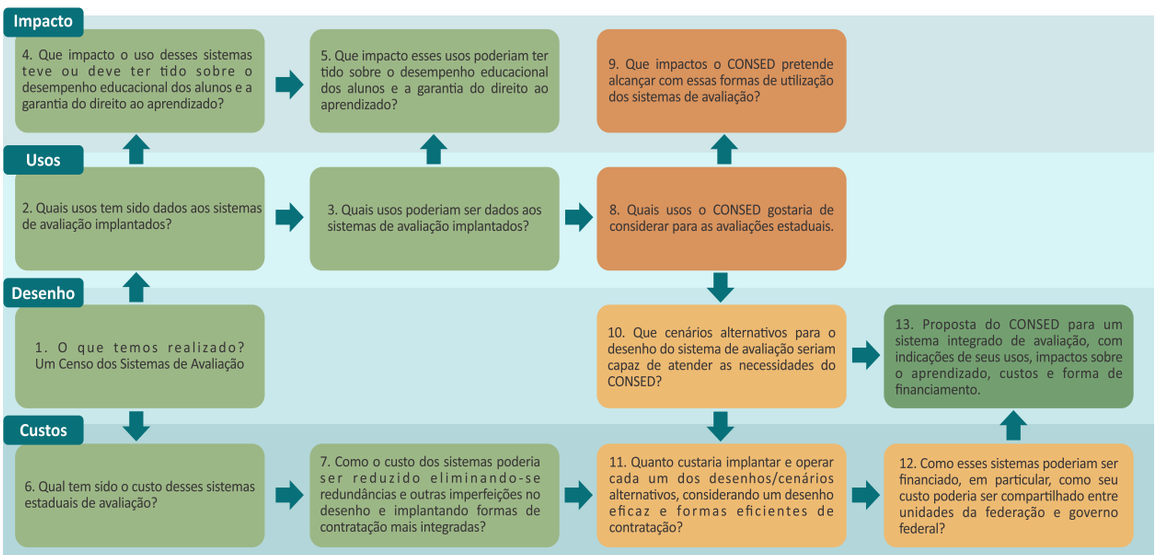
experiências entre as unidades da federação e entre elas e o governo federal. O novo sistema deve buscar ser mais bem coordenado e mais eficaz na promoção do direito ao aprendizado, preenchendo lacunas, ampliando a cobertura, aprimorando a comparabilidade ou adicionando componentes (que pode ser um conteúdo, uma nova série avaliada etc.) e, viabilizando, dessa forma, novos usos da avaliação, especialmente aqueles voltados diretamente à melhoria do aprendizado. Um novo sistema deve buscar eliminar redundâncias e ineficiências, além de promover maior articulação tanto interna quanto entre os diversos sistemas dos entes federados, possibilitando, assim, a redução de custos via o aproveitamento de sinergias, melhores práticas e economias de escala.

O momento atual, de construção da Base Nacional Comum Curricular (BNCC), requer a revisão dos conteúdos a serem avaliados. Em termos bem simples, um currículo é "uma declaração oficial sobre o que os estudantes devem saber e poder fazer"[1]. A BNCC afirma o compromisso com a promoção da educação integral – entendida como aquela que proporciona o desenvolvimento global do aluno, indo muito além da mera transmissão de conhecimento e de aspectos puramente cognitivos – e explicita o objetivo de desenvolvimento de competências do estudante. Nesse sentido, avaliar se o direito à educação está sendo de fato garantido requer avaliar se os objetivos de desenvolvimento de competências estão sendo alcançados. Para tanto, a avaliação precisa incluir novos conteúdos, que permitam acompanhar devidamente o desenvolvimento do estudante numa perspectiva integral.

Com esses desafios em mente, o Consed criou o Grupo de Trabalho – Avaliação. O GT foi constituído com o objetivo de fazer um diagnóstico das experiências das diversas unidades da federação com sistemas de avaliação, incluindo a análise: *i)* do conteúdo (o que é avaliado); *ii)* da arquitetura (quem deve ser incluído na avaliação, quais redes e séries, e com que frequência as avaliações devem ocorrer); *iii)* dos usos que têm sido feitos dessas avaliações; e *iv)* da estrutura de custos das avaliações, com particular atenção à possibilidade de economias de escala que parcerias entre os entes federados poderia trazer.

A missão do GT, para além de realizar esse diagnóstico da situação atual, incluía apresentar uma proposta para um novo sistema que buscasse, por um lado, modernizar e ampliar o escopo dos sistemas existentes, e, por outro, aproveitar o mais plenamente possível as sinergias e economias de escala que uma efetiva cooperação entre unidades da federação e governo federal poderiam trazer, conforme se percebe pela estrutura conceitual das questões tratadas (veja o quadro abaixo).

[1] *The SAGE Handbook of Curriculum and Instruction* (2008; p. 8).



Este relatório apresenta uma síntese dos resultados de 18 meses de trabalho do Grupo de Trabalho – Avaliação.

ORGANIZAÇÃO DO DOCUMENTO

O primeiro capítulo apresenta a experiência recente das unidades da federação na avaliação da educação básica. Como primeira atividade do Grupo de Trabalho – Avaliação, no início de 2016, foi realizado Censo das Avaliações Estaduais, com o propósito de gerar insumos para as posteriores discussões e formulações do GT. Foram coletadas informações sobre as experiências de todos os estados e Distrito Federal com avaliação e também os planos de realização futura dos entes federados.

O segundo capítulo apresenta proposta para aprimoramento dos sistemas estaduais de avaliação da educação básica elaborada pelo Grupo de Trabalho – Avaliação, que foi apreciada, discutida e validada pelos secretários estaduais em dois momentos: *i)* coletivamente nas reuniões do Consed; e *ii)* individualmente em pesquisa específica conduzida pelo GT. Para permitir a adequação da avaliação às múltiplas realidades das nossas unidades da federação, o GT chegou a desenho flexível, que vai de formatos mais extensos a mais enxutos para a avaliação, dando liberdade para cada estado escolher o modelo que considerar mais adequado, mas, ao mesmo tempo, estabelecendo parâmetros mínimos para a avaliação, que viabilizam a comparabilidade dos resultados, além de arranjos colaborativos entre as UFs para a realização das avaliações.

Os quatro capítulos seguintes (3 a 6) referem-se a elaborações realizadas pelo Instituto Ayrton Senna – apoiador técnico do GT, que tiveram como propósito informar e inspirar os membros do Grupo acerca dos possíveis: *i)* conteúdos avaliativos; *ii)* usos para avaliação, incluindo os riscos de mau uso das avaliações e cuidados que devem ser tomados ao avaliar e tornar públicos os resultados da avaliação, *iii)* variações na arquitetura do sistema de avaliação e o propósito de cada tipo de arranjo; e *iv)* ganhos de escala a serem explorados pela maior articulação entre os diversos sistemas de avaliação estaduais.

O sétimo e último capítulo apresenta as discussões realizadas no âmbito do Grupo de Trabalho – Avaliação, sendo que a primeira seção sintetiza a discussão por blocos temáticos e a segunda apresenta um relato detalhado do formato e do teor de cada reunião técnica do GT.

CAPÍTULO 1

A EXPERIÊNCIA RECENTE DAS UNIDADES DA FEDERAÇÃO COM AVALIAÇÃO DA EDUCAÇÃO BÁSICA

A EXPERIÊNCIA RECENTE DAS UNIDADES DA FEDERAÇÃO COM AVALIAÇÃO DA EDUCAÇÃO BÁSICA

INTRODUÇÃO

O Censo dos Sistemas Estaduais de Avaliação da Educação Básica, realizado em 2016 em todos os estados e no Distrito Federal, constitui importante fonte de informação sobre as avaliações de desempenho dos estudantes no Brasil[1]. O Censo teve como propósito gerar insumos para o Grupo de Trabalho (GT) instituído pelo Consed, cuja missão é aprimorar os sistemas estaduais de avaliação da educação básica, buscando um desenho que, embora flexível, acomode as especificidades de cada unidade da federação (UF), contemple as novas necessidades de conteúdos avaliativos e permita que as sinergias e possibilidades de economias de escala sejam aproveitadas, reduzindo esforço e custo para todos.

Foram coletadas informações sobre as experiências das diversas UFs com sistemas de avaliação nos anos de 2014 e 2015 e também com relação aos seus planos de realização futura, incluindo elementos tanto do conteúdo (o que avaliar) e arquitetura (quem é incluído na avaliação – quais redes e séries, e com que frequência essas avaliações ocorrem), como dos usos que essas avaliações têm recebido e sua estrutura de custos, com particular atenção para a possibilidade de economias de escala que parcerias entre entes federados poderiam trazer.

Segundo o Censo, das 27 UFs apenas dez não realizaram avaliações em 2014 e 2015, e apenas quatro dentre estas nunca realizaram uma avaliação. Ademais, o número de UFs que não realizam avaliações vem declinando. Em 2015 dez UFs não realizaram avaliação, mas, segundo os planos declarados no Censo pelas Secretarias de Educação esse número vai ser cada vez menor.

A utilidade das informações coletadas com o Censo, no entanto, transcende sua utilização dentro do GT. De fato, prefeitos, governadores, órgãos municipais e estaduais de educação, investidos hoje de maior autonomia e de novas responsabilidades, que têm ou pretendem ter algum sistema de avaliação, podem se beneficiar dessas informações em seus processos de planejamento e tomada de decisões.

[1] Outra referência importante é o relatório publicado em 2011, com o apoio da Fundação Victor Civita, que mostra o papel cada vez mais importante das avaliações subnacionais para a política educacional. GAME/FAE/UFMG (2011). *A avaliação externa como instrumento educacional nos estados*. Relatório Final.

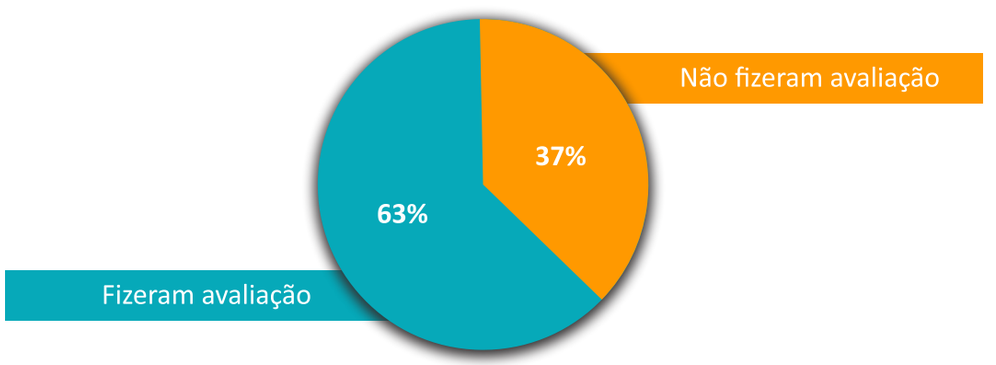
O presente capítulo tem como objetivo apresentar os principais resultados obtidos a partir do Censo dos Sistemas Estaduais de Avaliação. Seu conteúdo foi organizado seguindo, em grande medida, a estrutura de coleta de informações do próprio Censo[2].

RESULTADOS DO CENSO DOS SISTEMAS ESTADUAIS DE AVALIAÇÃO

QUEM FAZ AVALIAÇÃO

Todos os estados e o Distrito Federal responderam ao Censo dos Sistemas Estaduais de Avaliação da Educação Básica. Segundo as informações coletadas, apenas dez UFs não realizaram avaliações em 2014 e 2015, e apenas quatro dentre essas nunca realizaram uma avaliação.

Figura 1: Realização de avaliações pelas Unidades da Federação em 2014-2015



DISCIPLINAS, ANOS E SÉRIES CONTEMPLADOS NA AVALIAÇÃO

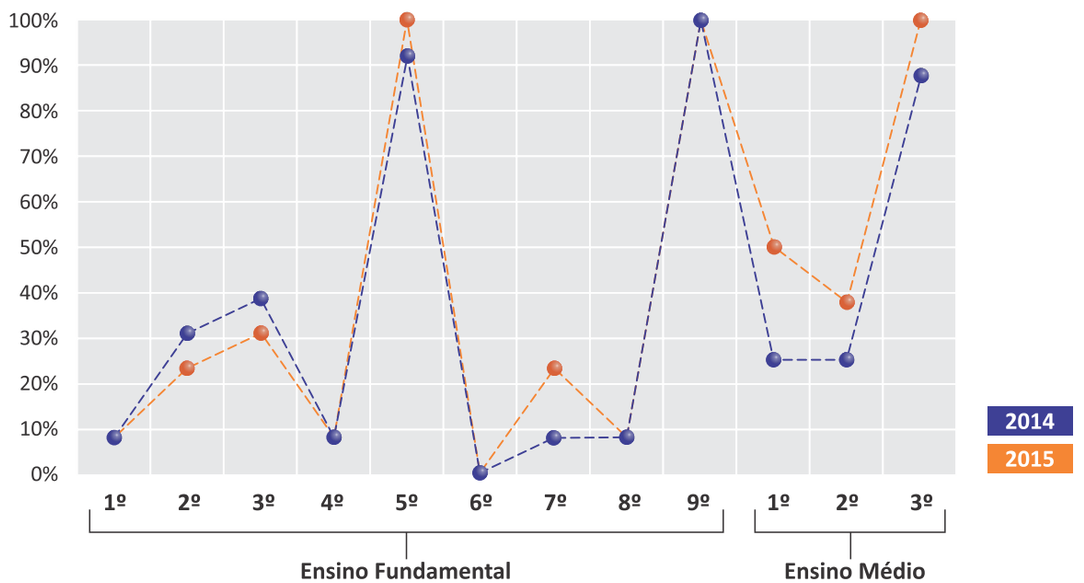
Todas as avaliações realizadas ou planejadas incluem as disciplinas de Língua Portuguesa e Matemática, e nove das 14 UFs que realizaram avaliações em 2014 e 2015 avaliaram as mesmas disciplinas e séries nas duas edições. Há, no entanto, para o conjunto das UFs, experiência com a avaliação de todas as disciplinas curriculares, exceto educação artística e educação física.

Tipicamente, como mostra a Figura 2, as avaliações estão concentradas no 5º e 9º anos do Ensino Fundamental e na 3ª série do Ensino Médio. Há, no entanto, experiência com

[2] Uma cópia do questionário do Censo dos Sistemas Estaduais de Avaliação encontra-se em anexo.

avaliações de Língua Portuguesa e Matemática em todos os anos e séries da Educação Básica (com exceção da 6ª série).

Figura 2: Avaliações em 2014-2015 por séries, dentre as Unidades da Federação que realizaram avaliação



MOMENTO DA AVALIAÇÃO

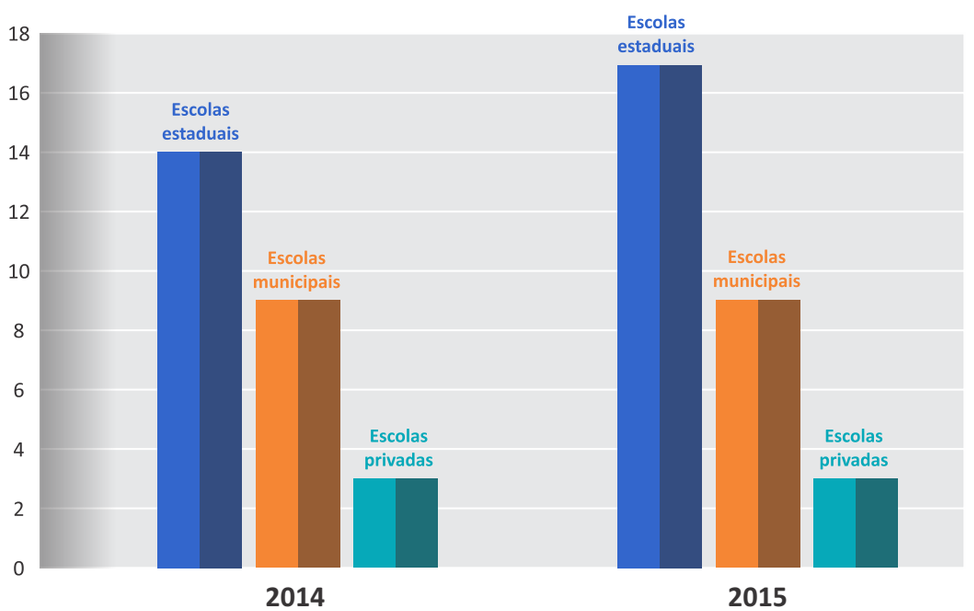
Em sua grande maioria essas avaliações ocorrem com periodicidade anual, no final do ano letivo. Em apenas duas UFs as avaliações foram realizadas também no início do ano letivo. As avaliações formativas[3], que têm periodicidade superior à anual, entretanto, são raras com apenas três UFs tendo realizado durante o ano letivo (bimestrais, trimestrais ou semestrais) e, segundo informações do Censo, apenas duas UFs planejam realizar esse tipo de avaliação no futuro. Vale ressaltar que, quando há avaliações formativas durante o ano letivo (bimestrais ou trimestrais), estas abrangem todas as séries do Ensino Médio.

[3] "Avaliação conduzida ao longo do processo educacional com o objetivo de potencializar a aprendizagem do estudante. Implica buscar evidências sobre a aprendizagem a fim de fechar o hiato entre desempenho atual e o desejado (de modo a permitir ações para fechar esse hiato); oferecer *feedback* aos estudantes; e envolver estudantes no processo de avaliação e aprendizagem (Fonte: CCSSO, 2008)". IBE Glossário de Terminologia Curricular. UNESCO, 2016.

ESCOLAS INCLUÍDAS NA AVALIAÇÃO

Tradicionalmente, mais da metade dos sistemas de avaliação estaduais existentes cobrem também as escolas municipais. Entretanto, praticamente todas as UFs que planejam realizar avaliações no futuro e que não avaliaram 2015 não têm a intenção de incluir as escolas municipais. Vale ressaltar que, quando a avaliação compreende os municípios, os anos avaliados são os mesmos do Ensino Fundamental incluídos na rede estadual. Com relação à inclusão das escolas particulares na avaliação, no entanto, o Censo revelou que apenas quatro UFs incluem ou planejam incluí-las. Quando estas são incluídas, comumente, os mesmos anos e séries são avaliados.

Figura 3: Avaliações em 2014-2015 por rede de ensino, dentre as Unidades da Federação que realizaram avaliação



AVALIAÇÃO CENSITÁRIA OU POR AMOSTRAGEM

Em sua grande maioria as avaliações são universais. Em apenas duas UFs as avaliações não foram universais em 2015, o mesmo estando previsto para 2016 segundo os planos declarados pelas Secretarias de Educação no Censo. Três UFs combinam avaliação universal na maioria dos anos e séries com avaliação amostral em algumas poucas séries, tipicamente, aquelas intermediárias, como o 7º ano do Ensino Fundamental e a 1ª série do Ensino Médio.

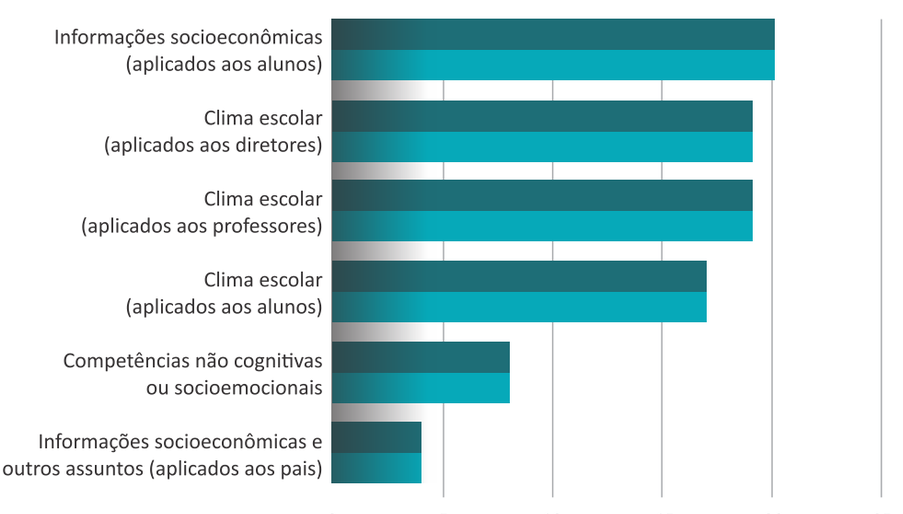
CONTEÚDOS AVALIATIVOS PARA ALÉM DAS DISCIPLINAS CURRICULARES

Além do desempenho dos estudantes, muitas UFs avaliam também as competências socioemocionais. De fato, quando perguntadas sobre os instrumentos de coleta e temas presentes na última avaliação estadual realizada, oito UFs afirmaram já ter realizado alguma avaliação dessa natureza.

Já com relação aos aspectos socioeconômicos das famílias dos estudantes, em 20 UFs as avaliações incluíram esse tipo de informação. Em alguns casos essas informações são obtidas diretamente com os pais.

Outro tema bastante investigado nas avaliações realizadas é o clima escolar. Em 19 UFs foi realizada alguma avaliação do clima escolar ou da dedicação dos professores, incluindo entrevistas com os mesmos (19 UFs), com os diretores (19 UFs) e com alunos (17 UFs).

Figura 4: Temas presentes na última avaliação realizada



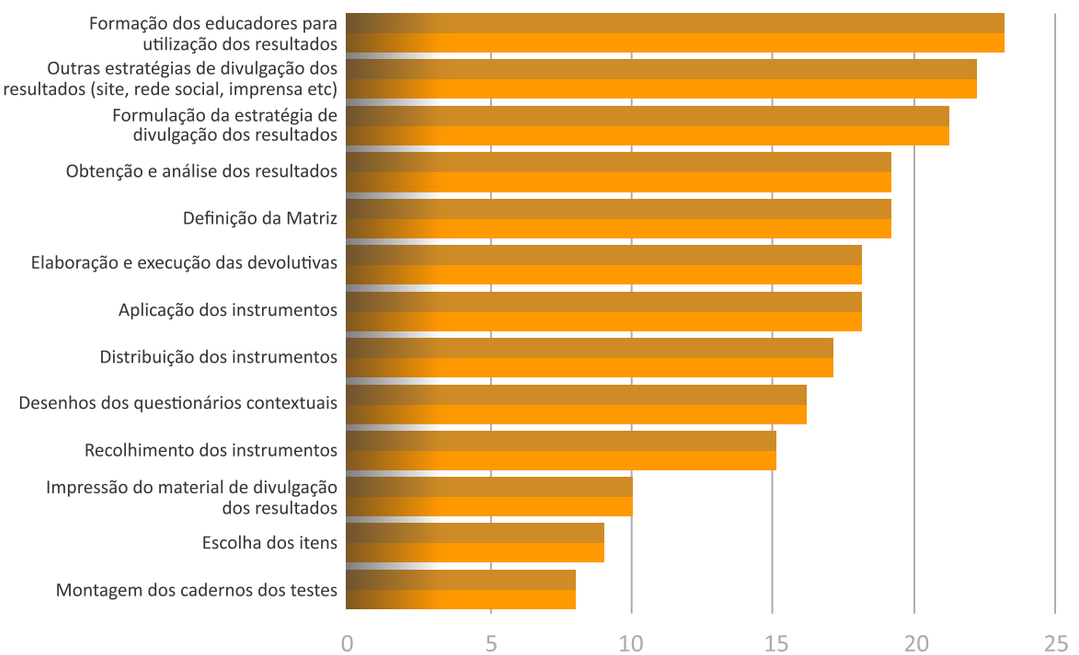
O PAPEL DAS SECRETARIAS DE EDUCAÇÃO NA AVALIAÇÃO

Todas as Secretarias de Educação que já realizaram alguma avaliação têm equipe dedicada em tempo parcial ou integral, e em mais de ¼ das UFs a legislação torna a avaliação compulsória e com periodicidade definida.

Quando perguntadas sobre a participação da equipe da Secretaria responsável pelo sistema de avaliação em diversas atividades, integral ou parcialmente, mais de ¾ das UFs declarou que essa equipe participa das seguintes atividades: *i)* definição da matriz e do desenho dos questionários contextuais; *ii)* distribuição, aplicação e recolhimento dos instrumentos de avaliação; *iii)* formulação da estratégia de divulgação dos resultados e análises; *iv)* elaboração e execução das devolutivas e formação dos educadores para utilização dos resultados e das análises produzidas.

Em menos da metade das UFs a equipe da Secretaria de Educação participa: *i)* da escolha dos itens e da montagem dos cadernos dos testes; e *ii)* da impressão do material de divulgação dos resultados e análises.

Figura 5: Participação das equipes da Secretaria de Educação na avaliação



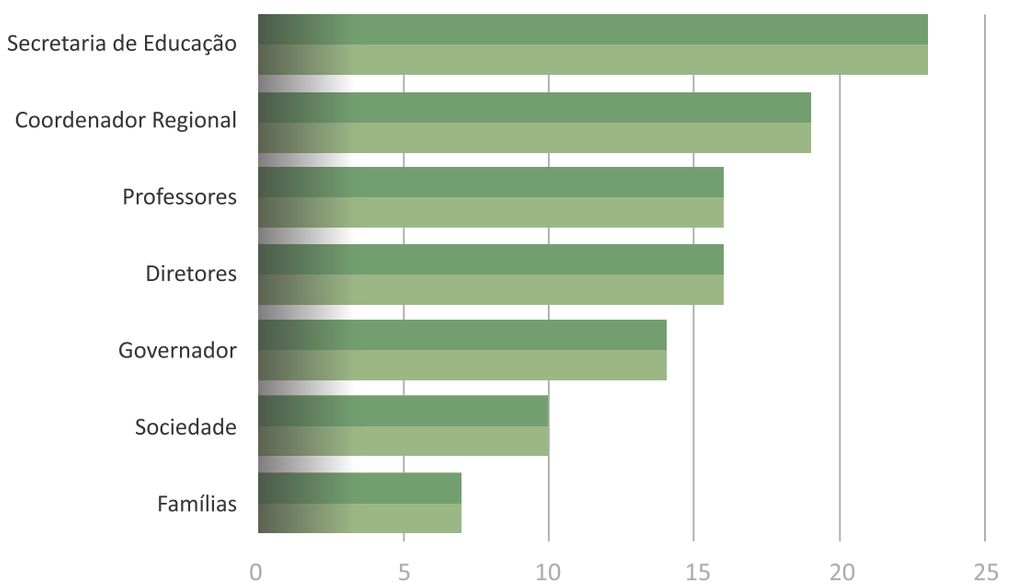
MENSURAÇÃO, DIVULGAÇÃO E UTILIZAÇÃO DOS RESULTADOS

Em 60% das UFs que têm sistema de avaliação, os resultados estão na mesma escala das avaliações federais (como Prova Brasil e Sistema de Avaliação da Educação Básica (SAEB)), e em todas as UFs onde há avaliação os resultados são apresentados para os professores, diretores, regionais e para a Secretaria. Um resultado importante de se observar, mostrado na Figura 6, é

a apresentação oficial dos resultados das avaliações também para as famílias (7 UFs), para a sociedade (10 UFs) e para o Governador do Estado (14 UFs).

Adicionalmente, em praticamente todas as UFs onde há avaliação, os resultados são utilizados para aprimorar a formação dos professores, e em cerca de 1/3 delas esses resultados são também empregados para definir a bonificação de professores e diretores.

Figura 6: Apresentação oficial de resultados das avaliações



PERSPECTIVAS FUTURAS PARA AS AVALIAÇÕES

Com relação ao escopo das avaliações futuras os resultados obtidos com base no Censo revelam que 17 Ufs, dentre as que já possuem um sistema de avaliação, pretendem ampliar o seu escopo. De fato, metade das UFs que já contam com um sistema de avaliação pretende incluir novas disciplinas, incluir medidas das competências socioemocionais, ou ampliar os anos e as séries cobertos pela avaliação.

Já no que diz respeito à cobertura e periodicidade, 10 UFs pretendem ampliar a cobertura incluindo escolas municipais, mas apenas 4 UFs pretendem expandir a cobertura para incluir também as escolas particulares. Entretanto, com relação à periodicidade ou a natureza censitária da avaliação, poucas declaram intenção de modificá-las.

Por fim, 15 UFs que realizam avaliações revelaram preocupação em reduzir seus custos no futuro, mas a imensa maioria (19 UFs) demonstra interesse em aprimorar a divulgação e a difusão dos resultados, além de ampliar suas possibilidades de utilização por meio, por exemplo, da identificação dos alunos ao longo dos anos, que viabiliza que avaliação somativa[4] de um ano sirva como diagnóstica para momentos subsequentes (16 UFs).

[4] "Determinação do desempenho do aluno ao final de um período, estágio, curso ou programa, usualmente – mas não necessariamente – envolvendo provas ou exames formais. Avaliações somativas são mais comumente usadas para classificar, dar notas e/ou aprovar estudantes, assim como para propósitos de certificação." IBE Glossário de Terminologia Curricular. UNESCO, 2016.

CAPÍTULO 2

PROPOSTAS PARA O SISTEMA DE AVALIAÇÃO DA EDUCAÇÃO BÁSICA

PROPOSTAS PARA O SISTEMA DE AVALIAÇÃO DA EDUCAÇÃO BÁSICA

INTRODUÇÃO

A definição de um modelo de Sistema de Avaliação da Educação Básica que seja capaz de funcionar em vários estados, respeitando a autonomia de cada ente federativo, com amplitude para oferecer respostas sobre o complexo conceito de desempenho educacional, reconhecendo os diversos contextos sociais, econômicos e políticos, é um dos maiores desafios da Educação Pública. Devemos buscar uma maior articulação entre as avaliações nacionais, estaduais e municipais, que possibilitarão maior acesso aos dados, buscando maior eficiência em relação aos custos.

A avaliação não pode ser encarada como uma etapa final da política educacional, pois seria um erro estratégico imaginar que a avaliação meramente indica o nível de qualidade da educação ofertada pelas redes e sistemas de ensino. É preciso ter consciência do caráter fundamental da avaliação enquanto base para o planejamento de novas ações e correção de rumo, quando observados resultados indesejados que reflitam falhas na política educacional. A avaliação funciona, portanto, como bússola para a revisão periódica à qual todas as ações de governo devem se submeter para garantir o esforço adequado, a aplicação otimizada de recursos e as decisões corretas e necessárias para a melhoria contínua da Educação Básica. As avaliações podem ser importantes ferramentas para o planejamento estratégico e para a elaboração de planos de ação de gestão e pedagógicos, fortalecendo as redes de ensino e melhorando os resultados.

A consciência da importância e complexidade da avaliação da Educação Básica e o reconhecimento do caráter democrático que deve revestir a concepção de políticas educacionais no país foram diretrizes determinantes para o processo de trabalho que baseou a construção da proposta que aqui apresentaremos para o Sistema de Avaliação da Educação Básica.

1. CONHECENDO NOSSO PONTO DE PARTIDA

Em outros capítulos desta obra, pode-se observar os aspectos predominantes no contexto atual da avaliação da Educação Básica no Brasil. Neste momento, nos dedicaremos a discorrer sobre o quanto esse contexto influencia a visão dos estados na construção de um sistema de avaliação e quais pressupostos mínimos devem ser garantidos para que um novo desenho não prejudique as conquistas já alcançadas em várias unidades da federação. Destacamos a importância de ampliar o acesso aos dados gerados pelas diferentes avaliações aplicadas em todas as redes.

A realização de um Censo da Avaliação nos estados pelo GT de Avaliação do Consed, com o apoio e brilhantismo do Instituto Ayrton Senna, constituiu-se em marco essencial para a devida compreensão dos instrumentos empregados pelos estados para avaliação de suas redes, em muitos casos de redes municipais e, eventualmente, mediante adesão, de escolas das redes privadas de ensino. A leitura que derivou desse levantamento ajudou, não somente a definir quais temáticas deveriam ser observadas no desenvolvimento dos trabalhos dos subgrupos do GT de Avaliação, mas também nos mostra o mínimo que devemos preservar. O Sistema de Avaliação da Educação Básica deve trazer novas ferramentas e possibilidades, mas também resguardar as conquistas dos estados e municípios e possibilitar que os diversos relatórios atualmente gerados e utilizados pelas redes como devolutivas pedagógicas e instrumentos de gestão encontrem amparo no novo modelo e amplo acesso a informações necessárias à consolidação das políticas educacionais de cada ente.

Embora observe-se diferenças marcantes, ao longo do país, na forma e profundidade empregadas na utilização das devolutivas pelos estados e municípios, é bastante perceptível que secretarias estaduais e municipais, pais, professores, gestores escolares, universidades e até mesmo órgãos de controle encontram, nas informações resultantes do processo de avaliação das redes, importantes subsídios para o exercício de suas respectivas funções.

A compreensão dessas particularidades é essencial para a definição de qual modelo de avaliação seria capaz de abarcar os principais anseios das redes estaduais, respeitando as particularidades regionais e garantindo uma marcante evolução da avaliação em todo o país.

QUAL A AVALIAÇÃO DESEJADA PELOS ESTADOS?

Embora a experiência e conhecimento sobre avaliação da educação seja bastante variada entre os estados, as discussões provocadas durante o GT de Avaliação ajudaram a fortalecer o amadurecimento de todas as instituições envolvidas.

Importante pontuar que os temas levantados durante os encontros do GT de Avaliação proporcionaram debates que sempre foram realizados sob o olhar atento dos respectivos Secretários e Secretárias de Educação, no âmbito do Fórum de Secretários do CONSED, envolvendo vários outros atores das secretarias estaduais que não compunham oficialmente o GT de Avaliação.

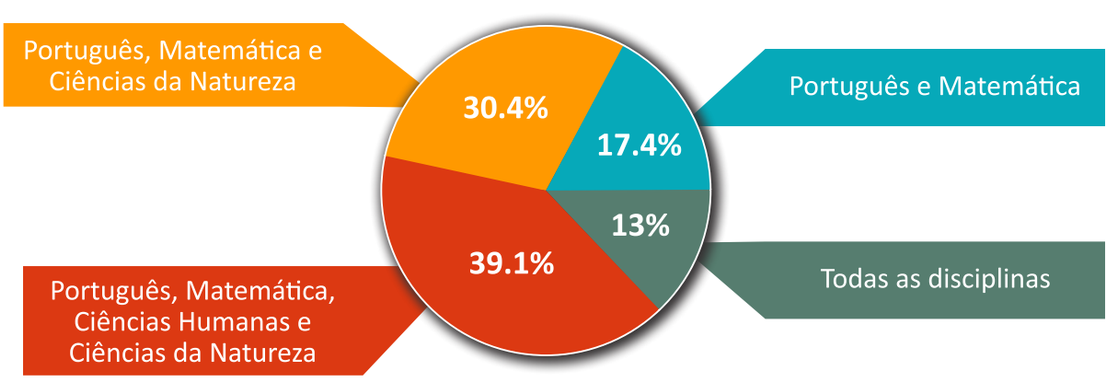
Esse amadurecimento sobre o tema foi essencial para a etapa mais importante do projeto de elaboração de uma proposta para o Sistema de Avaliação da Educação Básica: A escolha de diretrizes do modelo pelas secretarias de educação.

Para compreender as necessidades e anseios de cada estado sobre o formato e extensão da Avaliação, foram elaborados e distribuídos questionários para respostas dos Secretários de Educação contendo questionamentos sobre Conteúdo, Arquitetura e Usos da Avaliação da Educação. Os técnicos, enquanto representantes do GT de Avaliação nas secretarias, intermediaram discussões e apresentaram aos respectivos dirigentes as questões e implicações de cada escolha para o modelo[1]. O resultado da escolha pode ser retratado em alguns temas:

Conteúdo

Além do tradicional escopo presente nos sistemas de avaliação da maioria dos estados, que envolve originariamente a aplicação de avaliações em Língua Portuguesa e Matemática, os estados aspiram expandir as avaliações para comportar provas nas áreas de Ciências Humanas e Ciências da Natureza, conforme pode ser observado na figura abaixo:

Figura 1: Quais competências acadêmicas incluir?

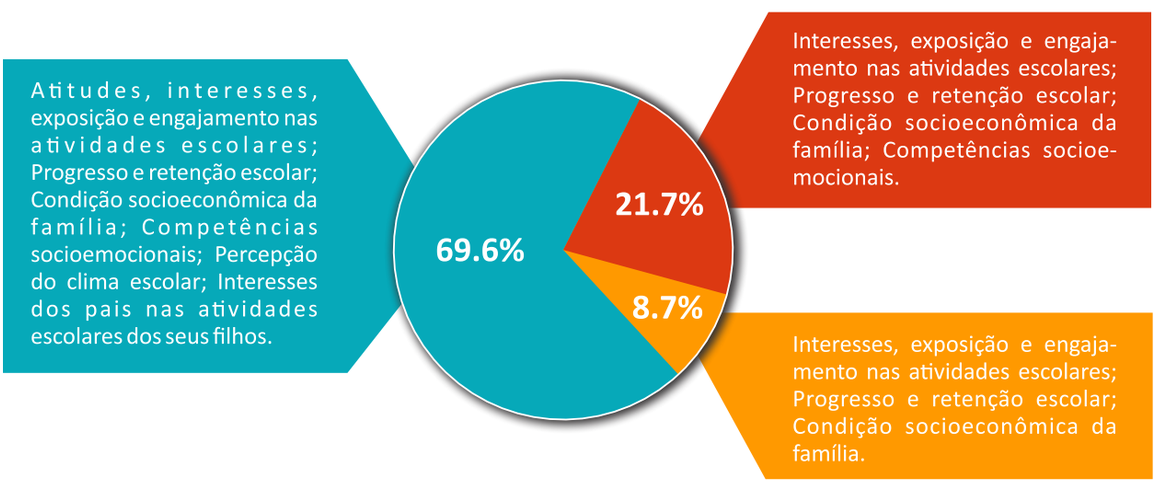


Nas discussões realizadas, muitos destacaram que a expansão do escopo dos sistemas de avaliação pode ser realizada de forma gradual, como forma de minimizar o impacto nos custos.

[1] A pesquisa realizada com os Secretários de Educação é distinta do Censo das Avaliações Estaduais. O Censo foi realizado no início dos trabalhos do GT, para colher insumos para as discussões e formulação da proposta. A pesquisa com os Secretários foi realizada como forma de discussão e validação da proposta formulada pelo GT.

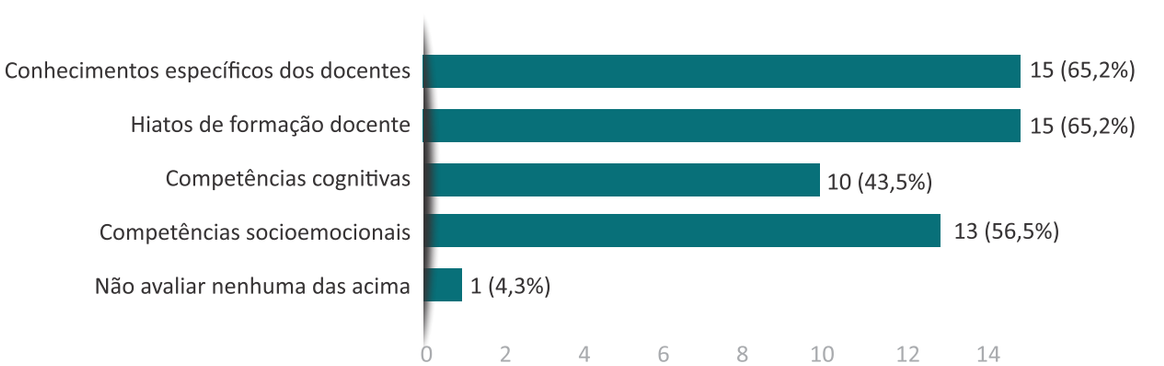
As reuniões do GT de Avaliação já antecipavam uma tendência de incluir as competências socioemocionais no universo das avaliações externas. A pesquisa com os Secretários consolidou esse entendimento.

Figura 2: Quais outras competências do estudante incluir?



A avaliação docente também figurou como um dos pontos bastante discutidos durante as reuniões do GT de Avaliação. O resultado da pesquisa com os Secretários pode ser analisado abaixo:

Figura 3: Quais tópicos incluir na avaliação docente?

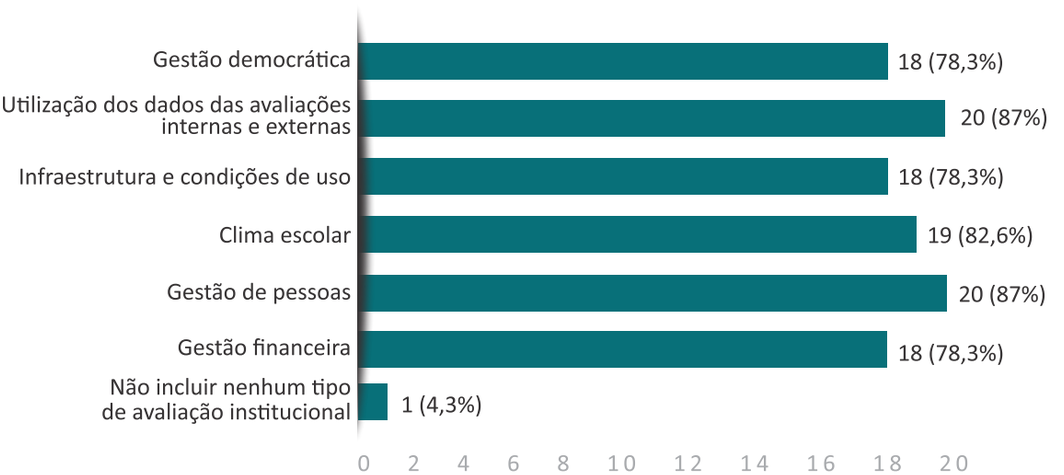


O resultado da pesquisa demonstra que os estados estão focados em avaliar quais conhecimentos específicos podem aprimorar o desempenho dos professores e contribuir para o processo de aprendizagem nas redes. Não há, por parte dos estados, interesse em avaliar

competências cognitivas do corpo docente. Percebe-se também a existência de menor consenso em relação à avaliação das competências socioemocionais dos professores.

Outro ponto abordado na pesquisa foi a avaliação institucional. Nesse aspecto, podemos observar abaixo que o tema tem ampla adesão dos estados:

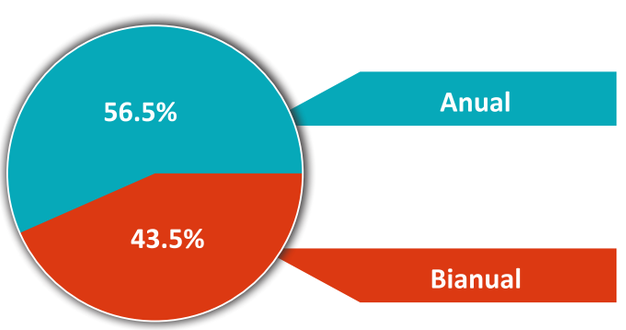
Figura 4: Quais tópicos incluir na avaliação institucional?



1. Arquitetura da Avaliação

Outra dimensão da avaliação retratada na pesquisa foi a Arquitetura desejada pelos estados para a Avaliação. A primeira questão abordada foi a periodicidade da Avaliação.

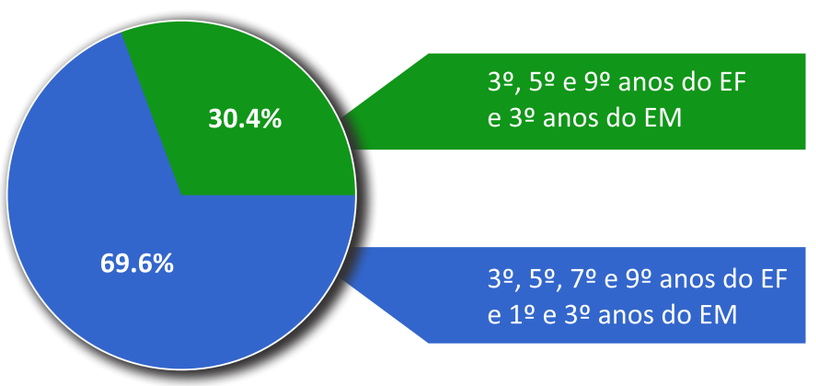
Figura 5: Qual deve ser a periodicidade da avaliação?



Conforme pode-se depreender da figura 5, a maioria dos estados apresenta preferência pela realização de avaliações todos os anos. Ressalte-se que a maior ou menor utilização dos dados obtidos pelas avaliações influencia diretamente na preferência do Estado por uma avaliação anual ou bianual. Os estados que, por diretriz de sua política educacional, utilizam microdados na formação de seus professores, elaboração de projetos pedagógicos e planos de ação de escolas, definição de estratégias de reforço, entre outras aplicações, demandam dados contínuos que se integram à rotina das escolas. Há estados, no entanto, que compreendem os resultados das avaliações como indicadores norteadores das políticas educacionais, não necessariamente utilizando-os como insumo contínuo do processo formativo. É importante destacar que nas discussões do GT de Avaliação e entre os Secretários, quando se inclui a possibilidade de construção de um Sistema de Avaliação em parceria entre MEC/INEP e estados, minimizando os custos e ampliando o acesso aos dados, a opção por avaliações anuais cresce bastante.

Tema complexo, porém essencial para a determinação da Arquitetura da Avaliação, é quais anos devem ser foco das avaliações.

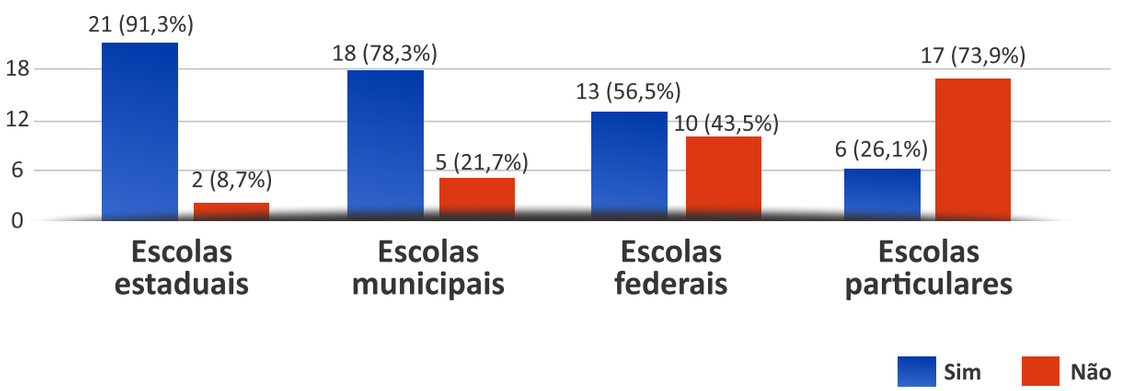
Figura 6: Quais anos devem ser incluídos?



Resultado surpreendente observou-se na escolha pelas secretarias de educação de um rol maior de anos para a avaliação. Esse certamente é um aspecto que demandará o estabelecimento de uma estratégia gradual de expansão do escopo das avaliações, considerados os impactos financeiros e operacionais que o aumento de anos avaliados poderia gerar. Ressaltamos que a possibilidade de permitir a integração das bases de dados das avaliações do 9º ano do Ensino Fundamental com as bases do Ensino Médio das redes estaduais, poderia reduzir a demanda por dados relativos ao 1º ano do Ensino Médio, o que poderá implicar não ser necessário realizar a avaliação em estudantes que estão iniciando o Ensino Médio.

O último aspecto abordado quanto à Arquitetura da Avaliação é a cobertura das avaliações.

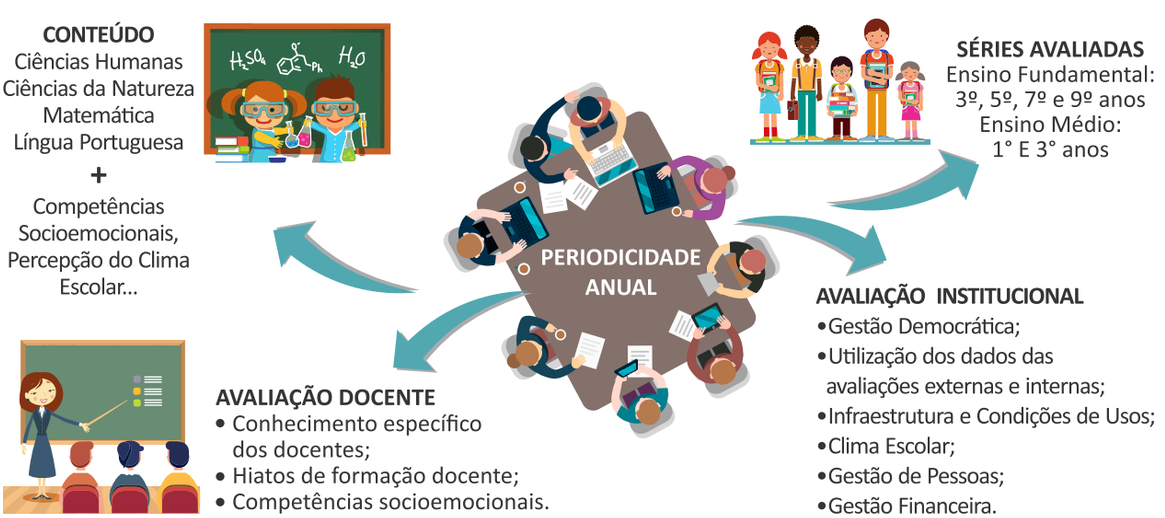
Figura 7: Qual deve ser a cobertura das avaliações?



Não há grandes surpresas quanto à opinião dos estados sobre a cobertura das avaliações. Os estados, em sua maioria, consideram que o foco da avaliação deve ser as escolas públicas. Há poucas experiências de avaliação das escolas particulares e quase sempre associadas ao pagamento de uma taxa pela instituição privada que deseje ser avaliada.

O diagrama abaixo resume o conjunto de deliberações dos estados, retratando qual o formato desejado para o Sistema Nacional de Avaliação da Educação Básica:

Figura 8: O sistema de avaliação – visão dos estados



2. COMO PODE SE ESTRUTURAR O SISTEMA DE AVALIAÇÃO DA EDUCAÇÃO BÁSICA

O amplo entendimento por parte dos estados de que a Avaliação em Educação precisa ser mais frequente e alcançar mais anos não redundará imediatamente em uma expansão da aplicação da avaliação. A limitação para o atingimento de todo o potencial desejado pelos estados está atrelada principalmente ao elevado impacto financeiro das avaliações em larga escala. Não é verdade, entretanto, que o custo de um modelo de avaliação mais abrangente torne inviável a expansão da avaliação. Elencaremos brevemente a seguir algumas alternativas que poderiam, ao aproveitar os recursos já disponíveis, proporcionar um crescimento relativamente rápido da Avaliação da Educação Básica no Brasil.

Alternativa 1: Articulação entre as avaliações estaduais e nacionais.

A ideia é bastante simples. Os estados poderiam promover as avaliações nos anos pares e o Governo Federal ficaria responsável por garantir as avaliações nos anos ímpares. Obviamente, as avaliações devem guardar coerência ao longo do tempo e prover a entrega de resultados e informações minimamente padronizadas em todos os anos. Dessa forma, seriam requisitos para a implementação dessa alternativa:

a) O alinhamento das avaliações entre governo federal e estados para atenderem às mesmas especificações de arquitetura e conteúdo;

b) Garantia do amplo acesso aos resultados das avaliações de ambas as redes e dos respectivos microdados;

c) As devolutivas das avaliações precisariam ser amplas e irrestritas para os estados e municípios, proporcionando todo o aparato informacional do qual dependem as estratégias pedagógicas e gerenciais das secretarias de educação;

d) Estabelecimento de parâmetros adequados para custeio das avaliações.

A ideia, ressalte-se, não é nova. No entanto, em nenhum outro momento os estados estiveram tão atentos e dispostos a expandir as avaliações. O momento é oportuno para empreender um esforço conjunto e articulado que viabilize o Sistema Nacional de Avaliação da Educação Básica – SINAEB, conforme previsto na Lei nº 13.005/2014, que aprova o Plano Nacional de Educação - PNE.

Alternativa 2: Formação de consórcio entre estados para organização conjunta da avaliação.

A criação de um consórcio entre os estados poderia oferecer grandes benefícios financeiros e gerenciais em relação à avaliação que cada Estado promove individualmente, a saber:

a) Unificação de serviços realizados paralelamente, como, por exemplo, formação de banco de itens;

b) Ganhos financeiros de escala produzidos pelo volume de cadernos de provas impressos e distribuídos;

c) Compartilhamento de conhecimento entre estados com diferentes estágios de evolução na Avaliação da Educação.

Evidentemente, nenhuma das alternativas apresentados possibilitaria o imediato atingimento de toda a extensão da avaliação desejada pelos estados. Seria necessário o incremento gradual de recursos para, no longo prazo, alcançar todo o escopo ideal de conteúdo e arquitetura das avaliações. Há, no entanto, de se reconhecer que as alternativas propostas poderiam resultar em importante avanço para o Sistema Nacional de Avaliação da Educação Básica sem gerar grandes impactos financeiros no curto prazo.

3. CONSIDERAÇÕES FINAIS SOBRE A PROPOSTA PARA O SISTEMA NACIONAL DE AVALIAÇÃO DA EDUCAÇÃO BÁSICA - SINAEB

O SINAEB, para ter legitimidade desde a sua criação, deve ser formulado em sintonia com os anseios das redes estaduais e municipais de educação. Dessa forma, a proposta consistiria em um Sistema Nacional de Avaliação da Educação Básica plenamente coerente com as políticas educacionais já estabelecidas pelos entes federativos.

As demandas dos estados refletem bastante a crescente utilização das informações oriundas das avaliações pelos municípios. Esse movimento é essencialmente benéfico para a melhoria da qualidade da informação. Estados e municípios, cientes da importância da avaliação e uso consciente das informações, fortalecem suas políticas educacionais e elevam a atenção às necessidades de estudantes e professores.

As avaliações podem e devem cumprir um importante papel de geração de dados para apoiar a definição de políticas públicas para a educação, bem como estratégias pedagógicas e de gestão das redes de ensino e das escolas, contribuindo para a avançarmos na qualidade da educação em nosso país.

CAPÍTULO 3

CONTEÚDOS POSSÍVEIS PARA UM SISTEMA DE AVALIAÇÃO DA EDUCAÇÃO BÁSICA

CONTEÚDOS POSSÍVEIS PARA UM SISTEMA DE AVALIAÇÃO DA EDUCAÇÃO BÁSICA

INTRODUÇÃO

Quem se posicionaria contra medir algo que se deseja transformar? Afinal, como disse Lorde Kelvin, "Aquilo que não se pode medir não se pode melhorar (saber se melhorou)"[1]. Assim, a questão não é se medir é *desejável*. A questão é se é *possível* medir com a necessária validade, confiabilidade e sensibilidade.

Avaliar significa medir com *validade*, *confiabilidade* e *sensibilidade*; onde validade significa medir o que se deseja medir; confiabilidade ou precisão significa que a medida não varia quando o que ela busca medir não se altera; e sensibilidade significa que a medida varia quando o que se deseja medir muda. Assim, a questão não parece ser sobre a desejabilidade de se avaliar. A questão é se avaliar (com a necessária validade, confiabilidade e sensibilidade) é ou não possível e, quando é possível, que uso será dado à avaliação e à mensuração que esta envolve.

TRÊS PROPÓSITOS PARA UMA AVALIAÇÃO

Em princípio, a mensuração de um resultado pode ser utilizada para três propósitos básicos. O primeiro seria o de saber se a situação atual é boa ou se houve melhoria. Se o que se deseja é apenas "avaliar" se a situação atual é boa, basta medir uma vez e comparar o resultado com um "padrão" ou "norma" (situação adequada, por exemplo), ou com o resultado de outras unidades (por exemplo, a situação de uma escola pode se mostrar melhor que a de ¾ de um grupo de referência dentro do universo das escolas do mesmo sistema educacional). Caso o objetivo seja "avaliar" se houve melhoria, exceto pelo fato de que são necessárias duas medidas (antes e depois), o processo é basicamente o mesmo e tudo que precisa ser feito é comparar a mudança ocorrida com a de um grupo de referência. Em ambos os casos é necessário ter em conta que o desempenho (situação atual ou progresso) de uma unidade é tão influenciado pelo contexto que pode se tornar pouco significativo comparar a situação de uma unidade com a de um grupo que não esteja sujeito ao menos a um contexto similar (por isso se restringe a

[1] O autor da frase é um matemático e físico britânico do Século XIX chamado William Thomson. Aos 68 anos de idade, recebeu o título de nobreza de Primeiro Barão Kelvin de Largs (ficou mais conhecido como "Lord Kelvin"), pela grande importância de seu trabalho científico.

comparação do desempenho de alunos àqueles no mesmo ano/série, ou entre escolas que atendem crianças provenientes de famílias com nível socioeconômico similar). Esse uso da mensuração para "avaliar" a situação inicial ou final, ou a magnitude da mudança, na linguagem da política pública, é comumente denominado *monitoramento* e pode utilizar para esse fim tanto o que se denomina "avaliação somativa", ao final do ano letivo, quanto uma "avaliação diagnóstica", realizada no início do ano, ou mesmo uma combinação das duas.

Essa, no entanto, não é a única razão para se medir um resultado. Uma segunda finalidade para a mensuração é o atrelamento de uma ação ou decisão, previamente acordada, ao resultado da avaliação. Esse é o caso, por exemplo, de um sistema de promoção atrelado à melhoria no aprendizado dos alunos, onde esses só são considerados aprovados quando atingem determinado patamar mínimo num dado conjunto de avaliações. Outros exemplos são os sistemas de incentivos morais ou materiais, atrelados a determinados resultados mínimos ou a melhorias mínimas nos resultados, como os prêmios dados aos professores ou para os alunos, vinculados aos resultados alcançados. Vale ressaltar que, independentemente da decisão atrelada, a mensuração dos resultados permanece indispensável à verificação do alcance de metas ou patamares mínimos estabelecidos. Tipicamente esse tipo de mensuração fica atrelado ao que se denomina avaliação somativa e ocorre ao final do ano letivo. A diferença entre os resultados da avaliação somativa presente e os alcançados na avaliação anterior serve como medida da eficácia conjunta de todas as ações, práticas e esforços perpetrados ao longo do período.

Por fim, a mensuração dos resultados pode ser utilizada para uma investigação ou diagnóstico das causas da presença ou da ausência de um nível adequado ou de uma mudança desejada. Nesse caso, a mensuração dos resultados é apenas o necessário e indispensável ponto de partida para a identificação e avaliação da importância dos fatores determinantes (causas) do nível ou da mudança nesses resultados. Por que o desempenho de um aluno, de uma escola ou de um sistema educacional, não está atingindo os níveis esperados? A mensuração dos resultados serve aqui como ponto de partida para a construção de um diagnóstico das causas do sucesso ou fracasso educacional.

Esse último talvez seja, dentre os três usos possíveis, o que mais completamente explora e utiliza a informação coletada. Conforme o próprio Lord Kelvin ressalta, "[...] apenas medindo [...] podemos testar se nossas conjecturas são falsas ou verdadeiras.[2]" e, dessa forma, decifrar os determinantes do sucesso ou fracasso.

[2] A citação original completa é a seguinte: "Although reason and imagination also advance knowledge, only measurement makes it possible to observe patterns and to experiment—to put one's guesses about what is and is not true to the test" William Thomson – Lord Kelvin (1824-1907).

No contexto da política pública, muitas vezes se reserva o termo “avaliação” a esse uso mais analítico da mensuração de resultados, em particular, quando a mensuração é o ponto de partida para a estimação da magnitude do impacto de uma dada ação ou prática. Para que esse uso avaliativo e diagnóstico seja possível, é indispensável que informações para um conjunto, o mais completo possível, de determinantes dos resultados educacionais também estejam disponíveis. Afinal, nesse caso, o objetivo é relacionar os resultados educacionais às suas possíveis causas, com vistas a quantificar a relativa importância de cada uma. Nesses termos, embora o Sistema de *Avaliação* da Educação Básica, SAEB, *avalie* no sentido de mensurar o aprendizado dos alunos, ele, por si só, *não avalia* ou identifica as causas do desempenho estar aquém ou além do esperado. O SAEB precisar estar conectado a outras informações para que seja possível a identificação e quantificação das causas do sucesso ou do fraco desempenho brasileiro.

MEDIR O QUE? QUAIS OS RESULTADOS DE INTERESSE?

Em princípio não deveria haver dúvidas: o objetivo de qualquer sistema de avaliação deve, em última instância, ser sempre a verificação da plena e efetiva garantia do direito à educação a cada uma das crianças, adolescentes e jovens do País. Por isso, é indispensável que se mensure os resultados educacionais relacionados ao acesso à escola, à permanência e ao progresso escolar e também aqueles ligados à proficiência nas mais diversas disciplinas e ao desenvolvimento das competências para a vida.

Mas, além disso, o direito à educação também trata do direito *na* educação. Assim, o acompanhamento da garantia do direito à educação requer também informações sobre o atendimento dado a cada estudante: foi respeitoso com relação a seus demais direitos humanos, estimulante e soube acomodar suas peculiaridades e necessidades? Em suma, todo sistema de vigilância do cumprimento do direito à educação deve tratar tanto dos indicadores tradicionais de resultado (acesso, permanência, progresso, proficiência e desenvolvimento de competências), como também de indicadores que avaliem a forma como os serviços foram oferecidos.

Via de regra, no entanto, todo sistema de avaliação e monitoramento educacional opta por mensurar não apenas os seus resultados finais, mas também os insumos utilizados, os processos adotados, as atividades realizadas. Como todos os serviços educacionais são produzidos necessariamente a partir de determinados insumos, utilizando determinados processos, caracterizados por uma dada sequência de atividades previamente especificadas, podemos dizer que uma escola ou um sistema educacional tem um bom funcionamento sempre que os insumos necessários estão disponíveis e as atividades previstas são realizadas de forma adequada. Um bom funcionamento não garante, contudo, que os resultados sejam alcançados;

mesmo que os insumos necessários estejam disponíveis e sejam empregados como planejado, pode haver inadequações no modelo de mudança adotado, levando a situações em que, a despeito do funcionamento perfeito, não se consiga alcançar os resultados finais esperados. É em boa medida por isso que todo sistema de avaliação e monitoramento educacional, via de regra, opta por mensurar não apenas os seus resultados finais, mas também os insumos empregados, os processos adotados, as atividades realizadas. Com isso, é possível saber em que medida as condições de funcionamento previstas foram alcançadas e, ao contrastarmos estas com os resultados alcançados, podemos entender melhor se o modelo de mudança adotado é adequado ou não.

1. LEGISLAÇÃO BRASILEIRA: O QUE DEVE SER AVALIADO

A legislação trata de forma separada de quatro tipos de avaliação. Em primeiro lugar, temos a avaliação do desempenho do aluno que professores e escolas devem fazer periodicamente para definir a promoção do aluno para o ano/série subsequente. Nesse caso, a Lei de Diretrizes e Base – LDB, em seus Artigos 24 (II e V) e 36 (§ 1º), afirma que a alocação dos alunos aos anos/séries mais adequadas deverá ocorrer com base em medidas de seu aproveitamento e desenvolvimento:

> *"Art. 24. A educação básica, nos níveis fundamental e médio, será organizada de acordo com as seguintes regras comuns:*
>
> > *II – a classificação em qualquer série ou etapa, exceto a primeira do ensino fundamental, pode ser feita:*
> >
> > > *a) por promoção, para alunos que cursaram, com aproveitamento, a série ou fase anterior, na própria escola;*
> > >
> > > *b) por transferência, para candidatos procedentes de outras escolas;*
> > >
> > > *c) independentemente de escolarização anterior, mediante avaliação feita pela escola, que defina o grau de desenvolvimento e experiência do candidato e permita sua inscrição na série ou etapa adequada, conforme regulamentação do respectivo sistema de ensino;*
> >
> > *V – a verificação do rendimento escolar observará os seguintes critérios:*
> >
> > > *a) avaliação contínua e cumulativa do desempenho do aluno, com prevalência dos aspectos qualitativos sobre os quantitativos e dos resultados ao longo do período sobre os de eventuais provas finais;*
> > >
> > > *b) possibilidade de aceleração de estudos para alunos com atraso escolar;*

c) possibilidade de avanço nos cursos e nas séries mediante verificação do aprendizado;

d) aproveitamento de estudos concluídos com êxito;

e) obrigatoriedade de estudos de recuperação, de preferência paralelos ao período letivo, para os casos de baixo rendimento escolar, a serem disciplinados pelas instituições de ensino em seus regimentos"

"Art. 36. O currículo do ensino médio observará o disposto na Seção I deste capítulo e as seguintes diretrizes:

§ 1º Os conteúdos, as metodologias e as formas de avaliação serão organizados de tal forma que ao final do ensino médio o educando demonstre:

I – domínio dos princípios científicos e tecnológicos que presidem a produção moderna;

II – conhecimento das formas contemporâneas de linguagem".

Embora exista a possibilidade de progressão automática, conforme revela o Artigo 32 (§ 2º) da LDB, ainda assim os estabelecimentos devem manter uma avaliação de desempenho dos alunos.

"Art. 32. O ensino fundamental obrigatório, com duração de 9 (nove) anos, gratuito na escola pública, iniciando-se aos 6 (seis) anos de idade, terá por objetivo a formação básica do cidadão, mediante:

§ 2º Os estabelecimentos que utilizam progressão regular por série podem adotar no ensino fundamental o regime de progressão continuada, sem prejuízo da avaliação do processo de ensino-aprendizagem, observadas as normas do respectivo sistema de ensino".

Ademais, conforme revelam os Artigos 13 e 36 da LDB, o professor deve dedicar parte de seu tempo à avaliação de seus alunos e fazer com que essa avaliação seja estimulante para estes.

"Art. 13. Os docentes incumbir-se-ão de:

V – ministrar os dias letivos e horas-aula estabelecidos, além de participar integralmente dos períodos dedicados ao planejamento, à avaliação e ao desenvolvimento profissional"

"Art. 36. O currículo do ensino médio observará o disposto na Seção I deste capítulo e as seguintes diretrizes:

II – adotará metodologias de ensino e de avaliação que estimulem a iniciativa dos estudantes"

Em segundo lugar, o Artigo 9º da LDB atribui à União a incumbência da *avaliação em escala* – via a colaboração com o Distrito Federal, estados e municípios – do rendimento escolar em todos os níveis educacionais. A importância da colaboração dos entes federados para o engajamento de todos os estabelecimentos de ensino na avaliação do rendimento dos alunos é ratificada no Artigo 87 da LDB.

> *"Art. 9º A União incumbir-se-á de:*
>
> > *VI – 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"*
>
> *"Art. 87. É instituída a Década da Educação, a iniciar-se um ano a partir da publicação desta lei.*
>
> > *§ 3º O Distrito Federal, cada estado e município, e, supletivamente, a União, devem:*
> >
> > *IV – integrar todos os estabelecimentos de ensino fundamental do seu território ao sistema nacional de avaliação do rendimento escolar"*

A importância central da avaliação da qualidade da educação básica, tanto em nível nacional como em cada escola, é ressaltada pelo Plano Nacional de Educação – PNE. Em seu Artigo 11, o PNE institui o Sistema Nacional de Avaliação da Educação Básica – SINAEB, que tem por atribuição produzir "informação para a avaliação da qualidade da educação básica e para a orientação das políticas públicas". Esse sistema deve conter informações tanto sobre o desempenho dos estudantes como sobre "o perfil do alunado e do corpo dos (as) profissionais da educação, as relações entre dimensão do corpo docente, do corpo técnico e do corpo discente, a infraestrutura das escolas, os recursos pedagógicos disponíveis e os processos da gestão".

> *"Art. 11. O Sistema Nacional de Avaliação da Educação Básica, coordenado pela União, em colaboração com os Estados, o Distrito Federal e os Municípios, constituirá fonte de informação para a avaliação da qualidade da educação básica e para a orientação das políticas públicas desse nível de ensino.*
>
> > *§ 1º O sistema de avaliação a que se refere o caput produzirá, no máximo a cada 2 (dois) anos:*
> >
> > > *I – indicadores de rendimento escolar, referentes ao desempenho dos (as) estudantes, apurado em exames nacionais de avaliação, com participação de pelo menos 80% (oitenta por cento) dos (as) alunos (as) de cada ano*

escolar periodicamente avaliado em cada escola, e aos dados pertinentes apurados pelo censo escolar da educação básica;

II – indicadores de avaliação institucional, relativos a características como o perfil do alunado e do corpo dos (as) profissionais da educação, as relações entre dimensão do corpo docente, do corpo técnico e do corpo discente, a infraestrutura das escolas, os recursos pedagógicos disponíveis e os processos da gestão, entre outras relevantes.

§ 2º A elaboração e a divulgação de índices para avaliação da qualidade, como o Índice de Desenvolvimento da Educação Básica - IDEB, que agreguem os indicadores mencionados no inciso I do § 1º não elidem a obrigatoriedade de divulgação, em separado, de cada um deles.

§ 3º Os indicadores mencionados no § 1º serão estimados por etapa, estabelecimento de ensino, rede escolar, unidade da Federação e em nível agregado nacional, sendo amplamente divulgados, ressalvada a publicação de resultados individuais e indicadores por turma, que fica admitida exclusivamente para a comunidade do respectivo estabelecimento e para o órgão gestor da respectiva rede.

§ 4º Cabem ao Inep a elaboração e o cálculo do Ideb e dos indicadores referidos no § 1º.

§ 5º A avaliação de desempenho dos (as) estudantes em exames, referida no inciso I do § 1º, poderá ser diretamente realizada pela União ou, mediante acordo de cooperação, pelos Estados e pelo Distrito Federal, nos respectivos sistemas de ensino e de seus Municípios, caso mantenham sistemas próprios de avaliação do rendimento escolar, assegurada a compatibilidade metodológica entre esses sistemas e o nacional, especialmente no que se refere às escalas de proficiência e ao calendário de aplicação.”

Em terceiro lugar, a LDB também estipula em seu Artigo 67 que a mensuração do desempenho dos professores seja utilizada como um dos elementos definidores de sua progressão funcional:

''Art. 67. Os sistemas de ensino promoverão a valorização dos profissionais da educação, assegurando-lhes, inclusive nos termos dos estatutos e dos planos de carreira do magistério público:

IV – progressão funcional baseada na titulação ou habilitação, e na avaliação do desempenho"

Por fim, em quarto lugar, o Artigo 7º da LDB também coloca como condição para a prestação de serviços de educação por organizações da iniciativa privada o uso da avaliação institucional da qualidade do ensino que oferecem.

"Art. 7º O ensino é livre à iniciativa privada, atendidas as seguintes condições:

II – autorização de funcionamento e avaliação de qualidade pelo Poder Público"

Em resumo, vimos que as possibilidades para a avaliação de um sistema educacional são extremamente amplas. A avaliação pode se voltar tanto para o acompanhamento e promoção de cada aluno individualmente, quanto para a avaliação do desempenho do conjunto dos alunos de uma escola ou rede de ensino. O desenho desse tipo de sistema de avaliação é o foco deste relatório, uma vez que trata daquilo que centra a atenção tanto dos atuais sistemas de avaliação estaduais como do futuro Sistema Nacional de Avaliação da Educação Básica – SINAEB, como proposto pelo Plano Nacional de Educação – PNE.

A avaliação, no entanto, como vimos, pode também ter outros objetivos. Pode ser dirigida para avaliar o desempenho dos professores, a qualidade dos serviços oferecidos pelas escolas e sistemas educacionais ou a eficiência com que escolas, regionais e secretarias funcionam. Assim, a despeito de sua reconhecida centralidade e importância, deixamos fora do escopo do relatório a definição de um sistema de avaliação institucional em todos os níveis.

No capítulo sobre usos da avaliação, argumentamos que é preferível coletarmos as informações na menor unidade de análise possível, dado que agregar é sempre mais fácil, e desagregar, nem sempre possível. O nível de agregação proposto neste relatório é a escola (ou talvez turma). Conforme procuramos mostrar, mesmo quando restritas a esse nível mais agregado, as possibilidades de avaliação permanecem extremamente amplas. Em todos os casos, porém, o sistema de avaliação deve sempre contar com alguma medida do desenvolvimento dos alunos, dado que este é o resultado mais diretamente relacionado à garantia do direito à educação, em geral, e do direito ao aprendizado, em particular.

Além disso, a utilidade do sistema de avaliação depende sobremaneira da presença de outros componentes avaliativos. Idealmente, o sistema deve contar com informações do contexto escolar (perfil do corpo docente, clima escolar, infraestrutura, recursos pedagógicos disponíveis, processo de gestão, entre outros) e dos contextos familiar e comunitário onde a escola e os estudantes se inserem. Essas informações são fundamentais tanto para a avaliação

dos determinantes dos sucessos e fracassos educacionais, como também para a realização de comparações mais justas entre o desempenho de escolas, na medida em que limitamos a comparação às escolas cujos contextos sejam similares.

Por conseguinte, neste relatório consideramos sistemas de avaliação em que a escola (eventualmente a turma) é a unidade mínima de análise, que incluem em seu conteúdo medidas do rendimento dos estudantes na escola, além do contexto escolar, familiar e comunitário em que se inserem. Assim, embora a unidade última de análise no sistema tratado neste relatório seja tipicamente a escola, o resultado central a ser avaliado é sempre o desempenho dos estudantes e seu contexto.

2. DIMENSÕES DO DIREITO À EDUCAÇÃO

A Constituição Brasileira, assim como a Declaração Universal dos Direitos Humanos e uma variedade de outros compromissos e aspectos da legislação nacional e internacional, apresenta a educação como um componente essencial dos direitos humanos mais básicos. Embora todos os direitos humanos sejam igualmente importantes e inalienáveis, o direito à educação desempenha reconhecidamente um papel central, já que permite que cada pessoa conheça, acesse e goze de seus (demais) direitos. Em função disso, a garantia do direito à educação é muitas vezes vista como uma condição necessária à garantia de outros direitos. A educação funciona como porta de entrada para o acesso aos direitos humanos.

Tanto a Constituição Brasileira como a Declaração Universal dos Direitos Humanos veem o direito à educação como o direito ao pleno desenvolvimento humano e a escola como sendo o instrumento que oferece, por excelência, as oportunidades necessárias para que este pleno desenvolvimento possa ocorrer.

> *"Art. 205: A educação, direito de todos e dever do Estado e da família, será promovida e incentivada com a colaboração da sociedade, visando ao pleno desenvolvimento da pessoa, seu preparo para o exercício da cidadania e sua qualificação para o trabalho" (Constituição da República Federativa do Brasil 1988).*
>
> *"Art. 26, Inciso 2: A educação será orientada no sentido do pleno desenvolvimento da personalidade humana e do fortalecimento do respeito pelos direitos do ser humano e pelas liberdades fundamentais. A educação promoverá a compreensão, a tolerância e a amizade*

entre todas as nações e grupos raciais ou religiosos e coadjuvará as atividades das Nações Unidas em prol da manutenção da paz" (Declaração Universal dos Direitos Humanos).

O direito à educação deve ser interpretado, portanto, como o direito ao pleno desenvolvimento, em contraste com o mero direito de acesso a uma vaga em uma escola. Essa diferenciação é muitas vezes sintetizada afirmando-se que o direito à educação não pode ficar restrito ao acesso à escola, mas deve garantir o desenvolvimento e o aprendizado. Daí a crescente distinção entre *direito de acesso à escola* e *direito ao aprendizado*.

Tradicionalmente, a distinção entre *direito de acesso à escola* e *direito ao aprendizado* tem sido reduzida a uma demanda por educação de qualidade, na medida em que um tratamento mais simplista do *direito ao aprendizado* não faz distinção entre acesso a uma escola de qualidade e aprendizado. Essa concepção não diferencia "ter uma oportunidade" e "alcançar um resultado". Assim, pressupõe que o aprendizado de um estudante requer o acesso às devidas oportunidades e que o acesso às devidas oportunidades leva necessariamente ao aprendizado. Dessa forma, qualquer aprendizado inadequado resulta sempre de deficiências na qualidade da escola, mas nunca de deficiências no engajamento do estudante frente a uma escola de alta qualidade e adequada às suas necessidades. Por essa linha de argumentação, em uma escola de boa qualidade, não existiriam estudantes desengajados ou desinteressados e o acesso a uma oportunidade adequada levaria necessariamente aos resultados esperados.

Quando oportunidades e resultados são tão fortemente conectados, a avaliação do grau de engajamento dos estudantes não exerce qualquer papel relevante. Entretanto, para que oportunidades levem a resultados desejados, é necessário que os estudantes aproveitem as oportunidades disponíveis. Assim, vale ressaltar que a conexão direta entre oportunidades e resultados requer não apenas que a simples disponibilidade de uma oportunidade de qualidade leve inexoravelmente à sua utilização, mas também que a plena utilização de uma oportunidade de qualidade leve necessariamente ao alcance de um resultado de interesse. Na sequência, vamos tratar da importância ou irrelevância de avaliarmos o grau de engajamento dos estudantes, sempre supondo que oportunidades de boa qualidade, quando devidamente utilizadas, levam ao alcance de objetivos de desenvolvimento e aprendizado.

Quando tratamos da educação da população jovem, talvez não seja possível ignorar que as escolhas feitas pelos estudantes têm motivações que vão além das que podem ser influenciadas pela escola, por mais atraente e adequada que esta seja. Nesse caso, acreditar que oportunidades levam inexoravelmente a resultados pode se tornar inadequado. O aprendizado (resultado desejado) não decorre de forma mecânica e diretamente apenas da disponibilidade e do acesso a oportunidades educacionais com a necessária qualidade e adequação; o aprendizado também depende do interesse e, portanto, do engajamento dos

estudantes no aproveitamento dessas oportunidades que, por sua vez, não é determinado exclusivamente pela qualidade das oportunidades educacionais oferecidas.

Nesse sentido, o engajamento também deve ser um elemento na avaliação da garantia do direito à educação. Consideremos o caso em que só ocorre engajamento quando existe disponibilidade de serviços adequados e de qualidade, e que o engajamento em serviços de qualidade é suficiente para alcançarmos os objetivos de desenvolvimento e aprendizado. Nesse caso, sabemos que os objetivos serão alcançados (e o direito à educação será assegurado) apenas quando há engajamento e disponibilidade de serviços adequados e de qualidade.

Embora a falta de engajamento certamente implique que os objetivos de desenvolvimento e aprendizado não serão alcançados, ela por si só não significa indisponibilidade de serviços adequados e de qualidade. Assim, quando o direito à educação é entendido como o acesso a serviços adequados e de qualidade, a falta de engajamento não indica falha na garantia do direito à educação. No entanto, quando o direito à educação requer que os objetivos de desenvolvimento e aprendizado sejam atingidos, a falta de engajamento é um indicativo de que o direito à educação não está sendo garantido. Portanto, tudo depende da forma como o direito à educação é medido: ou pelos resultados alcançados (aprendizado e conclusão, por exemplo) ou pelas oportunidades oferecidas. Afinal, permanece em debate se um jovem que teve oportunidades educacionais adequadas e de qualidade, mas não se engajou, teve seu direito à educação assegurado ou não.

Uma questão é líquida e certa: nenhum resultado educacional de destaque e, portanto, a garantia do direito à educação, será alcançado sem o engajamento dos jovens. A disponibilidade de oportunidades educacionais adequadas e de qualidade, em geral, não é suficiente para garantir que tais resultados sejam alcançados. O engajamento dos jovens é indispensável e este, geralmente, não é determinado exclusivamente pela disponibilidade de oportunidades, por melhor e mais adequadas que sejam. Fatores externos, fora do controle da escola, também influenciam o engajamento dos estudantes.

Em suma, concluímos que o engajamento dos estudantes tem em geral certa independência na determinação do sucesso educacional. Por um lado, não parece ser sempre plenamente determinado pela qualidade e adequação das oportunidades educacionais disponíveis (embora seja certamente influenciado por estes aspectos). Por outro, é um determinante crucial dos resultados educacionais. Assim, o engajamento dos estudantes deve ser também utilizado para indicar se o direito à educação está sendo garantido, seja porque é um indicador indireto da existência de uma oferta adequada e de qualidade dos serviços, seja porque aponta para o cumprimento dos objetivos de desenvolvimento e aprendizado. Nesse

sentido, ao elencarmos os aspectos necessários para verificação do direito à educação, começaremos pelo engajamento.

2.1. ENGAJAMENTO

Como a relação entre o jovem e a escola é bem diversa, existe uma ampla variedade de aspectos que precisam ser considerados para se avaliar o grau de engajamento dos jovens. No limite, temos pleno engajamento apenas quando o jovem se matricula na escola no início de cada ano letivo, frequenta todos os dias letivos e realiza todas as atividades escolares a ele atribuídas ou que dele se espera que busque realizar por iniciativa própria.

Assim como a participação em qualquer atividade, o engajamento dos jovens em atividades escolares tem um *componente extensivo* (ter ou não algum engajamento) e um *componente intensivo* (o nível de engajamento). Esses componentes, por sua vez, admitem decomposições adicionais de grande valor analítico. A seguir, iremos nos debruçar sobre eles.

MARGEM EXTENSIVA

Quando analisamos a margem extensiva, é fundamental diferenciar entre os fenômenos do abandono e da evasão. A evasão ocorre quando o aluno que foi à escola em um dado ano deixa de se matricular no início do ano letivo seguinte; já o abandono ocorre quando um aluno que se matriculou no início do ano deixa de frequentar a escola a partir de um dado momento durante o ano letivo. Portanto, um aluno pode evadir sem jamais ter abandonado a escola. Esse aluno pode ter sido aprovado no último ano letivo e decidir parar de estudar e, consequentemente, evadir – estes são os evadidos aprovados.

Da mesma forma, um aluno pode abandonar múltiplas vezes a escola sem que venha a evadir: um exemplo seria um aluno que sempre se matricula na escola no início do ano letivo, mas não permanece até o final, ou seja, abandona a escola em sucessivos anos letivos. Assim, mesmo quando apenas a margem extensiva é analisada, o escopo da análise depende de como a interpretamos: evasão (ausência de matrícula no início do ano letivo) *versus* abandono (desistência durante o ano letivo). Em boa medida, pode-se considerar a evasão como um caso mais grave de abandono, já que o jovem abandona a escola antes mesmo de o ano letivo iniciar.

MARGEM INTENSIVA

O componente intensivo é bem mais complexo, uma vez que envolve tanto a ida à escola (a frequência), como o efetivo engajamento nas atividades escolares. Consideremos em primeiro lugar a questão mais simples, referente à frequência. Se tudo que importa é se o jovem vai ou não à escola, ainda assim existem nuances quanto ao que se deve considerar maior ou menor engajamento. Um aluno que frequentou intermitentemente as aulas, mas nunca abandonou de fato os estudos, embora tenha sido reprovado por faltar inúmeras vezes, deve ser tratado de forma distinta do aluno que frequentou o primeiro semestre e logo após abandonou a escola, para retornar no ano seguinte? Em última instância, é preciso definir em que medida o abandono deve ser tratado de forma diferenciada da reprovação por faltas. Falta de engajamento deve ser definida em função da porcentagem das aulas que o aluno faltou ou em função da interrupção de sua conexão com a escola? Devemos focar a atenção na porcentagem de dias de aula que o aluno faltou ou na porcentagem de alunos que desistem de estudar antes do final do ano letivo?

Mesmo entre aqueles que frequentam a escola em um dado dia, o grau de engajamento também enfrenta grandes variações: alguns podem ter feito seus deveres de casa e estudado o material referente às aulas anteriores, enquanto outros, não; uns podem prestar grande atenção às aulas, enquanto outros permanecem dispersos ou concentrados em outras atividades.

Ao considerarmos medidas do grau de conexão dos jovens com a escola, é também importante diferenciar entre medidas do grau de desengajamento (de estoque) e medidas da mudança no grau de desengajamento (de fluxo). Por exemplo, ao analisarmos a evasão, o que importa é quantos jovens, dentre os que se matricularam na escola em um dado ano, deixaram de se matricular no ano seguinte (o fluxo, o grupo de novos evadidos), ou o total de jovens que em um dado ano não se matriculou na escola no ano seguinte (o estoque, o número total de evadidos)?

2.2. APRENDIZADO

"Na Base Nacional Comum Curricular - BNCC[3], o Ensino Fundamental (assim como o Ensino Médio) está organizado em quatro *áreas do conhecimento*": Linguagens, Matemática, Ciências

[3] A Base Nacional Comum Curricular (BNCC) é um documento de caráter normativo que define o conjunto orgânico e progressivo de aprendizagens essenciais que todos os alunos devem desenvolver ao longo das etapas e modalidades da Educação Básica. No momento de elaboração deste relatório, o documento referente à BNCC estava pendente de homologação pelo Conselho Nacional de Educação.

da Natureza e Ciências Humanas. Exceto pelo caso da Matemática, todas as demais áreas do conhecimento abarcam diversos *componentes curriculares*. Em Linguagens estão incluídos os seguintes componentes curriculares: Língua Portuguesa, Literatura, Língua Estrangeira Moderna, Arte e Educação Física. Ciências da Natureza, por sua vez, abarca Biologia, Física e Química. Por fim, pertencem às Ciências Humanas os seguintes componentes curriculares: Filosofia, Geografia, História e Sociologia.

Segundo a forma organizacional proposta pela BNCC, a cada área de conhecimento e componente curricular estão associadas *competências específicas*. "As competências específicas possibilitam a articulação horizontal entre as áreas, perpassando todos os componentes curriculares, e também a articulação vertical, ou seja, a progressão" (BNCC, 2016, p. 26) entre as etapas do Ensino Fundamental e Médio. "Para garantir o desenvolvimento das competências específicas, cada componente curricular apresenta um conjunto de *habilidades*. Essas habilidades estão relacionadas a diferentes *objetos de conhecimento* – aqui entendidos como conteúdos, conceitos e processos –, que, por sua vez, são organizados em *unidades temáticas*" (BNCC, 2016, p. 26). Assim, por exemplo, a Base Curricular referente ao 5º Ano em Matemática contempla 25 habilidades, 20 objetos do conhecimento e 5 unidades temáticas.

O aprendizado desse amplo conjunto de habilidades, objetos de conhecimento, temas e competências específicas e gerais é um direito. É o que denominamos do direito ao aprendizado. Verificar se esse direito está sendo assegurado requer avaliar em que medida os estudantes de fato adquiriram tais habilidades e competências. Para isso, é necessário definir operacionalmente o que significa ser proficiente com relação a cada uma delas. Essa definição se faz com base em Matriz de Referência que, diferentemente do que se espera de um currículo, não trata de todas as habilidades e competências importantes. Essa Matriz basicamente centra sua atenção nas habilidades que podem ser medidas por meio de uma prova escrita. Dessa forma, é evidente que o desempenho dos alunos em uma prova como essa não fornece ao professor indicações de todas as habilidades e competências desenvolvidas. Tipicamente, o foco da avaliação é medir competências básicas e essenciais. Nesse sentido, qualquer estratégia que desenvolva nos alunos a capacidade de responder corretamente aos itens da avaliação estará lhes permitindo consolidar competências fundamentais para o exercício do seu direito ao aprendizado.

Embora o currículo, em geral, seja mais amplo que a Matriz de Referência, muitas vezes seu conteúdo programático se assemelha. Por exemplo, enquanto a proposta de currículo de Matemática do 5º Ano, segundo a BNCC, cobre 5 unidades temáticas, a Matriz de Referência da avaliação nacional (Prova Brasil) feita nessa transição entre as duas fases do Ensino Fundamental cobre 4. No entanto, se é certo que a intenção da Prova Brasil é avaliar o domínio dos conhecimentos e habilidades do todo o ciclo educacional objeto da avaliação, seria

perfeitamente possível realizar a avaliação para averiguar exclusivamente se o aluno domina o conteúdo e as habilidades esperadas da última série concluída. No exemplo da avaliação ao final do 5º Ano, a Prova Brasil averigua se o aluno domina os conhecimentos e habilidades de todo o primeiro ciclo do ensino fundamental, mas poderia haver uma prova que avaliasse somente o domínio esperado para o 5º Ano. Quer num caso, quer noutro, diante da aprovação de uma Base Nacional Comum Curricular, a Prova Brasil precisaria ser ajustada para avaliar o domínio dos objetos do conhecimento e habilidades especificados na nova referência curricular.

No que se refere ao conteúdo das avaliações da garantia do direito ao aprendizado, ao menos duas questões merecem particular atenção. A primeira delas é a fantástica amplitude do leque de habilidades que um estudante precisaria dominar para podermos assegurar que seu direito ao aprendizado foi atendido. Por exemplo, no 9º Ano do Ensino Fundamental, a BNCC considera que um estudante deve adquirir mais de 200 habilidades, no conjunto dos 8 componentes curriculares (disciplinas) a que são expostos. E, se a intenção da avaliação ao fim do 9º Ano é averiguar o domínio de todo conteúdo programático dos anos finais do ensino fundamental, o número de habilidades que o aluno precisaria demonstrar seria de aproximadamente 800 habilidades.

Uma possível solução para diminuir o número de habilidades a serem avaliadas é, na medida em que as unidades temáticas se repetem ao longo dos anos escolares do mesmo ciclo educacional, considerar que as habilidades especificadas para o último ano do ciclo educacional em questão referem-se à ultima etapa de um contínuo de desenvolvimento e, portanto, se o aluno domina o conteúdo do último ano do ciclo, necessariamente domina o dos anos anteriores. Mas a despeito dessa solução, um número da ordem de 200 habilidades em um único ano escolar já inviabiliza a avaliação de todo conteúdo curricular; nenhuma avaliação em escala seria capaz de avaliar o desempenho dos estudantes em todas essas 200 habilidades. Existe, portanto, a necessidade de limitar a avaliação a apenas uma parte do currículo. É com base na Matriz de Referência da avaliação que essa escolha se torna explícita, permitindo que os resultados da avaliação possam ser propriamente interpretados.

A segunda questão a se considerar é que o direito ao aprendizado é individual. Assim, o fato de um estudante numa escola ter um amplo espectro de habilidades não compensa o limitado leque de habilidades de outro estudante. O estudante com habilidades limitadas tem seu direito de aprendizado, independentemente da riqueza de habilidades do restante de seus colegas na mesma escola. Portanto, a proficiência média por escola, quando usada como medida da garantia do direito ao aprendizado, deve ser interpretada com extrema cautela. Nesse particular, é sempre importante contarmos com informação por escola sobre a porcentagem de alunos que atendem um nível mínimo de proficiência considerado adequado.

2.3. COMPETÊNCIAS

Conforme ressalta a Base Nacional Comum Curricular[4], o direito ao aprendizado não pode se limitar à apropriação de "... conhecimento desinteressado e erudito entendido como fim em si mesmo" (BNCC, 2016, p. 17). A garantia do direito à educação e ao aprendizado pressupõe que "os resultados das aprendizagens precisam se expressar e se apresentar como sendo a possibilidade de utilizar o conhecimento em situações que requerem aplicá-lo para tomar decisões pertinentes" (BNCC, 2016, p. 17).

"A esse conhecimento mobilizado, operado e aplicado em situação se dá o nome de *competência*" (BNCC, 2016, p. 15). De fato, "No âmbito da BNCC, a noção de competência é utilizada no sentido da mobilização e aplicação dos conhecimentos escolares, entendidos de forma ampla (conceitos, procedimentos, valores e atitudes). Assim, ser competente significa ser capaz de, ao se defrontar com um problema, ativar e utilizar o conhecimento construído" (BNCC, 2016, p. 16).

A garantia do direito ao aprendizado requer, portanto, "a indicação clara do que os alunos devem saber, e, sobretudo, do que devem saber fazer como resultado de sua aprendizagem" (suas competências) (BNCC, 2016, p. 16). Em outras palavras, "a BNCC está comprometida [...] (com a) construção intencional de processos educativos que promovam aprendizagens... de modo a formar pessoas autônomas, capazes de se servir dessas aprendizagens em suas vidas" (competências) (BNCC, 2016, p. 17).

Assim, ao foco que não reside apenas na apropriação de conhecimentos, mas também na capacidade de mobilizar esses conhecimentos com a finalidade de buscar soluções, damos o nome de competência. Na BNCC os conteúdos curriculares estão a serviço do desenvolvimento de competências. Essa visão pedagógica, baseada no conceito de competência, já era ressaltada nas Leis de Diretrizes e Bases da Educação Nacional – LDB "... no já destacado Artigo 9º, Inciso IV, que menciona as diretrizes e competências que nortearão os currículos. A LDB adota de forma explícita o enfoque por competências" (BNCC, 2016, p. 15). A BNCC, no entanto, vai além, ao colocar o desenvolvimento de competências como o principal fundamento pedagógico.

[4] A Base Nacional Comum Curricular é um documento de caráter normativo que estabelece os objetivos de aprendizado (conhecimentos essenciais e competências), em nível nacional, que todos os alunos devem desenvolver na Educação Básica. BRASIL. Ministério da Educação. Base Nacional Comum Curricular. Proposta preliminar. Segunda versão revista. Brasília: MEC, 2016. Disponível em: <http://basenacionalcomum.mec.gov.br/documentos/bncc-2versao.revista.pdf>. Acesso em: 17 Jul. 2017.

3. A CENTRALIDADE DO CONTEXTO ESCOLAR E FAMILIAR

A garantia do direito à educação ou do direito ao aprendizado, em particular, não é de responsabilidade de um único agente. Por melhor que seja o professor ou a escola, eles sozinhos não são capazes de garantir que todos os alunos adquiram todo o tempo as habilidades que constam no currículo. Um bom professor e uma boa escola podem e devem compensar, em boa medida, as diferenças de ambiente familiar entre alunos garantindo que a grande maioria esteja, durante a maior parte do tempo, motivada para aprender e, por conseguinte, engajada. Garantir que todos estejam plenamente engajados o tempo todo e que todas as diferenças de contexto familiar sejam neutralizadas pela escola parecem metas inviáveis. Ademais, o engajamento e o desempenho dos alunos também dependem da vontade deles, bem como da qualidade e da quantidade de atenção e cuidado que recebem em casa e na comunidade onde vivem.

A educação é uma atividade em que, inexoravelmente, as responsabilidades são compartilhadas entre escola, alunos, famílias e comunidade. Consequentemente, os resultados educacionais dependem não apenas do contexto escolar e familiar, mas também das atitudes, valores, preferências e percepções dos alunos. Um aluno que não acredita que estudar vale a pena irá provavelmente estudar menos do que outro que está convencido da importância da educação para a sua vida e deverá, portanto, apresentar um pior desempenho acadêmico. Assim, para que os resultados de avaliações do desempenho educacional (seja do aprendizado ou do engajamento) possam ser adequadamente interpretados e comparados, e suas causas identificadas e mensuradas, é indispensável que, concomitantemente à avaliação do desempenho, se possa contar com informações de contexto, seja ele escolar, familiar ou de natureza pessoal.

3.1. CONTEXTO ESCOLAR: A “QUALIDADE” DA ESCOLA

Conforme ressaltado, o desempenho educacional de um aluno depende de uma diversidade de atributos relacionados a uma variedade de atores e instituições. Como a maior parte do aprendizado ocorre dento da própria escola ou orientado por ela, é natural que o contexto escolar seja o principal determinante do desempenho educacional (aprendizado, progresso e engajamento), em particular, do aprendizado.

De fato, a relação entre a garantia do direito à educação e o acesso a uma escola de “qualidade” tem demonstrado ser empiricamente tão próxima que, por vezes, os dois conceitos

se mesclam, com o direito à educação sendo na prática tratado como o direito de acesso a uma escola de “qualidade”. Embora não haja disputa de que o direito à educação seja mais que a “qualidade” da escola, essa corriqueira justaposição de conceitos deixa clara a importância do acesso a uma educação de “qualidade” para a garantia do direito à educação. Sem margem para qualquer dúvida, não é possível garantir o direito à educação sem a garantia de um efetivo acesso a uma escola de “qualidade”. A “qualidade” da escola pode não ser suficiente, mas é certamente indispensável.

A importância e indispensabilidade da “qualidade” da escola são seguramente consensuais; já o que se entende exatamente por “qualidade” da escola é bem mais controverso, ou ao menos medido de formas variadas. Ao menos três abordagens podem ser escolhidas quando se busca medir a “qualidade” da escola. Vale ressaltar que essas abordagens são complementares e, se por vezes, são utilizadas de forma isolada, é muito mais por simplicidade e limitações das fontes de informação do que por um julgamento de uma abordagem ser mais importante ou abarcar a outra.

Em primeiro lugar, temos a visão da “qualidade” que aponta para a disponibilidade e qualidade dos insumos. Nesse caso, uma escola “boa” é uma escola rica em insumos, isto é, com professores de todas as disciplinas (componentes curriculares) em quantidade suficiente e com a formação adequada. Assim, o tamanho das turmas poderia ser definido da forma pedagogicamente mais adequada e não na busca de otimizar o uso de recursos humanos escassos (caso em que turmas grandes precisam ser formadas).

Também é fundamental para a “qualidade” dos serviços oferecidos pela escola, a disponibilidade e qualidade dos recursos humanos que apoiam o trabalho dos professores em sala de aula, aí incluída a disponibilidade e qualidade dos coordenadores pedagógicos, da equipe responsável pela gestão da escola (diretores e equipe administrativa), além de toda equipe de apoio ao seu funcionamento, como o pessoal responsável pela alimentação, transporte e segurança dos estudantes, e também aqueles responsáveis pela manutenção, limpeza e organização dos espaços da escola. A abundância de insumos também deve se traduzir numa infraestrutura adequada com ampla disponibilidade de laboratórios, bibliotecas, áreas para esporte, refeitório e outras facilidades necessárias ao funcionamento da escola.

A “qualidade” vista pelos insumos também abarca salas de aulas e demais espaços da escola de fácil acesso a todos (em particular, pessoas com deficiências), amplos, limpos, arejados e bem iluminados. A “qualidade” desses espaços também requer que estejam adequadamente equipados para que funcionem de forma eficaz: as bibliotecas precisam de um acervo variado e adequado às necessidades dos alunos, da mesma forma como os laboratórios devem ter os equipamentos e materiais que precisam para operarem de forma efetiva. Também

inclusa entre os insumos deve estar a oferta de condições básicas para os alunos frequentarem a escola e dela se beneficiarem como transporte, merenda, uniforme, livros e materiais escolares.

Mas, evidentemente, uma escola com uma ampla disponibilidade de insumos de qualidade não se constitui necessariamente numa escola de "qualidade". A "qualidade" da escola não está imobilizada em seus insumos, sendo sua disponibilidade e qualidade essenciais, mas não suficientes. É também necessário que esses insumos sejam mobilizados e utilizados, e que essa utilização seja feita de forma adequada e eficiente.

Pode-se dizer que uma escola, com recursos em quantidade e qualidade adequados, tem um "bom" funcionamento quando mobiliza esses recursos de forma intencional com o explícito propósito de realizar de forma adequada e efetiva uma série de atividades desenhadas para promover o aprendizado e o desenvolvimento dos seus estudantes. Em particular, para que uma escola tenha um bom funcionamento não basta que os professores tenham uma boa formação e dominem um amplo leque de práticas pedagógicas; é necessário que escolham as práticas mais adequadas a cada contexto e a cada estudante, implementando-as de forma eficaz. Assim, uma escola com "bom" funcionamento é uma escola que executa em sua plenitude, no momento apropriado e da forma adequada e efetiva, o amplo conjunto de atividades (em particular, as práticas pedagógicas em sala de aula) requeridas para que seus estudantes tenham plenas condições para se desenvolverem e se apropriarem dos conhecimentos e habilidades previstas no currículo escolar. Portanto, com relação à mensuração da "qualidade" da escola, uma visão complementar a da disponibilidade e qualidade dos insumos escolares é a do "bom" funcionamento. Nesse caso, uma escola de "qualidade" seria uma escola que "funciona", onde funcionar significa que as atividades que deveriam acontecer de fato acontecem com a abrangência, frequência, qualidade, adequação e efetividade previamente especificadas e indispensáveis ao pleno desenvolvimento e aprendizado dos estudantes.

Por fim, não basta que a escola tenha os recursos em quantidade e qualidade adequadas, e que os mobilize de forma planejada, intencional e adequada. É também essencial que se construa um ambiente escolar acolhedor e estimulante, isto é, um ambiente escolar que seja seguro, justo, fraterno, participativo e cooperativo, e onde exista crença no potencial de cada membro da comunidade escolar (estudantes e professores, entre outros atores) e na possibilidade de desenvolver o potencial de cada um. Por vezes, esse ambiente é denominado de clima escolar. Um bom clima escolar é certamente indispensável para um bom desempenho educacional dos estudantes e particularmente importante para o engajamento de toda a comunidade escolar com as atividades programadas.

Conforme já ressaltado, numa escola e, no processo de desenvolvimento e de aprendizado do aluno de forma mais geral, não existe um agente com poder discricionário sobre todos os insumos. O resultado educacional é decorrente da ação de agentes (professores, alunos e famílias) com substancial independência. Assim, o sucesso da educação e da escola, em particular, depende sobremaneira da capacidade de coordenação de esforço entre esses agentes que formam a comunidade escolar: o professor precisa ensinar de forma efetiva; o aluno precisa interagir positivamente e estudar, e a família apoiar o aprendizado e o engajamento dos estudantes das mais variadas formas.

A coordenação entre esses agentes depende, evidentemente, de se poder contar com uma escola com os recursos adequados e que funcione, mas certamente essa coordenação exige também um clima escolar adequado. De fato, a coordenação de esforços e cooperação requer um senso de objetivo e causas comuns, e um sentimento de pertencimento à comunidade escolar e um sentimento de confiança para os quais não basta que a escola funcione; é também indispensável que o clima escolar seja adequado. Para que uma cooperação eficaz possa de fato acontecer é indispensável que a relação entre os membros da comunidade (aluno, professores, pais, etc.) seja fraterna e justa, onde os conflitos sejam infrequentes e resolvidos pelo diálogo e de forma amigável, levando a que exista confiança e um senso de objetivo comum.

Em suma, o contexto escolar é (como deveria ser) o fator mais importante na determinação do desempenho educacional. Assim, seja quando se deseja investigar as causas de diferenças no desempenho educacional, seja quando se almeja realizar comparações significativas ou interpretar diferenças em desempenho entre escolas, é sempre indispensável contar com medidas do contexto escolar. Afinal, seria irrealista esperar que alunos em escolas com contextos mais vulneráveis alcancem o mesmo desempenho que alunos em escolas com contextos menos vulneráveis. Como vimos existem ao menos três visões complementares e, portanto, formas de medir a “qualidade” do contexto escolar: disponibilidade e qualidade dos recursos (insumos), “qualidade” do funcionamento (em que medida as atividades que deveriam ser realizadas o foram) e “qualidade” do clima escolar. Em princípio, qualquer medida abrangente de contexto escolar deveria abarcar medidas desses três componentes.

3.2. CONTEXTO FAMILIAR E COMUNITÁRIO

Quando o objetivo é alcançar uma descrição abrangente do contexto familiar e comunitário, em geral, ao menos três aspectos devem ser considerados. Em primeiro lugar, tem-se o grau de vulnerabilidade material das famílias e comunidades envolvidas. Esse aspecto inclui tanto medidas diretas de insuficiência de renda quanto manifestações indiretas, como a

escolaridade dos pais, as condições habitacionais, a posse de bens duráveis e a participação em programas sociais.

Em segundo lugar, há o aspecto relacionado a quanto o ambiente familiar e comunitário promovem, estimulam ou acolhem as atividades escolares. Esse aspecto envolve tanto a adequação do ambiente familiar às práticas escolares (informações sobre a disponibilidade de local iluminado e silencioso para estudar, alimentação e local para descansar adequados, entre outras), como também a disponibilidade de conhecimento (livros, revistas, jornais e conectividade à internet). Inclui também informações sobre a adequação dos espaços comunitários, que compreende a disponibilidade de oportunidades culturais, locais públicos seguros e estimulantes, em particular, a presença de biblioteca pública ou comunitária com amplo leque de atividades de promoção da leitura.

Por fim, há o aspecto relacionado à valorização da educação pela família e comunidade e o seu interesse e engajamento em atividades escolares. Esse aspecto inclui, portanto, informações sobre: *i)* o quanto famílias e comunidade valorizam e atribuem importância à educação, *ii)* o interesse dos pais e os incentivos que oferecem ao desempenho e engajamento dos seus filhos nas atividades escolares, e *iii)* o envolvimento dos pais e da comunidade nas atividades escolares, em particular, o quanto eles leem para os filhos, os ajudam nas tarefas escolares e criam um ambiente familiar e comunitário acolhedor para as atividades escolares.

A relação entre escola, família e comunidade é marcada por dois objetivos e que por vezes são difíceis de conciliar. Por um lado, como a família e a comunidade exercem grande influência sobre o desempenho educacional, seu apoio e sua participação nas atividades escolares deveriam ser os maiores possíveis e, portanto, deveriam ser estimulados ao máximo pela escola. Por outro lado, para garantir que exista igualdade de oportunidades, a escola deve garantir que o desempenho dos seus alunos seja independente do grau de vulnerabilidade e do engajamento de seus familiares e da comunidade nas atividades escolares.

Embora seja desafiadora, a missão da escola não é contraditória. Sua incumbência é encorajar o apoio e a participação das famílias e da comunidade e, ao mesmo tempo, garantir que todos os alunos, independentemente do apoio e da participação de seus familiares e comunidade, tenham as mesmas chances de se desenvolver, adquirir habilidades e se apropriar de conhecimentos.

A despeito de ser papel de a escola encorajar a participação das famílias e compensar sua falta, ao final, contudo, a escola é incapaz de neutralizar completamente o efeito da família e da comunidade. Por consequência, invariavelmente, os resultados educacionais acabam dependendo do contexto familiar. Assim sendo, é fundamental contar sempre com informações sobre o contexto familiar e comunitário, seja com vistas a avaliar o quanto a escola não foi

capaz de compensar a influência da família e da comunidade, garantindo igualdade de oportunidades, seja com vistas a permitir que comparações mais significativas do desempenho entre estudantes e escolas possam ser realizadas.

3.3. CONTEXTO PESSOAL

O contexto pessoal também influencia o desempenho educacional. Doenças crônicas e deficiências certamente terão impactos deletérios sobre o desempenho escolar sempre que a escola não for capaz de acomodar diferenças em necessidades. O mesmo ocorre com a presença de importantes atividades que concorrem com as escolares pelo tempo dos alunos, como aquelas decorrentes da gravidez e maternidade, e da inserção no mundo do trabalho. Invariavelmente a gravidez e a maternidade na adolescência levam a necessidades especiais (seja na garantia de atendimento ao bebê, seja em termos de flexibilidade de horários e locais das atividades escolares) que, se não forem atendidas pela escola, devem acarretar a queda no desempenho escolar das estudantes mães e grávidas.

De forma similar, o interesse pelo trabalho interage com as atividades escolares, podendo tanto estimular o estudo, como concorrer com ele pelo tempo do aluno. Assim, é importante que qualquer sistema de avaliação do desempenho escolar também conte com informações sobre gravidez na adolescência e engajamento com o trabalho, entre outras atividades concorrentes com a escola.

Também não seria de se esperar que alunos com significativos déficits de aprendizado ou com elevada defasagem série-idade tivessem o mesmo desempenho educacional daqueles que frequentam determinada série na idade correta e sem qualquer déficit de aprendizado. Assim, a trajetória e o histórico escolar importam para o desempenho presente e, portanto, essas informações precisam ser levantadas e sistematizadas.

As expectativas e percepções dos alunos também são importantes determinantes de seu interesse e engajamento nas atividades escolares e, consequentemente, de seu desempenho educacional. Por isso, informações das expectativas dos alunos sobre a continuidade dos estudos e pretensões com relação à educação superior são fundamentais. Da mesma forma, a sua percepção sobre o valor da educação para a vida e, em particular, do que se ensina na escola, e o seu interesse por essas atividades é essencial para podermos interpretar corretamente as diferenças em desempenho e em engajamento escolar entre estudantes, turmas, escolas e redes de ensino.

Por fim, vale também ressaltar a importância de contar com informações sobre o perfil socioemocional dos alunos. As atividades escolares são, na sua vasta maioria, atividades

coletivas. Tipicamente, é a partir de sua interação com professores e colegas que um aluno aprende, se desenvolve e adquire habilidades e competências. Não é surpreendente, portanto, que o desempenho de um aluno na escola dependa da sua facilidade e habilidade de criar, desenvolver e manter relacionamentos saudáveis.

Também o aprendizado e a aquisição de habilidades envolvem incertezas e retrocessos, podendo gerar sentimentos de ansiedade e frustação. Por isso, não surpreende o fato de o desempenho de um aluno na escola depender da sua habilidade de lidar e gerenciar seus pensamentos, emoções e comportamentos. Nesse sentido, o desempenho escolar de um aluno depende de suas competências socioemocionais; estas, portanto, devem ser devidamente monitoradas se queremos compreender as causas do sucesso escolar e realizar comparações significativas entre alunos, turmas e escolas.

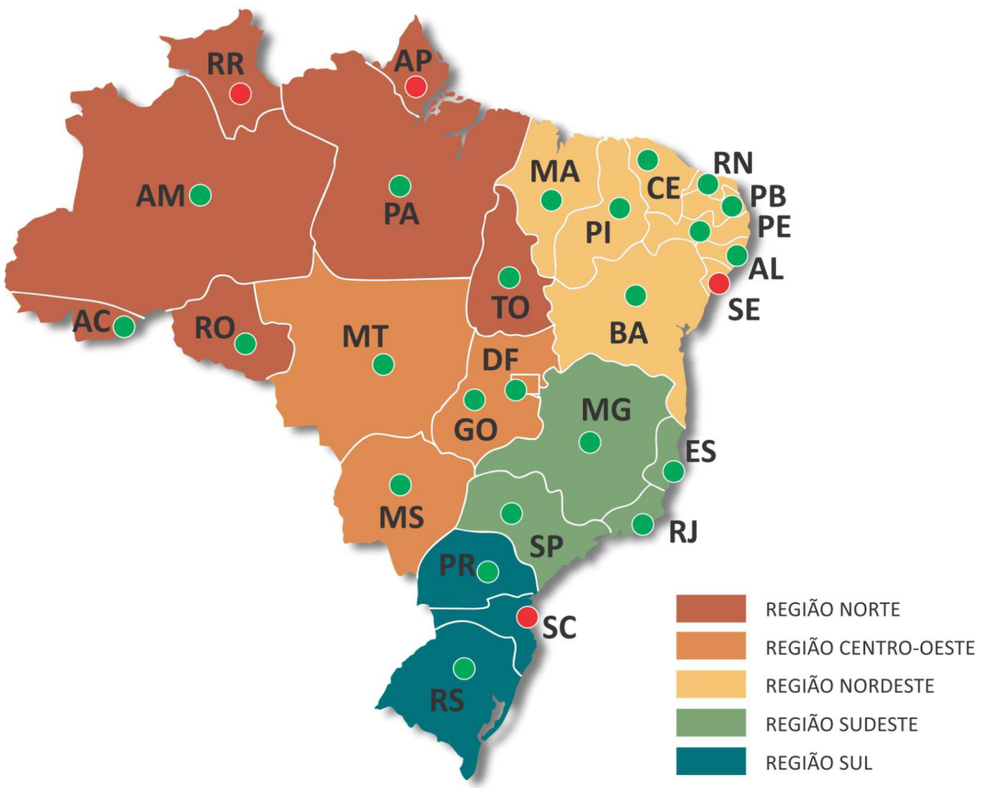
Em suma, informações sobre o contexto pessoal são fundamentais sempre que o objetivo da avaliação de desempenho educacional é a identificação de seus determinantes ou quando existe a necessidade de realizar comparações significativas que levem em consideração diferenças em oportunidades, condições ou tratamento.

4. EXPERIÊNCIAS EM AVALIÇÃO

4.1. AVALIANDO O APRENDIZADO

O Brasil tem ampla experiência com avaliações em escala da proficiência dos alunos. Em nível nacional temos a Prova Brasil, o Sistema Nacional da Avaliação da Educação Básica – SAEB e o Exame Nacional do Ensino Médio – ENEM, além da *Avaliação Nacional* da *Alfabetização* – *ANA*, da Provinha Brasil e do Exame Nacional para Certificação de Competências de Jovens e Adultos – ENCCEJA. Em nível internacional participamos, a cada três anos, do Programa Internacional de Avaliação de Alunos – PISA. Além disso, a maior parte dos Estados tem (ou já tiveram) um sistema próprio de avaliação da proficiência de crianças, adolescentes e jovens. De fato, de todas as 27 unidades da federação, apenas quatro não têm um sistema próprio de avaliação (veja Tabela 1). Por fim, muitos municípios de grande porte também já contam com um sistema próprio de avaliação de aprendizado.

Tabela 1: **Unidades da Federação com experiência em avaliação externa da educação básica**



Embora a experiência em nível nacional esteja historicamente concentrada na avaliação da proficiência em Língua Portuguesa (essencialmente leitura) e em Matemática (resolução de problemas), focos exclusivos da Prova Brasil e do SAEB[5], o ENEM e o ENCCEJA já compreendem todas as áreas do conhecimento presentes nas Diretrizes Curriculares Nacionais e no documento da Base Nacional Comum Curricular. Os sistemas estaduais também têm grande concentração na avaliação da proficiência em Língua Portuguesa e Matemática. De fato, apenas quatro sistemas estaduais atualmente cobrem outros componentes curriculares, além de Língua Portuguesa e Matemática. Mas, de toda forma, a experiência conjunta das 27 unidades da federação é muito rica – avaliações explícitas e específicas de 10 dos 14 componentes curriculares foram realizadas nos últimos anos por algum sistema estadual de avaliação. Apenas a proficiência em Ensino Religioso, Arte e Educação Física não foram objeto de avaliação específica em algum sistema estadual de avaliação.

[5] Vale ressaltar que Língua Portuguesa e Matemática também são os focos da Provinha Brasil e da ANA.

Em nível internacional, o PISA, do qual o Brasil participa, avalia a proficiência com relação a três componentes curriculares: Linguagem, Matemática e Ciências da Natureza. Uma análise por países revela que, em todos os casos, a avaliação inclui Linguagem (incluindo, na vasta maioria das vezes, Idioma Estrangeiro – geralmente o Inglês) e Matemática. Em grande parte dos casos a área de Ciências da Natureza também é incluída na avaliação.

Idealmente a avaliação da proficiência deveria cobrir todos os componentes curriculares que congregam hoje a base curricular nacional. A Matriz de Referência da avaliação (descritores) deve ser compatível com os objetivos de aprendizado e de desenvolvimento das Diretrizes Curriculares Nacionais – DCNs, passando a ser representativa dos objetos de conhecimento e habilidades presentes na Base Nacional Comum Curricular – BNCC, assim que esta for aprovada[6]. Os componentes curriculares que formam tanto as DCNs como a BNCC são tradicionalmente organizados em quatro áreas do conhecimento (Linguagem, Matemática, Ciências da Natureza e Ciências Humanas). A questão para o desenho do Sistema Nacional de Avaliação da Educação Básica – SINAEB e para o desenho dos sistemas de avaliação estaduais, portanto, não é se a proficiência em todos os componentes curriculares deve ser avaliada, e sim se deve existir uma avaliação (escala) para cada um desses componentes em separado, ou até mesmo se a avaliação e as escalas devem ser estruturadas por área de conhecimento.

4.2. AVALIANDO AS COMPETÊNCIAS GERAIS

A BNCC, assim como as DCNs e os PCNs, considera que o desenvolvimento de um amplo leque de competências gerais é parte central dos objetivos da educação. Na BNCC essas competências gerais estão organizadas em dez itens, que, por sua vez, se traduzem numa variedade de competências específicas para cada área do conhecimento e componente curricular.

Embora existam experiências, tanto nacional como estaduais[7], no que se refere a avaliação de competências gerais, não existem Matrizes de Referência, descritores ou bancos de itens, ao contrário do que ocorre com a avaliação da proficiência nas áreas do conhecimento tratadas na Prova Brasil, SAEB ou ENEM.

[6] A BNCC é fruto de amplo processo de debate e negociação com diferentes atores do campo educacional e com a sociedade brasileira encontrando-se hoje em sua terceira versão (a primeira versão do documento foi disponibilizada para consulta pública entre outubro de 2015 e março de 2016).

[7] Oito unidades da federação já vêm avaliando, ainda que de forma incipiente, esse tipo de competência (CE, ES, MA, PA, PB, PI, RJ, RS).

Segundo a BNCC, as DCNs e PCNs, em última instância, o objetivo das competências gerais é formar um jovem capaz de conhecer o mundo e a si mesmo, capaz de mobilizar esses conhecimentos para formular e avaliar alternativas para a sua vida, escolhendo e implementando de forma autônoma, responsável e independente uma dentre as alternativas disponíveis. Para isso, é indispensável que os jovens adquiram na escola valores democráticos, sustentáveis, inclusivos e solidários, valorizando a diversidade e desenvolvendo a identidade social e cultural. Além disso, a BNCC enfatiza a necessidade de a escola desenvolver no aluno o raciocínio crítico e a criatividade, além da capacidade de focar em seus objetivos com perseverança e resiliência, sendo, no entanto, flexível e sempre aberto a seguir caminhos novos para alcançar os objetivos finais almejados. Segundo a BNCC, as competências gerais devem servir também para que os jovens sejam capazes de participar de forma autônoma, eficaz e responsável da escolha e implementação de decisões coletivas. Para isso, eles precisam desenvolver ao longo de sua vida escolar a empatia, a assertividade e as capacidades de diálogo e cooperação.

Em suma, se devemos monitorar a garantia do direito à educação e se as competências gerais já estão presentes na legislação educacional brasileira há algum tempo – enfatizando que o desenvolvimento do pensamento crítico e criativo, a adoção de uma série de valores éticos (como democracia, sustentabilidade, solidariedade e apreciação pela diversidade) e um amplo leque de competências socioemocionais (como empatia, assertividade, abertura ao novo, determinação e flexibilidade) também constituem os objetivos da educação – estas precisam ser monitoradas.

Resta, no entanto, especificar o conjunto dessas competências gerais de forma a possibilitar o desenvolvimento de medidas que permitam avaliá-las em escala e também o acúmulo de experiência em sua mensuração e em uso na sala de aula. Essa é a grande lacuna atual dos sistemas de avaliação nacional, estadual e também internacional. Se valores éticos, formas de raciocínio e competências socioemocionais devam ser medidas e incluídas no sistema de avaliação, duas questões permanecem abertas. Por um lado, quais devem ser as competências e os valores a serem avaliados e, por outro, qual a melhor forma de realizar esta avaliação. Esses são desafios enfrentados pelos sistemas estaduais de avaliação para os quais não existe um amplo leque de experiências nacional ou internacionais.

4.3. AVALIANDO OS CONTEXTOS

Vimos, anteriormente, a importância de reconhecer a relevância e coletar informações de ao menos quatro tipos de contextos: escolar, familiar, comunitário e pessoal. Quanto ao contexto escolar, vimos que este é formado de ao menos três componentes: *i)* disponibilidade e qualidade dos insumos, *ii)* qualidade do funcionamento da unidade escolar, e *iii)* qualidade do

clima escolar. No Brasil, informações sobre disponibilidade e qualidade dos insumos já vêm sendo coletadas há décadas, nos Censos Escolares. Essa informação, embora rica com relação à disponibilidade dos insumos, em geral, é muito menos informativa sobre a qualidade dos insumos disponíveis. Seguindo essa linha de argumentação, não precisamos aumentar a quantidade de informações sobre a disponibilidade, e, sim a quantidade de informações sobre a qualidade. E enquanto há ampla limitação de informações disponíveis sobre a qualidade do funcionamento da escola, as informações sobre a qualidade do clima escolar, obtidas a partir de entrevistas com alunos, professores e diretores, estão disponíveis para uma amostra significativa de escolas.

Tanto no Brasil como no exterior, os contextos familiares e comunitários continuam a influenciar o desempenho educacional dos estudantes, a despeito da busca dos sistemas educacionais por maior igualdade de oportunidades. Informações sobre o contexto familiar/comunitário permitem documentar melhor a sua associação com o desempenho dos estudantes, desenhar políticas para combater esta associação, realizar comparações significativas entre escolas e estabelecer metas específicas. Atualmente, todos os sistemas de avaliação nacional, como Prova Brasil, SAEB e ENEM, coletam informações sobre o contexto familiar, mas não sobre o comunitário.

O mesmo ocorre em nível internacional. Como no caso do PISA, sempre que uma avaliação do desempenho educacional é realizada, informações sobre o contexto familiar são também coletadas. No caso dos sistemas estaduais, embora a coleta de informações sobre o contexto familiar não seja universal[8], em geral, essas informações são coletadas sempre que a proficiência é avaliada. Como vimos, informações sobre o contexto familiar podem ser divididas em três grandes grupos: *i)* disponibilidade de recursos, *ii)* ambiente familiar promotor e acolhedor do engajamento dos estudantes nas atividades escolares, e *iii)* engajamento dos familiares nas atividades escolares, estímulo à participação de suas crianças e adolescentes nessas atividades, e crença na importância da educação e das atividades escolares, em particular. No que se refere à coleta dessas informações, as experiências nacional e estadual, assim como a mundial, são similares. Em todos esses níveis, a disponibilidade de recursos (em particular, a escolaridade dos pais) é quase que universalmente investigada. O engajamento e incentivo dos pais (por exemplo, se os pais leem para os filhos) estão presentes em boa parte das avaliações realizadas. Já as informações sobre a capacidade do ambiente familiar em promover as atividades escolares (ambiente silencioso, iluminado, etc) são menos comuns.

Uma vez que praticamente todos os sistemas estaduais de avaliação incluem, de alguma forma, aspectos do contexto familiar em suas avaliações, o ponto controverso passa a ser a

[8] Existem três sistemas estaduais de avaliação que coletam informações sobre o aprendizado em Matemática e Língua Portuguesa, mas não coletam informações sobre o contexto familiar.

adoção de um padrão comum de levantamento. Estabelecer esse padrão comum de coleta é fundamental para tornar as informações coletadas comparáveis entre todos os sistemas em que as avaliações estão sendo conduzidas. Além disso, pode ser também necessário ampliar o leque de informações sobre a natureza promotora e acolhedora de atividades escolares características do ambiente domiciliar (ambiente silencioso, iluminado, arejado, etc.).

Por fim, ao menos os aspectos centrais do contexto pessoal também são tradicionalmente incluídos nas avaliações da proficiência dos alunos. Esses aspectos incluem as características demográficas básicas do aluno (sexo, idade e cor), assim como informações sobre o seu histórico escolar e interesse/motivação para continuar os estudos. Informações dessa natureza estão presentes na Prova Brasil, SAEB, ENEM e na maioria dos sistemas de avaliação estaduais. É surpreendente que, apesar da reconhecida influência que as condições de saúde, a inserção no mundo do trabalho e a gravidez e maternidade na adolescência exercem sobre o desempenho educacional, muito raramente informações dessa natureza são coletadas. Essa é uma falha grave e totalmente desnecessária que precisa de atenção no desenho dos novos sistemas estaduais de avaliação.

No que se refere ainda ao contexto pessoal, existem determinadas competências socioemocionais que são indispensáveis para um engajamento eficaz nas atividades escolares (gestão dos sentimentos, pensamentos, atenção e comportamento) e que, consequentemente, afetam o desempenho escolar futuro. Essas competências socioemocionais não têm estado presentes de forma sistemática nas avaliações nacionais ou estaduais. Nesse caso, no entanto, a mensuração pode ser complexa, requerendo o desenvolvimento de novas medidas.

CAPÍTULO 4

PROPÓSITOS, USOS, UTILIDADES E POTENCIAIS CONSEQUÊNCIAS NEGATIVAS DE UM SISTEMA DE AVALIAÇÃO DA EDUCAÇÃO BÁSICA

PROPÓSITOS, USOS, UTILIDADES E POTENCIAIS CONSEQUÊNCIAS NEGATIVAS DE UM SISTEMA DE AVALIAÇÃO DA EDUCAÇÃO BÁSICA

1. O DIREITO AO APRENDIZADO E A SUA VERIFICAÇÃO

Educação é tradicionalmente entendida como a busca ***intencional*** por objetivos ***explícitos*** relacionados ao desenvolvimento e ao aprendizado dos alunos, objetivos esses que são particulares a cada uma das etapas da educação e que, idealmente, devem ser alcançados em idades específicas. Afinal, o direito ao aprendizado nada mais é do que garantir que todos os alunos alcancem esses objetivos de desenvolvimento e aprendizado na idade correta e na etapa adequada do ciclo educacional.

De fato, é precisamente essa característica constitutiva da educação que leva a que existam currículos e bases curriculares cujas funções são definir, de forma explícita e específica, quais devem ser os objetivos de desenvolvimento e aprendizado dos alunos que cada professor deve intencionalmente buscar alcançar ao longo de cada série/ano letivo.

Mas qual o sentido em buscar alcançar intencionalmente objetivos que não somos capazes de averiguar se foram alcançados? Faz sentido definir e procurar garantir que cada criança, adolescente ou jovem alcance certos objetivos explícitos de desenvolvimento e aprendizado (especificados nos currículos) na idade correta, mesmo quando esses objetivos não são verificáveis? Faz algum sentido definirmos como direito humano algo que não somos capazes de verificar? O fato de que os objetivos de desenvolvimento e aprendizado precisam ser organizados em camadas, uma vez que o aprendizado de um conteúdo ou o desenvolvimento de uma competência invariavelmente requer o domínio de outras, é um fator a mais a demandar que o alcance de cada um desses objetivos seja verificável. Caso contrário, não é possível saber quando prosseguir para uma nova camada.

Para os educadores, que perseguem intencionalmente esses objetivos, é essencial saber se foram alcançados ou não, seja como uma medida de êxito ou como uma forma de avaliar a situação em cada etapa e definir de que forma redirecionar os esforços para que os objetivos da próxima etapa sejam atingidos com êxito.

O desenvolvimento de competências e o aprendizado de conteúdos, em geral, requerem substancial esforço e atenção de crianças, adolescentes e jovens. Assim, dada a sua importância na vida futura desses indivíduos, nada mais acertado do que eles também saberem se os objetivos de desenvolvimento e aprendizado propostos foram alcançados. De fato, a confirmação de que esses objetivos foram atingidos poderia fornecer a essas crianças, adolescentes e jovens, além de um importante senso de realização, a devida autoconfiança para realizarem tarefas que utilizem esses conhecimentos e competências.

Adicionalmente, na medida em que o direito ao aprendizado é definido como a garantia de que toda criança, adolescente e jovem alcance determinados objetivos de desenvolvimento e aprendizado na idade correta, passa também a ser função de todos os envolvidos na garantia desse direito, em particular dos gestores educacionais, que o alcance desses objetivos seja verificável. Caso contrário, não seria possível saber nem em que extensão o direito ao aprendizado está sendo garantido nem quando ou de que forma a política educacional deve ser alterada para que eventuais falhas na garantia desse direito sejam identificadas e corrigidas.

Em princípio, nada impede que a verificação da garantia do direito ao aprendizado tenha efeitos danosos sobre o bem-estar de alguns grupos de educadores ou estudantes. Nesse caso, é necessário ponderar os ganhos de se poder verificar a extensão em que o direito ao aprendizado é garantido e os efeitos perversos que essa verificação possa causar. Faz pouco sentido garantir um direito que não pode ser verificado e é pouco provável que a forma de verificação não possa ser ajustada de forma a mitigar a maior parte de seus efeitos danosos.

Supondo que os benefícios da verificação da garantia do direito ao aprendizado superam eventuais efeitos danosos, o que resta saber é ***como aferir*** e o ***que fazer*** com essa informação de forma a maximizar os benefícios dessa mensuração e minimizar seus efeitos colaterais perversos. O ***como aferir*** e o ***que fazer*** essa informação estão mais inter-relacionados do que se poderia imaginar *a priori*. Neste capítulo, tratamos da questão dos usos da informação gerada pelo natural processo de verificação do alcance dos objetivos de desenvolvimento e aprendizado e também de como esses possíveis usos dependem da forma como essa informação é obtida. Em outras palavras, mostramos como os possíveis usos são delimitados pela natureza da informação coletada e, portanto, de forma reversa, como o interesse por determinados usos estabelece critérios sobre que informações devem ser coletadas e sobre como realizar essa coleta.

Sem dúvida que existe certa diferença entre identificar se um objetivo foi alcançado e avaliar a extensão em que foi alcançado. A identificação do êxito no alcance de um objetivo é uma informação de natureza binária: ou o objetivo foi ou não foi alcançado. Já a avaliação da extensão em que foi alcançado dá a ideia de uma medida contínua. Na verdade, tudo depende

da possibilidade de se estabelecer uma gradação na aquisição de uma competência ou conhecimento e da forma como os objetivos da educação são especificados. Se não há essa gradação, então identificar e avaliar são coincidentes. Mas se existe gradação, então resta saber se o objetivo de desenvolvimento (ou aprendizado) especifica ou não um nível mínimo a ser alcançado. Caso especifique, identificar e avaliar voltam a coincidir e correspondem à verificação de que ao menos o nível mínimo foi alcançado.

2. POR QUE NÃO MEDIR?

A necessidade de medir é intuitiva. Como disse William Edwards Deming, "Em Deus nós acreditamos; para todo o resto tragam a evidência". E, segundo Lorde Kelvin: "Aquilo que não se pode medir não se pode melhorar". Embora essas afirmações sejam um tanto exageradas, elas captam a ideia de que informação é essencial para a tomada de decisão, fazendo com que o debate gire, em geral, muito mais sobre qual informação coletar e como melhor utilizá-la na tomada de decisões do que propriamente se alguma informação deve ser coletada ou não. Conforme o próprio Lorde Kelvin ressalta, "... apenas medindo … podemos testar se nossas conjecturas são falsas ou verdadeiras"[1].

Mas evidentemente não existe nada mágico ou sagrado a respeito de medir ou avaliar. Como qualquer outra, medir pode ser também uma má ideia. Certamente existem casos em que não medir seria melhor. Negar que situações dessa natureza possam existir não ajuda em nada o uso da mensuração; é preciso identificar esses casos e mitigar os fatores que podem tornar a mensuração indesejável.

Devemos ter em mente que, em alguns casos, nem ao menos sabemos medir e, portanto, essa não é sequer uma possibilidade viável. É sempre possível, porém, contra argumentar que alguma medida poderá ser desenvolvida. Mesmo aceitando esse argumento (não necessariamente correto), restaria a questão da qualidade dessa medida. Como medir tem custos e pode ter consequências danosas, quando a medida é muito ruim, pode ser preferível

[1] A citação original completa é a seguinte: "*Although reason and imagination also advance knowledge, only measurement makes it possible to observe patterns and to experiment—to put one's guesses about what is and is not true to the test*" William Thomson – Lorde Kelvin (1824-1907).

não medir. Qual é o limite aceitável de perda de qualidade da medida? O quão ruim pode ser uma medida que mesmo assim ainda vale a pena utilizá-la? Depende do custo de coletá-la e da importância dos efeitos perversos que o seu uso possa provocar.

Embora os efeitos danosos da mensuração possam ter naturezas distintas, invariavelmente decorrem ou diretamente da mensuração em si ou da qualidade da medida ou, indiretamente, do uso que se faça dela. De fato, se o ato de medir não causasse nenhum tipo de efeito e os resultados obtidos fossem imediatamente descartados, então a avaliação não teria qualquer consequência. Em qualquer mensuração, devemos sempre atentar para três tipos de cuidados.

Em primeiro lugar devemos ter cuidado para que a participação no processo de mensuração seja realizada de forma a não causar efeitos negativos sobre os participantes, o que poderia inclusive comprometer a qualidade das informações prestadas. Embora toda mensuração ou avaliação envolva algum esforço físico e mental/emocional, o ideal é que a mensuração seja feita da forma menos penosa possível. Daí a necessidade de se limitar o tempo da entrevista, de evitar questões constrangedoras e de contar sempre com o consentimento explícito daqueles envolvidos na avaliação.

Embora parte dos efeitos danosos de uma avaliação resulte diretamente da coleta das informações, a inadequação da medida ao uso que dela se pretende fazer também responde por boa parte de eventuais consequências negativas de uma avaliação. Assim, em segundo lugar, em toda avaliação as medidas utilizadas precisam ser válidas para medir o construto de interesse e terem um grau de precisão conhecido e adequado aos propósitos da avaliação. É importante que haja uma análise detalhada da adequação entre o que vai ser medido, por um lado, e os objetivos da avaliação e os usos que dela serão feitos, por outro. Medidas adequadas para um tipo de uso não necessariamente serão úteis para outras funções.

Em terceiro lugar, a principal fonte de consequências perversas de uma avaliação invariavelmente é o seu uso indevido, que atua via três canais. Para começar, temos a questão do acesso impróprio. Nem sempre as informações de uma avaliação devem se tornar plenamente públicas. Em muitos casos a privacidade de estudantes, professores e escolas deve ser preservada. Portanto, para evitar a ocorrência de uso inapropriado é necessário que exista um protocolo apontando quem pode ter acesso a cada informação e, em cada caso, para que tipo de uso tem esse acesso liberado. Muito do debate sobre os efeitos danosos da avaliação em educação resultam de discordâncias sobre quem deveria ter acesso às informações e para que tipo de uso. Além disso, temos as situações que resultam da falta de conhecimento dos usuários sobre a medida, suas propriedades e limitações. Nesse caso, os usuários acabam ou utilizando ou interpretando de forma inadequada as informações disponíveis ou, dada a forma

como divulgam a informação, induzem os outros aos mesmos erros. Por fim, pode ocorrer que o usuário seja conhecedor da medida, suas propriedades e limitações, mas o que é indevido é a forma de divulgação ou o uso que é feito da informação. Por exemplo, mesmo que seja possível identificar com absoluta precisão quais são os alunos de uma escola com maior déficit de aprendizado, pode ser um mau uso dessa informação anunciar quem são eles na reunião com todos os alunos da escola presentes. Em outras palavras, um uso conceitualmente, estatisticamente e logicamente correto das informações não necessariamente leva sempre a um uso adequado e produtivo dessa informação e que, ao menos no longo prazo, pode levar a uma melhoria no desempenho do sistema educacional ou garantir o direito ao aprendizado para um leque mais amplo de estudantes.

Em resumo, esses são três tipos de fatores que podem levar a que uma avaliação tenha efeitos perversos, eventualmente significando que não avaliar seja preferível a avaliar.

Vale ressaltar que outro fator que pode pesar nesse balanço contra ou a favor da realização de uma avaliação, embora não perverso em si, é o uso limitado dos resultados obtidos. Uma vez que toda avaliação tem custos e consequências perversas, é importante que os resultados sejam utilizados da forma mais intensa e adequada possível. De fato, toda mensuração e toda avaliação podem ser utilizadas numa variedade de formas.

Neste capítulo organizamos os usos de uma avaliação em sete grandes grupos. O significado de cada um desses usos é explorado em detalhe na seção 3. No momento, o importante é ressaltar que a ausência de certas condições pode limitar ou mesmo inviabilizar alguns dos usos possíveis e, portanto, limitar desnecessariamente os benefícios da avaliação. O uso integral de avaliações requer *i)* a disponibilidade de recursos humanos que entendam e saibam utilizar a avaliação, *ii)* a disponibilidade de atividades e ações que permitam corrigir desvios ou insucessos identificados pela avaliação, e *iii)* práticas institucionais de planejamento e gestão com base em evidência que desperte o interesse pelas informações produzidas a partir da avaliação. Muitas vezes essas condições não estão presentes, limitando sobremaneira a utilidade e os benefícios da avaliação. Pode ocorrer, no entanto, que a experiência de avaliações influencie a própria existência das condições necessárias ao seu pleno uso. O fato de que a disponibilidade das avaliações pode também levar ou induzir ao surgimento das condições necessárias ao seu pleno uso pode ser considerado como uma razão adicional para que avaliações sejam realizadas.

Dada a possibilidade de efeitos danosos, avaliar e não avaliar são em princípio opções que precisam ser investigadas caso a caso. Nenhuma avaliação pode ou deve ser realizada sem que suas vantagens e desvantagens sejam escrutinadas. Um procedimento universalmente recomendado é o de nunca iniciar ou realizar uma avaliação ou mensuração sem que antes

fique bem estabelecido *i)* qual o custo da avaliação, em particular os físicos e mentais impostos aos que serão avaliados, *ii)* qual a qualidade da medida que se pretende utilizar, *iii)* quais os usos que serão feitos dos resultados, *iv)* quem terá acesso às informações e para que tipo de uso, e *v)* quais as consequências danosas potencias e esperadas da avaliação ou mensuração. De posse dessas informações, planos para mitigar os riscos das consequências perversas devem ser implantados, e uma análise dos custos e benefícios da avaliação deve ser realizada com vistas a decidir sobre a propriedade de se conduzir a avaliação.

Dados os delicados aspectos éticos envolvidos em qualquer avaliação envolvendo seres humanos, como no caso da educação, é imperativo que toda avaliação tenha um desenho e um plano de uso das informações pré-definido e transparente. É de grande importância que seja publicamente conhecido o que será medido, quem será avaliado, quando e por quem. Também deve ser transparente quem terá acesso às informações e com que propósito. Afinal, um sistema de avaliação educacional não pode medir o que quiser, sobre quem quiser, a hora que quiser e, tampouco, permitir que qualquer um possa usar a informação coletada com o propósito que quiser.

É importante também selecionar com propriedade quem decide, e com que critérios, a realização ou não de uma avaliação. Dado que a avaliação pode ser inconveniente para determinados grupos de educadores e estudantes, é imprescindível que os responsáveis pela decisão sobre a realização ou não de uma avaliação estejam livres desse tipo de conflito de interesse e, na medida do possível, protegidos da pressão dos grupos de interesse. Via de regra, a avaliação é de grande valia para aqueles que participaram do processo educacional (estudantes e educadores) que gostariam de saber em que medida seus esforços levaram aos resultados esperados; também é de grande interesse para os responsáveis por garantir que os objetivos educacionais sejam alcançados (gestores educacionais); por fim, a sociedade como um todo, na medida em que investe no processo educacional, deve também se interessar em saber se os objetivos esperados foram de fato alcançados. É, portanto, vital considerar os interesses desses grupos no momento de decidir pela avaliação ou não.

3. USOS DA AVALIAÇÃO

O objetivo final de toda avaliação deve ser a promoção do desenvolvimento e do aprendizado dos estudantes. No entanto, cada avaliação influencia o desenvolvimento e o aprendizado de diferentes formas. De maior importância é o fato de que uma mesma avaliação pode ser utilizada de distintos modos e, portanto, exerce sua influência por distintos canais.

Em princípio, uma vez realizada a avaliação, a informação obtida pode ser utilizada para os mais variados usos. Na verdade, quanto mais essa informação for utilizada, melhor será a relação custo-benefício da avaliação, uma vez que o custo referente a cada uso adicional deverá ser limitado e o benefício significativo. Em outras palavras, cada uso adicional deve elevar os benefícios da avaliação de forma significativa, mas seus custos apenas de forma marginal. Assim, quanto maior a variedade de usos da avaliação, mais diluído estará seu custo fixo[2] e mais favorável deverá ficar sua relação custo-benefício.

Há uma infinidade de formas de como as avaliações podem ser utilizadas. Nesta seção organizamos os possíveis usos em sete grupos, que passamos a descrever brevemente. Esses usos vão desde os mais sistêmicos, voltados para a prestação de contas, até os mais pedagógicos, dentro da sala de aula.

1. Prestação de contas: iniciando pelo uso mais sistêmico, a avaliação pode ser utilizada como um instrumento de prestação de contas. Nesse caso, o que se deseja extrair da avaliação é apenas a evolução dos indicadores de desenvolvimento e de aprendizado dos alunos, isto é, em que medida o direito ao aprendizado está ou não sendo garantido. A avaliação, portanto, serve para que todos possam acompanhar como têm evoluído o desenvolvimento e o aprendizado dos alunos. A partir desses indicadores é possível identificar se houve progresso ou retrocesso, qual a sua magnitude, ou, ainda, se os resultados permaneceram estagnados.

A depender do interesse e, consequentemente, da forma como é feita, a avaliação precisará gerar tanto indicadores agregados do desempenho do sistema educacional, como também indicadores desagregados por regional administrativa, escola, turma ou mesmo por aluno. Para a sociedade como um todo, importa o desempenho agregado, indicando se há evolução, e se a velocidade de progresso está abaixo ou acima do esperado. Para os gestores centrais e diretores de escola, além dos resultados agregados, é importante também analisar os resultados em nível de regionais e de escolas, de forma a identificar aquelas com limitado progresso e que requerem, portanto, maior assistência e, possivelmente, mais recursos. É importante que se identifique as regionais e as escolas com melhor desempenho, uma vez que essas podem apontar para as melhores práticas a serem adaptadas e utilizadas por todas. Coordenadores pedagógicos podem querer conhecer o desempenho de cada uma das turmas sob sua direta responsabilidade, de forma que sejam capazes de identificar aquelas que mais necessitam de sua atenção e aquelas que, por seu desempenho, podem servir como exemplos a serem seguidos. Professores podem se interessar pelo resultado de suas turmas para saberem como tem sido seu desempenho em comparação com seus pares na mesma escola ou com

[2] No campo das ciências econômicas, a Teoria da Firma define que custos fixos são aqueles que independem do nível de atividade e, portanto, não sofrem alterações de valor em caso de aumento ou diminuição da produção.

outras escolas do mesmo sistema educacional. Por fim, alunos e pais podem se interessar pelos resultados individualizados na medida em que permitem conhecer o desempenho de um aluno e compará-lo com os dos demais na mesma turma, escola, regional ou sistema educacional em que se insere.

Vale ressaltar que é precisamente com base na própria experiência da sociedade, dos sistemas educacionais, das regionais e das escolas que realizam essas avaliações que metas e expectativas de progresso para os anos seguintes são dimensionados. Invariavelmente, as metas são baseadas na capacidade de evoluir no passado e na magnitude das diferenças entre sistemas educacionais e escolas. Assim, pode-se dizer que um dos subprodutos da realização de avaliações e de seu uso para a prestação de contas é o desenvolvimento da capacidade da sociedade, dos sistemas educacionais, regionais e escolas de formularem metas desafiadoras, porém factíveis, que permitam criar expectativas sobre a evolução futura e, posteriormente, julgar o desempenho alcançado. Portanto, avaliações com o objetivo de prestar contas, na medida em que desenvolvem a capacidade dos sistemas educacionais, regionais e escolas para formularem suas próprias metas, têm como importante efeito colateral o desenvolvimento da capacidade e do interesse em formas de gestão orientadas por resultados.

2. Sistema de incentivos: a educação, como a maioria dos serviços públicos, é intensiva em recursos humanos. Tanto é assim que, embora o desenvolvimento e o aprendizado dos alunos possam depender de uma miríade de fatores, é do desempenho, do envolvimento do professor e da sua interação com eles que esses resultados realmente dependem. Nenhum outro fator compete com o professor quando o que está em discussão é o impacto sobre o desenvolvimento e o aprendizado dos alunos. Além disso, a educação precisa ir aonde os alunos estão, para que tenha a capilaridade que a garantia do direito à educação requer e, por esse motivo, ocorre necessariamente de forma espacialmente dispersa. Lidar com a diversidade e os costumes locais requer considerável descentralização administrativa e pedagógica, na medida em que a educação precisa não apenas ter capilaridade, mas também se adequar o máximo possível às peculiaridades e necessidades locais.

Toda instituição com administração altamente descentralizada e que opera com alta capilaridade – que deve se adequar às especificidades locais e invariavelmente depende da qualidade do envolvimento de seus recursos humanos – necessita de um sistema de incentivos muito bem desenhado. Independentemente se esse sistema irá trabalhar a motivação intrínseca ou a extrínseca, via incentivos morais ou materiais, a disponibilidade de informações confiáveis e comparáveis sobre os resultados alcançados em cada escola tem se mostrado indispensável. De fato, caso o objetivo seja promover a motivação intrínseca dos educadores e estudantes para perseguirem com mais afinco os objetivos de desenvolvimento e de aprendizado, é necessário que eles tenham informação de qualidade que permitam o acompanhamento, em

tempo real, da medida do alcance dos objetivos estabelecidos. Caso a estratégia seja trabalhar a motivação via incentivos morais, incentivando a preocupação de todos em alcançar resultados iguais ou acima de grupos de referência considerados relevantes e adequados, então mais uma vez são imprescindíveis medidas precisas do desempenho local (aluno, turma ou escola), comparáveis às dos grupos de comparação.

Além disso, quando um dado resultado é afetado pelas ações de múltiplos agentes, como tipicamente ocorre na educação, é desejável que o sistema de incentivos seja desenhado considerando todos os elos da cadeia que ligam as ações ao resultado desejado. Um sistema que premia ou penaliza exclusivamente um agente (p.e., o professor) pelo resultado alcançado por outro agente (p.e., o aluno) possivelmente não produzirá os objetivos almejados, se o resultado em questão depender não apenas da ação do professor, mas do próprio aluno ou do corpo escolar como um todo. Ou seja, para que sistemas de incentivos funcionem bem, é desejável que o seu desenho contemple todos os agentes que têm a capacidade de afetar o resultado desejado.

3. Diagnóstico da situação atual e do progresso educacional: anteriormente ressaltamos que um dos objetivos principais da avaliação é a identificação da situação atual e do progresso ocorrido nos principais resultados educacionais com vistas à prestação de contas. Para tanto, é também fundamental identificar fortalezas, fraquezas, eventuais retrocessos e estagnação. De fato, o uso diagnóstico da avaliação tem dois objetivos. Por um lado, identificar e priorizar os principais desafios educacionais e, por outro, determinar quais as causas e consequências da lentidão do progresso em alguns indicadores, ou mesmo do fracasso em outros. Com vistas a priorizar os desafios educacionais, todo diagnóstico precisa combinar: *i)* uma boa avaliação da situação atual e da velocidade com que progressos educacionais estão sendo alcançados, com *ii)* medidas da magnitude das consequências dos déficits educacionais. A importância de um déficit educacional depende tanto da sua magnitude como da consequência que cada unidade desse déficit deverá provocar. De posse da magnitude e consequência de cada aspecto do atraso educacional, é possível tanto calcular a sua consequência agregada como também identificar os aspectos que merecem prioridade, caso não seja possível atuar sobre todas simultaneamente.

Para que esses déficits sejam ajustados, é preciso desenhar ações eficazes que visem corrigi-los e, para isso, é indispensável conhecer as suas causas. Daí a importância de se investigar as causas e determinar a magnitude de suas influências sobre os déficits educacionais existentes. O conhecimento das causas irá permitir aos gestores públicos não só desenhar ou redesenhar ações voltadas a mitigar os déficits educacionais, como também selecionar, dentre as ações existentes, aquelas que são mais adequadas em função das causas que endereçam. Por fim, vale ressaltar que todas essas atividades diagnósticas são vitais no momento do

planejamento e replanejamento, quando prioridades e metas são definidas ou redefinidas. Para isso é fundamental conhecer a importância relativa de cada um dos componentes do atraso educacional, a viabilidade e os custos envolvidos na sua superação. Em boa medida, essa superação decorre da facilidade com que a origem desses déficits pode ser modificada e, portanto, do conhecimento detalhado de suas causas.

4. Fatores de sucesso e alocação de recursos: o sucesso e o fracasso certamente não são homogêneos. Em um dado momento há sistemas educacionais ou regionais e escolas de um mesmo sistema com desempenhos radicalmente distintos. O mesmo pode ocorrer com turmas de uma mesma escola e estudantes em uma mesma turma. Desempenho diferenciado de sistemas, regionais, escolas, turmas e alunos é uma característica típica em educação, como também na maior parte dos demais serviços públicos. Essa heterogeneidade é certamente indesejável na medida em que reflete desigualdade de oportunidade entre estudantes e implica que, mesmo quando (em média) há sucesso, nem todos estão se beneficiando dele. Mas, por outro lado, essa heterogeneidade revela que existem casos de sucesso, mesmo num momento em que (em média) o progresso esteja sendo lento. Esses casos, uma vez identificados e com seus fatores de sucesso reconhecidos, podem servir de referência para os demais sistemas ou escolas, indicando que medidas poderiam ser adotadas com o objetivo de superarem seus déficits ou atrasos. Assim, um dos usos de grande importância da avaliação é o mapeamento dos casos de maior sucesso e a identificação das suas causas e, em particular, a identificação dos fatores que poderiam ser replicados em outros contextos e instantes no tempo (os ditos fatores universais). Dessa forma, a avaliação estaria ajudando na difusão das melhores práticas, tanto permitindo a sua identificação como revelando e explicitando os fatores determinantes do sucesso.

Mas, além de viabilizar a identificação dos casos de sucesso, a avaliação permite também que os sistemas, regionais, escolas, turmas ou alunos com progresso mais lento sejam também identificados. A identificação desses casos é tão importante para a política educacional quanto a dos casos de sucesso, permitindo que os recursos existentes sejam alocados com maior prioridade. Ademais, na medida em que a avaliação é um insumo fundamental para que se identifique a razão para o relativo insucesso dessas unidades, ela não somente viabiliza a focalização da assistência técnica naqueles sistemas, regionais, escolas que mais precisam como também, na medida em que informa sobre as causas específicas do insucesso, torna essa assistência bem mais eficaz. Em suma, a avaliação permite identificar as melhores práticas, fazendo com que seja possível adequá-las às carências identificadas pela avaliação dos sistemas, regionais e escolas com pior desempenho.

5. Identificação de áreas em que novas práticas, ações e programas educacionais precisam ser desenhados: como vimos, as avaliações são vitais tanto para identificar e priorizar

medidas contra os déficits educacionais, além de identificar em que sistemas, regionais e escolas se concentram e quais as suas causas. De posse dessas informações, é possível trabalhar num correspondente mapa de soluções. Para isso, no entanto, é necessário também contar, para cada aspecto do déficit educacional, com um levantamento estruturado das práticas e ações disponíveis, que envolva uma clara identificação da causa em que cada uma dessas ações, práticas ou programas busca atuar. Além disso, é fundamental contar com informações sobre a eficácia de cada uma dessas intervenções (um dos usos da avaliação que será discutido na sequência).

De posse desse mapa das soluções, organizadas segundo as causas em que procuram operar, e de medidas da sua eficácia, pode-se, então, identificar as lacunas em ações, práticas ou programas educacionais. Ou seja, é possível identificar causas para as quais não existem ações com eficácia adequada e, portanto, para as quais é necessário desenhar novas ações. Assim, a identificação das áreas de atuação (causas de determinados déficits educacionais) em que novas ações, práticas e programas educacionais precisam ser desenhados é um importante uso das avaliações.

6. Avaliações *ex-ante* e *ex-post* da eficácia de ações, práticas e programas educacionais: o progresso educacional depende, em última instância, da eficácia das práticas, ações e programas educacionais implantados. Essa eficácia, por sua vez, é em grande medida função da qualidade intrínseca do desenho de cada uma dessas práticas, ações e programas. No entanto, como nem todas acabam tendo a mesma eficácia, o desempenho educacional também depende da alocação de recursos entre as ações implantadas: quanto mais os recursos são destinados às ações de maior impacto, maior será a eficácia do conjunto de ações.

As avaliações contribuem para a eficácia das ações por meio de dois canais: *i)* promovendo um desenho inicial intrinsecamente mais eficaz, e *ii)* fornecendo a informação necessária sobre a eficácia das ações existentes, o que permite uma melhor alocação dos recursos. Assim, a mensuração dos resultados educacionais oferece informações essenciais para a realização de avaliações *ex-ante* da eficácia de desenhos alternativos para políticas em gestação. Isto é, a mensuração dos resultados educacionais permite estimar o impacto de desenhos alternativos de uma ação, mesmo antes de a ação ser implantada e, dessa forma, permite que o desenho da ação seja, ao menos teoricamente, mais eficaz.

Além da contribuição para a definição do desenho das ações em gestação, a mensuração e o monitoramento dos resultados educacionais é um instrumento essencial à avaliação *ex-post* da eficácia das políticas implantadas. Em geral, a magnitude do impacto de uma ação, uma vez implementada, tende a ser distinta daquela prevista em avaliações *ex-ante*, que precederam sua implantação. Consequentemente, para que se avalie o efetivo impacto de cada ação

implantada, é importante também que sejam conduzidas avaliações *ex-post*. Esse tipo de avaliação requer a mensuração dos resultados educacionais de interesse, uma vez que é sobre esses resultados que o impacto das ações estará sendo avaliado. Em suma, a mensuração dos principais resultados educacionais é um insumo fundamental tanto às avaliações *ex-ante,* que irão possibilitar desenhos mais eficazes para as ações educacionais, como para as avaliações *ex-post*, que irão avaliar o efetivo impacto dessas ações, sendo, portanto, requerimento indispensável a uma alocação adequada dos recursos educacionais públicos.

7. Planejamento das ações, práticas e programas educacionais: o sucesso de um sistema educacional depende, em última instância, da adequação das ações implantadas às necessidades desse sistema e da eficácia dessas ações. De nada adiantaria podermos contar com um leque de ações eficazes em eliminar déficits que não representam aqueles que de fato existem. Tampouco seria relevante contarmos com ações voltadas para eliminar os déficits existentes se elas não são eficazes em desempenhar essa missão. Assim, para que as ações implantadas verdadeiramente eliminem os déficits existentes e promovam um bom desempenho educacional, é essencial que elas sejam simultaneamente adequadas e eficazes.

Para alcançar essa condição, é necessário se fundamentar num processo de planejamento que incorpore também novas e promissoras ações, sobre as quais recaiam excelentes expectativas em função de avaliações *ex-ante* que apontem para um alto grau de eficácia. Já na alocação dos recursos existentes entre as ações disponíveis, é essencial que se leve em consideração as necessidades do sistema identificadas no processo diagnóstico. Portanto, a definição de um plano de ação eficaz requer que o processo de planejamento contemple um profundo e detalhado diagnóstico das condições educacionais atuais e de sua evolução. Esse panorama explicita quais são os principais desafios a serem enfrentados e, dessa forma, quais os déficits educacionais que precisam ser eliminados e quais as principais causas desses déficits que precisam ser atacadas. Adicionalmente, para que o leque de ações seja eficaz, o processo de planejamento deve selecionar, dentre as ações novas e as já implantadas, aquelas que respondem aos desafios identificados pelo diagnóstico e que apresentam o maior impacto, segundo as avaliações de impacto *ex-ante* e *ex-post* disponíveis.

Em suma, o planejamento de sistemas, regionais e escolas depende das avaliações de resultados educacionais, na medida em que requer um diagnóstico detalhado da realidade e na medida em que essas avaliações permitem estimar o impacto das diversas ações, permitindo que o plano de ações estratégicas incorpore e aloque recursos às ações mais eficazes.

8. Adequação das práticas pedagógicas: segue, da própria forma como a educação é organizada, que a maior parte do desenvolvimento e aprendizagem dos alunos ocorre nas salas de aula, com o professor. Assim, para que o uso da avaliação dos alunos influencie de forma

significativa o seu desenvolvimento e aprendizado, os resultados da avaliação têm que chegar à sala de aula. Anteriormente vimos como diferentes usos da avaliação se complementam para, ao final, aprimorarem o planejamento das ações a serem implantadas. A avaliação influencia o desenho, a seleção de quais ações serão implementadas e a forma como essa implementação será conduzida.

Além desses usos, a avaliação também pode promover o desempenho educacional modificando diretamente o que ocorre dentro da sala de aula. Esses usos mais pedagógicos da avaliação podem ser organizados em três grupos. Em primeiro lugar, informações sobre os déficits de desenvolvimento e de aprendizado dos alunos de uma turma podem ser utilizadas para identificar lacunas e deficiências na formação dos professores e, dessa forma, serem de extrema valia na definição dos programas e tipos de formação que cada professor mais necessita. Assim sendo, a avaliação permite aprimorar o desenho e o conteúdo dos programas de formação de docentes de toda a rede de ensino. Em segundo lugar, a informação sobre o desempenho dos alunos permite identificar os déficits de desenvolvimento e de aprendizado específicos a cada turma, permitindo adequar as práticas pedagógicas às peculiaridades dos alunos. Por fim, em terceiro lugar, o conhecimento do processo de desenvolvimento de cada aluno e de sua proficiência nas mais diversas áreas do conhecimento permite identificar aqueles com maiores dificuldades. Uma vez identificados, é possível dar-lhes a atenção especial de que necessitam, em particular, adequando as práticas às suas especificidades, preferências e dificuldades.

Nesta seção argumentamos que a avaliação dos resultados educacionais alcançados pelos alunos pode ser utilizada de oito formas, com oito objetivos complementares. Na próxima, detalhamos o que cada uso da avaliação requer para ser efetivo. Esses requerimentos envolvem dois aspectos: *i)* as informações necessárias a cada uso, e *ii)* os recursos humanos e desenho institucional necessários para que cada um desses usos seja verdadeiramente eficaz.

4. REQUERIMENTO PARA EFETIVIDADE

Na seção anterior vimos que a avaliação dos resultados educacionais pode ser empregada de oito formas diferentes. Cada um desses usos, no entanto, requer insumos específicos para ser efetivo. Assim, nesta seção detalhamos quais são esses insumos. Mais especificamente, organizamos a seção em duas partes. Na primeira apresentamos as informações necessárias a cada uso apresentado no capítulo anterior e, na segunda, tratamos

dos recursos humanos e desenho institucional necessários para que cada um desses usos seja verdadeiramente eficaz.

4.1. INFORMAÇÕES NECESSÁRIAS

Todo sistema de avaliação deve definir o tipo de informação a ser coletada (por exemplo, proficiência em Matemática), que anos/séries incluir, com que periodicidade coletar essas informações e que cobertura adotar (apenas as escolas públicas ou também as particulares; todas as turmas de todas as escolas ou apenas uma amostra de turmas ou escolas?). Todas essas decisões dependem do uso que se queira fazer da avaliação e, portanto, cada uso tem uma demanda específica do tipo de informação que necessita. Na sequência, identificamos que tipo de informação é necessária para cada um dos oito usos propostos para a avaliação vistos na seção anterior.

1. Prestação de contas e estabelecimento de metas: quando o objetivo é simplesmente a prestação de contas, então a informação necessária é definida em consonância com as metas estabelecidas, isto é, os resultados educacionais a serem avaliados devem ser aqueles presentes nas metas. Afinal, é com relação a esses resultados que se deseja prestar contas. Por exemplo, como no Plano Nacional de Educação e na quase totalidade dos Planos Estaduais e Municipais de Educação, as metas são definidas em termos do Ideb. Assim, o sistema de avaliação deve informar os indicadores que compõem o Ideb (proficiência em Matemática e Língua Portuguesa e taxas de aprovação de todos os anos/séries).

Há também a questão dos anos/séries que o sistema de avaliação precisa cobrir. Novamente, esses anos/séries estão relacionados às metas que foram estabelecidas. Se as metas se referem a todos os anos/séries, então deverá ser coletada informação para todos eles. No entanto, como ocorre quando as metas estão atreladas ao Ideb, apenas avaliações da proficiência nos anos finais de cada ciclo (5º e 9º ano do Ensino Fundamental e 3ª série do Ensino Médio) precisam ser coletadas.

Quanto à periodicidade com que o sistema de avaliação coleta informações, essa deve refletir ao menos a periodicidade presente nas metas. Assim, se as metas são anuais, a avaliação deverá também ser capaz de gerar a informação com periodicidade ao menos anual; se as metas forem bianuais, então, a informação poderá ser obtida a cada dois anos.

Por fim, a capilaridade e a cobertura das informações devem refletir a correspondente desagregação do sistema de metas adotado. Se as metas são apenas para o sistema estadual

como um todo, como no caso das metas do PNE para o Ensino Médio, basta coletar informações para uma amostra de escolas de Ensino Médio em cada unidade da federação. Por outro lado, se existem metas para cada escola pública, é necessário que as informações sejam coletadas de forma a serem representativas de cada uma dessas escolas embora, nesse caso, nem todas as turmas precisem ser avaliadas – uma amostra de turmas deve ser suficiente. Por fim, se existem metas para o conjunto das escolas (públicas e particulares) num dado município ou estado, então, a avaliação irá precisar cobrir, além das escolas públicas, ao menos uma amostra representativa de escolas particulares.

2. Sistema de incentivos: para que um sistema de incentivos funcione de forma adequada é necessário que os indicadores utilizados na definição das metas (usadas para definir os incentivos morais ou materiais) estejam sendo coletados com a periodicidade compatível e com a abrangência e a capilaridade (desagregação) necessária à identificação do desempenho individual daqueles que se deseja incentivar. Ou seja, como incentivos em geral se baseiam em metas, todos os requisitos acima definidos para o seguimento de metas também se aplicam ao bom funcionamento de sistemas de incentivos. No entanto, para funcionarem adequadamente, alguns aspectos mais delicados da informação precisam ser observados.

Em primeiro lugar, é necessário que a informação utilizada tenha excelente, senão excepcional, fidedignidade e precisão. Todo indicador se altera devido às mudanças reais no construto que busca medir ou a erros de medida. Assim, quanto mais mudanças num indicador resultarem de efetivas mudanças no construto que busca medir, melhor ele servirá como instrumento para incentivar os responsáveis por sua mudança. De fato, mudanças num indicador cujas variações podem, em grande medida, ser atribuídas a erros de medida não são uma boa fonte de incentivo, uma vez que frente a uma mudança os agentes não saberão se é uma mudança real no construto que buscam modificar ou se é uma flutuação aleatória devido a um erro de medida.

Em segundo lugar, um sistema de incentivos precisa de informações mais rápidas que o necessário para a prestação de contas. Quanto maior a defasagem entre o momento que se busca promover uma mudança e o momento em que se verifica se a modificação de fato ocorreu, menor a eficácia do sistema de incentivos. Num sistema de incentivos ideal, a informação sobre mudanças no construto de interesse deve ser instantânea.

Em terceiro lugar, os incentivos e, portanto, as informações coletadas devem se adequar à unidade de decisão que se deseja incentivar. Assim, se a escola de fato funciona como uma unidade de decisão, então, informações de desempenho em nível da escola são suficientes. Mesmo nesse caso é importante que as informações sejam abrangentes e que cubram os diversos aspectos do desempenho da escola. Invariavelmente esse não é o caso. As informações

sobre escolas na verdade se referem ao aprendizado dos alunos de apenas algumas séries e em apenas algumas disciplinas ou áreas do conhecimento. Os resultados medidos não levam em consideração, de forma balanceada, a contribuição de todos os membros da comunidade escolar. Mas nem sempre a escola deve ser pensada como a unidade de decisão a ser incentivada. Pode ser que o professor ou o aluno sejam aqueles que se deseja incentivar. Nesse caso, informações por turma ou por aluno tornam-se indispensáveis.

Finalmente, é preciso se tomar muito cuidado com implementação de mecanismos de incentivos. Em um primeiro momento, prioridades devem ser dadas a formas de incentivos positivos, prêmios para escolas ou regionais com desempenho notável. Embora algumas redes já contem com sistemas de incentivos baseados em remuneração variável atrelada ao desempenho dos alunos, do ponto de vista do estímulo de maior uso das avaliações pelos professores, é preferível dar prioridade a usos pedagógicos da avaliação. Apenas em sistemas maduros, em que todos os envolvidos entendam e estejam certos dos benefícios de avaliações externas, devem ser implementadas iniciativas dessa natureza.

3. Diagnóstico da situação atual e do progresso educacional: como no caso da prestação de contas e dos incentivos, também a realização de diagnósticos educacionais depende de informações sobre os avanços e os retrocessos dos indicadores educacionais. A necessidade de desagregação das informações, no entanto, vai depender se o diagnóstico é nacional, estadual, municipal ou em nível de escola. Mas, ao contrário dos usos anteriores, no caso dos diagnósticos a informação necessária precisa ir muito além do acompanhamento dos resultados educacionais. Como um diagnóstico, em geral, visa identificar e mensurar as causas e as consequências de eventuais ou potenciais fracassos, eles também requerem informações sobre potenciais causas e consequências dos resultados educacionais.

No que diz respeito a fatores determinantes (causas), o diagnóstico busca relacionar os resultados educacionais alcançados a possíveis fatores determinantes e, portanto, informações sobre esses fatores são também imprescindíveis. Esses fatores combinam informações sobre: *i)* os mais diversos aspectos da escola (como infraestrutura, gestão, clima e funcionamento); *ii)* aspectos ligados à formação, envolvimento e dedicação dos professores e clima dentro de sala de aula, *iii)* grau de vulnerabilidade socioeconômica e ambiente familiar em que as crianças se inserem, aí incluído a preocupação, atenção e apoio dos pais com as atividades escolares de seus filhos, e *iv)* características demográficas dos estudantes (idade, gênero, etc.), relacionadas as suas atitudes e preferências frente ao aprendizado (como preferem aprender, que tipo de ambiente escolar preferem, etc.) e a seus déficits de desenvolvimento e aprendizado, além das dificuldades correntes com o que está sendo oferecido ou ministrado pelos professores.

Já com relação à magnitude das consequências do sucesso ou fracasso educacionais, os diagnósticos precisam relacionar o desempenho educacional com os resultados que devem influenciar. Por esse motivo, precisam de informação não apenas do desempenho educacional, mas também dos resultados que devem influenciar, desde os mais imediatos, como o acesso aos níveis educacionais subsequentes, aos mais distantes como a formação de famílias, amizades e capital social, empregabilidade, remuneração, condições de saúde, longevidade e participação comunitária e pública, entre outros.

Investigar a relação entre indicadores envolve relacionar a variação em um indicador à variação no outro, isto é, o quanto a variação em um acompanha a variação no outro. Além disso, para fazermos relações entre indicadores, é preciso que eles estejam na unidade desejada. Se quisermos relacionar o desempenho dos alunos no ensino fundamental com a conclusão do ensino médio, por exemplo, precisamos dos resultados de desempenho por aluno e da informação sobre se esse mesmo aluno concluiu ou não o ensino médio. Assim, para abrir as possibilidades de exploração das informações coletadas, é desejável que cada indicador seja coletado no nível mais desagregado possível. Se existirem informações nos menores níveis de desagregação possíveis (aluno, professor, etc.), teremos as melhores chances de produzir bons diagnósticos, afinal, agregar é sempre mais fácil, mas desagregar nem sempre é possível.

4. Fatores de sucesso e alocação de recursos: a identificação dos casos de sucesso e também daqueles que não vêm apresentando um bom desempenho e, por isso, precisam de maior apoio, requer informação sobre o desempenho de cada unidade cujo sucesso e fracasso se queira analisar. Assim, se o objetivo é identificar sistemas educacionais com melhor ou pior desempenho, as informações precisam ser desagregadas apenas em nível de sistema educacional (seja ele estadual ou municipal). Entretanto, se o objetivo é aprender com o desempenho das regionais de maior sucesso, então, a informação precisa estar desagregada por regional. Já quando o intuito é identificar melhores práticas escolares, é preciso ter informação desagregada por escola. Se a finalidade é aprender com a experiência dos melhores professores e oferecer apoio àqueles com pior desempenho, então, a informação deve estar desagregada por professor. Por fim, se a finalidade é identificar as melhores práticas pedagógicas contrastando o desenvolvimento e o aprendizado de alunos, então, a informação sobre desempenho deve estar desagregada por aluno.

Como, nesse caso, o uso se baseia inteiramente na identificação de unidades com melhor e pior desempenho, é fundamental que os indicadores utilizados na avaliação sejam estritamente comparáveis. Em alguns casos, o desempenho precisa ser condicionado às condições locais e a missão passa a ser identificar, dentre as unidades sujeitas às mesmas condições (igualmente vulneráveis), aquelas com melhor e pior desempenho.

Nesse contexto, além da importância da comparabilidade, é necessário também contar com medidas fidedignas das condições locais de vulnerabilidade. Quando o objetivo é identificar melhores práticas e as unidades que apresentam fraco desempenho, com vistas a prestar-lhes apoio, é fundamental identificar e documentar que aspectos do seu funcionamento podem ser os responsáveis pelo sucesso ou pelo fraco desempenho. Em suma, quando o objetivo da avaliação é a identificação de melhores práticas e unidades com fraco desempenho, é indispensável contar com indicadores que possuem elevado grau de comparabilidade, medidas de contexto acuradas e informações sobe o funcionamento das unidades que permitam compreender porque umas têm sucesso enquanto outras, no mesmo contexto, têm desempenhos muito piores.

5. Identificação de áreas em que novas ações e programas educacionais precisam ser desenhadas: o direcionamento do desenho de novas ações depende do conhecimento das causas dos principais desafios educacionais para as quais não existem ações voltadas à sua superação. Dessa forma, requer o conhecimento e o mapeamento dos desafios educacionais e a identificação de suas causas, sendo essas atividades o objeto do uso diagnóstico das avaliações de resultados educacionais. Ademais, a identificação das causas para as quais novas ações precisam ser desenhadas depende também de informações sobre as ações já desenhadas e implementadas, que tipo de causa essas ações procuram superar e qual o seu grau de eficácia. Assim, a identificação de áreas em que novas ações educacionais precisam ser desenhadas também necessita, além de bons diagnósticos, de informações detalhadas sobre as ações existentes que incluam seu modelo lógico de mudança[3] (e que permitem identificar sobre que causas atuam), além de estimativas sobre a sua eficácia. Em suma, esse uso da avaliação requer um diagnóstico que identifique adequadamente os desafios educacionais e disponibilize um mapeamento das ações existentes, tanto em termos da natureza de cada uma delas (que causas atacam) como da eficácia com que são capazes de influir sobre as causas que focam sua atenção.

6. Avaliações *ex-ante* e *ex-post* da eficácia de ações, práticas e programas educacionais: a eficácia de uma ação educacional é, via de regra, obtida comparando-se os resultados alcançados por unidades beneficiadas com o de unidades não beneficiadas comparáveis (equivalentes). As unidades beneficiadas e não beneficiadas devem ser comparáveis no sentido de que, na ausência da ação, teriam o mesmo desempenho[4]. Ou seja,

[3] "O modelo lógico busca configurar um desenho do funcionamento do programa, que seja factível em certas circunstâncias, para resolver os problemas identificados. Pode ser a base para um convincente relato do desempenho esperado, ressaltando onde está o problema objeto do programa e como este se qualifica para enfrentá-lo" (Ipea, 2010). Adicionalmente, segundo ENAP (2015), o "modelo lógico tem como função primordial explicitar a teoria de funcionamento por detrás de uma ação pública, apresentando de forma esquemática a transformação social que os gestores públicos pretendem atingir e o caminho que precisa ser trilhado".

[4] O impacto de uma ação é exatamente a diferença entre a situação de unidades beneficiadas com a ação (por exemplo, escolas) e a situação que essas mesmas unidades teriam na ausência da ação. A ideia é que, se

se há comparabilidade, então, os resultados observados para as unidades não beneficiadas podem servir como "modelo" do que aconteceria caso as unidades beneficiadas não tivessem recebido a ação.

Para que as avaliações da eficácia de ações sejam realizadas, é fundamental que avaliações dos resultados educacionais estejam disponíveis para um vasto universo de unidades (sistemas educacionais, regionais, escolas ou turmas). Essas avaliações devem ser capazes de oferecer estimativas estritamente comparáveis entre as unidades avaliadas, caso contrário, a não comparabilidade irá se confundir com o impacto da ação levando a medidas enviesadas da eficácia da ação[5].

É importante também que as medidas de desempenho sejam as mais precisas possíveis, uma vez que toda imprecisão irá se traduzir em imprecisões na mensuração da magnitude do impacto das ações e, consequentemente, em um poder de detecção[6] mais limitado da eficácia das ações. De fato, quanto mais imprecisas as medidas de resultados obtidas nas avaliações, maior deve ser o impacto para que ele possa ser detectado, ou seja, não é possível identificar a eficácia de ações pouco eficazes.

Além da precisão, é também importante que as informações sobre os resultados educacionais estejam disponíveis para os momentos e nos anos/séries que para os quais se deseja avaliar a eficácia de uma ação. Como a eficácia de ações educacionais tipicamente devem ser avaliadas ao final do ano letivo, é fundamental que a avaliação do desempenho educacional esteja disponível para todo final de ano letivo (ao invés de cada dois anos, por exemplo). Como as ações educacionais podem ser específicas para anos/séries e disciplinas específicas, se o objetivo é apoiar avaliações de impacto, então, a mensuração dos resultados educacionais deveria ocorrer ao final de todo ano letivo, em todos os anos/séries, e cobrindo todas as disciplinas curriculares.

É preciso ressaltar que a unidade de intervenção de uma ação educacional pode ser tanto o conjunto de um sistema, as regionais de um sistema, escolas em um dado grupo, ou turmas de um conjunto de escolas, ou mesmo estudantes de um grupo selecionado de turmas. Assim, para que um sistema de avaliação seja útil à avaliação da eficácia de ações educacionais

---

conseguirmos gerar um grupo de unidades comparáveis (equivalentes), então, na ausência da ação, as unidades beneficiadas teriam o mesmo resultado alcançado pelas unidades que não foram beneficiadas.

[5] Quando não há comparabilidade os resultados observados para as unidades beneficiadas não servem de "modelo" para as unidades não beneficiadas. Daí a importância da comparabilidade.

[6] Numa avaliação, o poder de detecção se refere à capacidade de verificação do impacto de uma ação sobre um determinado resultado com confiança estatística e resulta do tamanho da amostra utilizada e do quanto o resultado em questão varia na população.

ele deve se adequar às unidades de intervenção dessas ações. Como essas unidades podem ser as mais variadas possíveis e agregar medidas de desempenho é sempre possível, o ideal para o apoio às avaliações de impacto seria poder contar com sistemas de avaliação que mensurem o desempenho de todos os estudantes em todos os anos/séries com relação a todas as disciplinas curriculares.

Por fim, como as avaliações de impacto inevitavelmente se baseiam na comparação de unidades beneficiadas e não beneficiadas, é indispensável que o sistema de informação contenha também dados sobre que unidades foram beneficiadas pela ação (e em que extensão) e quais não foram atendidas. Além disso, para que o impacto de uma ação seja efetivamente isolado não basta comparar o desempenho de unidades beneficiadas com o de unidades não beneficiadas; é também essencial que se garanta a comparabilidade entre essas unidades, conforme dissemos anteriormente. Ou seja, na ausência da ação, unidades beneficiadas e não beneficiadas teriam o mesmo desempenho. Seja para verificar esse requisito de comparabilidade entre unidades *ex-ante* selecionadas para serem beneficiadas e não beneficiadas, ou para a construção *ex-post* de um grupo de unidades não beneficiadas que sirvam de comparação, é fundamental contar com um amplo leque de informações sobre as características das unidades de intervenção da ação e seu contexto (por exemplo, características da escola e do grau de vulnerabilidade das famílias dos estudantes).

7. Planejamento das ações, práticas e programas educacionais: o planejamento parte da identificação dos desafios educacionais a serem enfrentados e da identificação de quais são os seus principais determinantes (causas), que é uma das funções do uso diagnóstico das avaliações. Em seguida, o planejamento utiliza informações advindas da identificação e do modelo lógico de mudança das práticas existentes, das novas ações desenhadas e de quais dentre as existentes são as mais eficazes para selecionar como os recursos disponíveis deveriam ser mais bem alocados frente aos desafios a serem enfrentados. Ademais, o planejamento deve, também, utilizar a informação proveniente da identificação das unidades com pior desempenho ou mais vulneráveis, que precisam atenção redobrada ou atendimento melhor adequado as suas especificidades. Por fim, a partir dos desafios educacionais identificados, do leque de ações educacionais disponíveis e sua eficácia, e das unidades vulneráveis para atendimento prioritário, o planejamento educacional é capaz de redefinir rumos e metas. Assim, em resumo, o planejamento educacional usa as avaliações de resultados de forma indireta na medida em que se baseia no uso diagnóstico das avaliações, no seu uso para mensurar a eficácia das ações educacionais, e para identificar as unidades mais vulneráveis que precisam receber atendimento prioritário e de que tipo.

8. Adequação das práticas pedagógicas: conforme já ressaltado, uma avaliação de resultados educacionais realmente promove o desenvolvimento e o aprendizado dos alunos

quando efetivamente modifica as práticas pedagógicas em sala de aula. Esse uso da avaliação voltado para aprimorar o clima e a eficácia das atividades dentro da sala de aula, denominado de uso pedagógico, requer informações rápidas e bem desagregadas em nível de turma ou aluno. Como o desenvolvimento dos alunos é muito dinâmico, para que sejam de utilidade prática em sala de aula, as avaliações dos alunos precisam ser processadas com muita celeridade, o que consiste em um crônico problema para toda avaliação externa.

A avaliação precisa também permitir que resultados representativos, ao menos em nível de turma, sejam disponibilizados; se resultados para cada aluno estiverem disponíveis a utilidade pedagógica da avaliação deverá ser ainda mais significativa. Ademais, para que tenha utilidade pedagógica, as avaliações devem permitir que análises desagregadas por item/descritor[7] curricular possam ser realizadas, caso contrário, seu uso para adequar e aprimorar as práticas pedagógicas em cada turma fica extremamente limitado. Para que esse uso pedagógico da avaliação beneficie a todos os estudantes, ela precisaria ser feita para todos os anos/séries, e deveria cobrir todas as disciplinas individualmente, ou ao menos todas as áreas do conhecimento. Do ponto de vista da periodicidade, o ideal seria poder contar com várias avaliações ao longo do ano letivo, de tal forma que os déficits de desenvolvimento e de aprendizado de cada aluno pudessem ser identificados e acompanhados, permitindo mais rapidamente a implementação de ações corretivas adequadas.

Em suma, o uso pedagógico das avaliações em sala de aula é o mais demandante em termos de desagregação e periodicidade, uma vez que idealmente requer que o desempenho dos alunos de cada turma deva não só ser avaliado em vários momentos ao longo do ano letivo como, também, em todos os anos/séries e disciplinas curriculares, além de permitirem análises pedagógicas detalhadas ao nível de descritores ou itens curriculares. Existe, no entanto, um aspecto que é vital para a maioria dos usos da avaliação, mas que para seu uso pedagógico em sala de aula é de menor importância: trata-se da comparabilidade dos resultados entre escolas, sistemas educacionais e contextos socioeconômicos. De fato, para o uso pedagógico da avaliação, o que importa é saber com precisão os conteúdos curriculares que o aluno domina e aqueles que ele ainda precisa desenvolver. Do ponto de vista da sala de aula, para que o professor saiba o que seu aluno precisa desenvolver, não é muito útil saber que conteúdos os alunos da escola vizinha dominam.

[7] "O descritor é uma associação entre conteúdos curriculares e operações mentais desenvolvidas pelo aluno, que traduzem certas competências e habilidades" (PDE : Plano de Desenvolvimento da Educação : SAEB : ensino médio : matrizes de referência, tópicos e descritores; 2008, p. 18).

4.2. CONDIÇÕES NECESSÁRIAS PARA UM USO EFICAZ

Para que uma avaliação dos resultados educacionais seja eficaz, é necessário que certas condições estejam presentes. Essas condições dizem respeito tanto à disponibilidade de recursos humanos apropriados, como às abordagens do planejamento educacional, à focalização dos recursos e à gestão escolar. Cada uso da avaliação tem uma demanda específica de condições, para que seja eficaz. Na sequência, identificamos as condições necessárias para que cada um dos oito usos possíveis para a avaliação seja eficaz na promoção do desenvolvimento e do aprendizado dos alunos.

1. Prestação de contas e estabelecimento de metas: o uso das avaliações de resultados educacionais para prestação de contas e verificação de metas tende a ter maior utilidade e eficácia na promoção do desenvolvimento e do aprendizado dos alunos num contexto em que já existam metas bem estabelecidas, de amplo conhecimento da sociedade, e onde a sociedade já tem a cultura e a experiência de acompanhá-las.

A atenção da sociedade ao cumprimento das metas, contudo, não se traduzirá em maior desenvolvimento e aprendizado dos alunos, caso os responsáveis por seu cumprimento não estejam bem definidos, isto é, caso a governança do sistema educacional não esteja bem desenhada. Apenas quando existem responsáveis bem definidos pelo cumprimento das metas, o seu acompanhamento torna-se eficaz. Ademais, importa para a eficácia da prestação de contas e do acompanhamento de metas que os responsáveis pela gestão do sistema educacional e, consequentemente, por seu cumprimento, tenham a devida autonomia para tomar decisões que, de fato, influenciem o desempenho educacional e, assim, possam efetivamente influenciar o cumprimento das metas. Por gestores educacionais incluem-se os gestores da escola e da sala de aula (os professores). Assim, o processo de estabelecimento de metas e sua verificação deve envolver todos esses agentes.

Por fim, é indispensável para a eficácia da prestação de contas e o acompanhamento das metas que os gestores responsáveis por seu cumprimento tenham tradição na realização de sua gestão com base em evidência e que toda evidência disponível esteja sempre à sua disposição de forma amigável, no momento adequado.

2. Sistema de incentivos: para que o uso da avaliação dos resultados educacionais seja eficaz em subsidiar o funcionamento de um sistema de incentivos à promoção do desenvolvimento e do aprendizado dos alunos, é necessário, em primeiro lugar, que o desenho desse sistema seja adequado. Em particular, é indispensável que os agentes que estão sendo incentivados tenham autonomia, informação necessária e tradição em tomar decisões com base

em evidência, permitindo, assim, que os incentivos levem a que melhores decisões sejam tomadas. Portanto, o impacto de um sistema de incentivos sobre o desenvolvimento e o aprendizado dos alunos será tão mais bem-sucedido quanto melhor for o conhecimento dos agentes que o sistema busca incentivar sobre a forma como os resultados educacionais são alcançados, e que influência os fatores e ações sob o seu controle têm sobre esses resultados. Em suma, a eficácia do uso da avaliação em apoio a um sistema de incentivos depende da qualidade do desenho desse sistema e também da autonomia, experiência e conhecimento dos agentes que o sistema visa incentivar com relação às decisões que de fato podem influenciar o desempenho dos alunos.

Como mencionado anteriormente, é preciso extremo cuidado com implementação de mecanismos de incentivos. Em um primeiro momento, prioridades devem ser dadas a formas de incentivos positivos, prêmios para escolas ou regionais com desempenho notável. Apenas em sistemas maduros, em que todos os envolvidos entendam e estejam certos dos benefícios de avaliações externas, devem ser implementadas iniciativas baseadas em remuneração variável atrelada ao desempenho dos alunos.

3. Diagnóstico da situação atual e do progresso educacional: a construção de diagnósticos da situação educacional não é uma tarefa trivial. Ela requer equipes com conhecimentos específicos, que demandam formação de longa duração, em particular, para a identificação e a mensuração das causas e consequências dos sucessos e fracassos observados.

A construção de bons diagnósticos requer também, além de informações de qualidade (envolvendo tanto resultados educacionais, quanto informações sobre seus determinantes), recursos humanos com conhecimento específico para a realização dessa atividade. No entanto, um bom diagnóstico da situação educacional é apenas um primeiro passo para a melhoria do desempenho dos alunos. A real eficácia do diagnóstico irá depender também da cultura de gestão existente. Quanto mais os gestores utilizarem esses diagnósticos para identificarem os desafios a serem priorizados e planejarem suas ações, maior será a importância do uso diagnóstico das avaliações para a melhoria do desempenho dos alunos. Em outras palavras, quanto mais acostumados os gestores estiverem com uma gestão baseada em evidência, mais eficaz será o uso diagnóstico das avaliações. Embora este uso dependa da qualificação e da prática da equipe específica, na medida em que boa parte da gestão da educação é feita na escola e na sala de aula, é extremamente desejável que os processos de diagnóstico envolvam ativamente esses agentes.

4. Fatores de sucesso e alocação de recursos: a identificação dos casos de sucesso e daqueles com pior desempenho é tarefa relativamente simples, embora a facilidade e a frequência de sua realização dependam, em alguma medida, da experiência da equipe gestora do sistema educacional. Tanto a documentação dos casos de sucesso e identificação dos seus

fatores determinantes quanto a identificação dos fatores responsáveis pelo fraco desempenho de certas unidades, dos recursos e da assistência de que necessitam para melhorar seu desempenho requerem equipes experientes e com extensa formação voltada para a execução dessas tarefas.

A identificação e documentação das melhores práticas, no entanto, é apenas um primeiro passo na sua difusão. Para que as melhores práticas sejam efetivamente difundidas, adaptadas e adotadas pelas unidades com pior desempenho, é necessário que o sistema educacional tenha tradição na sua difusão, que existam incentivos para que as melhores práticas sejam adotadas e que as unidades com pior desempenho recebam a necessária assistência técnica para que possam conhecer essas práticas e melhor adaptá-las a seu contexto específico.

Em suma, o uso das avaliações com vistas a documentar e difundir as melhores práticas e adaptá-las às necessidades das unidades mais vulneráveis demanda recursos humanos com formação e experiência adequadas e um bom sistema de incentivos e de apoio para que essas práticas sejam difundidas e adotadas. Além disso, na medida em que boa parte da gestão da educação é feita na escola e na sala de aula, é extremamente desejável que os processos de diagnóstico envolvam ativamente esses agentes. Em outras palavras, a comunidade escolar deve ter papel ativo na identificação dos fatores de sucesso e na escolha de novas práticas.

5. Identificação de áreas em que novas ações e programas educacionais precisam ser desenhadas: para ser efetivo na promoção do desenvolvimento e do aprendizado dos alunos, esse uso da avaliação deve envolver: *i)* a identificação da necessidade de novas ações, *ii)* o desenho de novas ações, e *iii)* a implementação das novas ações.

Para que essas três formas de atuação ocorram de forma eficaz, é necessário que o sistema conte com: *i)* um bom diagnóstico da situação e do progresso educacional e, portanto, que os desafios educacionais a serem enfrentados estejam bem definidos, e *ii)* um inventário abrangente que inclua o conhecimento do modelo de mudança e do grau de eficácia das ações em curso, e também daquelas que estão em andamento em outros sistemas e que poderiam ser adaptadas ao contexto local se necessário.

Assim, caso o sistema educacional disponha de recursos humanos com a formação adequada, frente a um bom diagnóstico e um bom inventário das ações existentes, ele será capaz de identificar novas ações que precisam ser criadas e, então, desenhar essas novas ações com eficácia. Caso existam incentivos e o apoio necessário para a adoção de novas práticas, elas deverão ser implantadas, tornando o uso da avaliação para o desenho de novas ações eficaz na

promoção do desenvolvimento e do aprendizado dos alunos. Para garantir as melhores condições de efetividade de novas iniciativas, é desejável que a comunidade escolar seja também envolvida nas diferentes etapas do processo.

6. Avaliações *ex-ante* e *ex-post* da eficácia de ações, práticas e programas educacionais: para que a mensuração sistemática de resultados educacionais seja eficaz na promoção do desenvolvimento e do aprendizado dos alunos via sua contribuição para a realização de avaliações de impacto, é necessário que avaliações de impacto sejam realizadas e que os seus resultados sejam utilizados para aprimorar a política educacional adotada. Mais especificamente, em primeiro lugar, é imperativo que exista uma cultura voltada para a demanda, realização e uso dessas avaliações. A cultura, no entanto, serve apenas como um primeiro passo. É preciso, também, contarmos com os recursos materiais necessários para a condução dessas avaliações, e com recursos humanos capazes de formular a demanda pelas avaliações, conduzi-las tecnicamente, e compreender e utilizar os resultados encontrados.

Por fim, vale ressaltar que tanto o redesenho das ações já existentes como o desenho de novas ações com base nos resultados das avaliações de impacto dependem não apenas de uma cultura voltada para o desenho e operação de políticas públicas com base em evidência, mas, também, da existência de incentivos eficazes, morais ou materiais, para que a evidência disponível seja efetivamente incorporada ao desenho das ações educacionais. Dado que a maior parte das ações implementadas chega nos níveis da escola e da sala de aula, avaliações de impacto *ex-ante* ou *ex-post* serão tão mais completas se o processo de avaliação envolver a escuta ativa da comunidade escolar.

7. Planejamento das ações, práticas e programas educacionais: identificar e dimensionar os desafios educacionais a serem enfrentados e ter um conhecimento detalhado das ações disponíveis, de seu modelo de mudança e do seu grau de eficácia são elementos fundamentais para o desenho de uma política educacional efetivamente capaz de promover o desenvolvimento e o aprendizado dos alunos. É, porém, a articulação desses conhecimentos que realmente constrói uma política educacional eficaz. Não basta conhecer quais são os desafios prioritários e quais as ações disponíveis; é necessário servir-se de toda evidência disponível para selecionar criteriosamente as ações mais eficazes para tratar dos principais desafios que precisam ser enfrentados.

Assim, para que a mensuração dos resultados educacionais tenha real impacto sobre o desenvolvimento e o aprendizado dos alunos, é necessário que o sistema tenha arraigada cultura de planejamento, com efetivo uso de evidência, e disponha de recursos humanos com a formação, os incentivos e o apoio necessários para realizar um planejamento com plena utilização de toda a informação disponível, adequando os desafios identificados pelo uso

diagnóstico das avaliações com o conhecimento das ações, de seus modelos de mudança e da eficácia de cada uma.

Como em casos anteriores, dado que a maior parte das ações implementadas chega nos níveis da escola e da sala de aula, o aproveitamento da avaliação no processo de planejamento será tão mais efetivo se envolver a escuta ativa da comunidade escolar.

8. Adequação das práticas pedagógicas: para que o uso pedagógico das avaliações seja efetivo na promoção do desenvolvimento e do aprendizado dos alunos, é imprescindível que os educadores, e os professores em particular, tenham formação e experiência que lhes permita interpretar com facilidade e de forma adequada os relatórios pedagógicos das avaliações. Caso contrário, por mais perfeita e bem estruturada que sejam as informações extraídas das avaliações, elas acabarão não sendo utilizadas, nem pelos responsáveis pelo desenho dos cursos de formação de professores, nem pelos próprios professores em sala de aula.

Além da formação e da experiência necessárias à leitura dos relatórios, é fundamental que os educadores percebam o valor dessa informação para o aprimoramento da formação dos professores e de suas práticas em sala de aula. É preciso também que tenham os devidos incentivos e o devido apoio para utilizar essa informação em sua plenitude. Os incentivos e a percepção de utilidade elevam o interesse dos educadores pelas informações pedagógicas da avaliação e os motivam para utilizá-las. O seu uso efetivo, no entanto, irá também depender da forma como essas informações são apresentadas. Quanto mais amigável, estimulante e de fácil compreensão e manuseio, maior a chance dessas informações serem efetivamente utilizadas da forma como se pretendia que fossem.

Em suma, para que o uso pedagógico das avaliações seja pleno, é indispensável que os educadores tenham a formação, a experiência, o apoio e os incentivos necessários e que os relatórios de apresentação dos resultados dessas avaliações sejam amigáveis e estimulantes.

CAPÍTULO 5

ARQUITETURA DO SISTEMA DE AVALIAÇÃO DA EDUCAÇÃO BÁSICA

ARQUITETURA DO SISTEMA DE AVALIAÇÃO DA EDUCAÇÃO BÁSICA

INTRODUÇÃO

A construção de um sistema de avaliação requer, em primeiro lugar, a definição de sua arquitetura. De fato, quem avalia necessariamente avalia alguém em algum momento. Assim, todo sistema de avaliação requer que se defina ***quando*** a avaliação irá acontecer e ***quem*** será avaliado e. A arquitetura de um sistema de avaliação é, portanto, uma resposta combinada a essas duas questões: ***quem*** será avaliado e ***quando***.

A abordagem a cada uma dessas questões pode naturalmente ser organizada numa sequência de subquestões. Para definir ***quando*** a avaliação deverá acontecer, é necessário que se determine, por um lado, a periodicidade com que ela deve ocorrer. A periodicidade deve ser bimestral, anual, bianual (como faz o INEP) ou a cada três anos (como faz o PISA)? Por outro lado, em particular quando a periodicidade é anual, é preciso definir o momento da avaliação. A avaliação deve ocorrer no início ou final do ano letivo?

Para definir ***quem*** será avaliado, é preciso definir quais as redes a serem incluídas na avaliação. Todas as redes, inclusive as escolas privadas, devem ser avaliadas ou somente as escolas públicas? O sistema de avaliação estadual deve incluir todas as redes municipais e também as escolas federais? Uma vez definidas as redes, resta saber quais escolas devem participar da avaliação. As principais opções seriam: todas, apenas aquelas com um tamanho mínimo ou uma amostra representativa. Definido o universo de escolas, é preciso, então, definir quais anos/séries devem ser avaliados em cada um dos ciclos educacionais. Todas as séries devem ser incluídas? Ou apenas as séries terminais dos ciclos educacionais (3º, 5º e 9º anos do Ensino Fundamental e 3ª série do Ensino Médio)? Ou será que as séries de transição (6º ano do Ensino Fundamental e 1ª série do Ensino Médio) também deveriam ser incluídas? Após a definição das escolas e os anos/séries a serem avaliados, é preciso determinar se todas as turmas dos anos/séries selecionados nas escolas escolhidas devem participar da avaliação ou se a participação deve se limitar a uma amostra de turmas. Uma vez determinadas quais as turmas selecionadas para avaliação, resta ainda determinar se todos os alunos da turma devem ser avaliados em todos os conteúdos, ou se a avaliação deve se limitar ao conhecimento de uma parcela dos alunos. Por fim, é necessário determinar em que medida alunos com alguma deficiência que requeira adaptação da avaliação também farão parte da mesma.

O objetivo deste capítulo é apresentar as possibilidades existentes para a arquitetura de um sistema de avaliação estadual, ressaltando as vantagens e desvantagens de cada uma dessas possibilidades. Iniciaremos a análise considerando as possibilidades para ***quando*** avaliar e, em seguida, passaremos a uma análise das alternativas para ***quem*** participa da avaliação.

Antes de iniciarmos a análise dessas alternativas, é importante ressaltar que a arquitetura mais adequada depende diretamente do uso que se pretende dar à avaliação. A cada uso particular da avaliação corresponde uma arquitetura específica para o seu desenho. Optamos por tratar da arquitetura mais adequada para uma avaliação com múltiplos usos, isto é, quando a avaliação deve servir tanto para a prestação de contas, como para aprimorar e avaliar a gestão, estimar o impacto de ações específicas, ajustar e aprimorar as práticas pedagógicas.

2. QUANDO AVALIAR?

Conforme já ressaltado, a definição do ***quando*** avaliar tem duas dimensões. Primeiramente, é necessário definir a periodicidade da avaliação. O Programa Internacional de Avaliação de Estudantes – PISA avalia a cada *três* anos o desempenho dos jovens de 15 anos de idade em Leitura, Matemática e Ciências[1]. O Sistema de Avaliação da Educação Básica (SAEB), no caso do Ensino Médio, e a Prova Brasil, realizam avaliações bianuais (nos anos ímpares) em Matemática e Língua Portuguesa. Já a Avaliação Nacional de Alfabetização (ANA) investiga a alfabetização e o letramento em Língua Portuguesa (leitura e escrita) e Matemática, com periodicidade ainda não plenamente estabelecida. Dessas avaliações, mesmo as realizadas por amostragem (PISA e SAEB) certamente permitem o acompanhamento do desempenho educacional do País e de cada unidade da federação. No caso das avaliações realizadas censitariamente (Prova Brasil e ANA), os resultados obtidos permitem acompanhar e prestar contas do desempenho em nível de escola.

É evidente, porém, que a periodicidade não anual dessas avaliações (à exceção da ANA) limita o seu uso para fins pedagógicos. Dada a natureza anual dos ciclos educacionais, para que uma avaliação possa ser utilizada com essa finalidade de forma sistemática, é necessário que sua periodicidade seja ao menos anual. Só assim é possível diagnosticar, ao início de cada ano letivo, a proficiência dos alunos e identificar seus déficits de aprendizado. Com base nessas informações, as práticas em sala de aula e o currículo dos cursos de formação, além de outros aspectos da política de formação continuada, podem ser ajustados. Mas, para que essas

[1] Seis avaliações foram realizadas entre 2000 a 2015: 2000, 03, 06, 09, 12 e 15.

informações sejam úteis do ponto de vista pedagógico, elas precisam ser processadas e analisadas em no máximo algumas poucas semanas após o início do ano letivo. Nesse sentido, tão importante quanto definir a periodicidade da avaliação é definir o momento em que ela ocorre e em que seus resultados são apresentados.

Além da celeridade com que as avaliações precisam ser analisadas para terem utilidade pedagógica, vale ressaltar que também importa a desagregação com que as informações podem ser disponibilizadas. De fato, uma medida agregada da proficiência numa dada disciplina para o conjunto dos alunos numa dada unidade da federação serve apenas como instrumento de prestação de contas e para apoiar algumas decisões gerenciais. Para que uma avaliação tenha utilidade pedagógica, é necessário que também ofereça informações desagregadas por tema e, na medida do possível, também por descritor da Matriz de Referência[2]. Se essa informação está disponível apenas em nível agregado para o conjunto dos alunos na unidade da federação, a avaliação será capaz de influenciar apenas a política pedagógica dessa unidade da federação. Por isso é importante que a avaliação seja capaz de oferecer resultados ao menos em nível de escola, uma vez que apenas dessa forma irá reconhecer a heterogeneidade entre elas e, assim, poderá influenciar as práticas em sala de aula e as necessidades de formação específicas a cada escola.

Portanto, para que uma avaliação possa ser utilizada de forma a cobrir todos os seus potenciais usos, em particular o pedagógico, ela precisa ocorrer no mínimo anualmente, e seus resultados devem ser disponibilizados antes ou no início do ano letivo. Por vezes, um importante uso das avaliações é a averiguação do impacto de determinados programas, projetos ou ações. Embora essas intervenções ocorram tipicamente ao longo de todo o ano letivo, seu término pode acontecer em anos pares ou ímpares. Para que o impacto das intervenções possa ser isolado, é imprescindível contar com avaliações anuais. Por esses motivos, avaliação com periodicidade anual é a norma[3].

Pode ocorrer, no entanto, alguma variação na periodicidade a depender do conteúdo a ser avaliado. Embora a avaliação de algumas disciplinas ou áreas do conhecimento em determinados anos/séries devam necessariamente ser realizadas de forma anual, outras disciplinas ou áreas do conhecimento ou outros anos/séries podem ser avaliadas de maneira mais esporádica. Nesse caso, o uso pedagógico mais sistemático ficaria restrito às disciplinas e anos/séries em que a avaliação é anual. Já em relação às disciplinas e anos/séries avaliados com

[2] "...a Matriz de Referência é o referencial curricular do que será avaliado em cada disciplina e série, informando as competências e habilidades esperadas dos alunos" (PDE : Plano de Desenvolvimento da Educação : SAEB : ensino médio : matrizes de referência, tópicos e descritores; 2008, p. 17).

[3] Todas as 21 unidades da federação que realizaram avaliação em 2016 (apenas 5 estados e o Distrito Federal – Roraima, Amapá, Sergipe, Bahia, Santa Catarina – não realizaram avaliação nesse ano) têm a intenção de realizá-la ao menos anualmente.

menor periodicidade o uso da avaliação estaria mais voltado para definições e ajustes nos rumos mais agregados. Ainda assim, é possível fazer uso pedagógico da avaliação como, por exemplo, na definição de conteúdo de programas de formação dos educadores com base em déficits de desenvolvimento dos estudantes, na definição de metodologias e material didático para uso em sala de aula. No caso dessas avaliações feitas com menor periodicidade, referentes a processos de ciclos maiores, a celeridade na análise e divulgação dos resultados é uma questão menos importante. O sistema de algumas unidades da federação, a exemplo do sistema de avaliação educacional chileno (*Sistema de Medición de la Calidad de la Educación* (Simce)), adota essa estratégia, em que se combinam algumas disciplinas e anos/séries com avaliação anual, com outras disciplinas e anos/séries avaliadas em menor periodicidade[4].

Pode parecer surpreendente que as unidades da federação planejem avaliações anuais quando existem avaliações federais bianuais. Em princípio, caso o escopo e abrangência das avaliações federais e estaduais fossem alinhados, bastaria que as estaduais ocorressem bianualmente, nos anos pares, para complementar as avaliações federais. No entanto, diferenças no conteúdo e na cobertura, assim como na velocidade com que os resultados das avaliações federais são liberados, têm impedido que as evidentes sinergias entre as avaliações estaduais e federais sejam exploradas. Por conseguinte, as unidades da federação que utilizam algum sistema de avaliação terminam por adotar uma periodicidade anual, a despeito da existência de um sistema federal bianual em paralelo. Assim, é imprescindível que os sistemas de avaliação estaduais e o sistema federal sejam compatibilizados, evitando a sobreposição entre os dois nos anos ímpares. Caso essa compatibilização não ocorra a tendência é de duplicação de esforços e gastos, sobrecarregando, em particular, os estudantes que passam a realizar duas avaliações nos anos ímpares. Importante ressaltar que, na medida em que o sistema federal é padronizado, produzindo resultados comparáveis entre as unidades da federação, a compatibilização dos sistemas estaduais com o federal implica necessariamente a compatibilização entre os próprios sistemas estaduais.

No entanto, para que o uso pedagógico seja plenamente atendido e ajustes de percurso ao longo do ano letivo possam ser realizados, o ideal seria poder contar também com avaliações de cunho formativo[5], com periodicidade bimestral ou trimestral. Essas avaliações formativas ainda são raras no cenário nacional. Poucas unidades da federação têm alguma experiência com esse tipo de avaliação, com apenas duas tendo realizado alguma avaliação desse tipo ao longo

[4] Os representantes das unidades da federação designados para compor o Grupo de Trabalho de Avaliação do Consed manifestaram preferência pela avaliação de todas as disciplinas com a mesma peridicidade.

[5] "Avaliação conduzida ao longo do processo educacional com o objetivo de potencializar a aprendizagem do estudante. Implica buscar evidências sobre a aprendizagem a fim de fechar o hiato entre desempenho atual e o desejado (de modo a permitir ações para fechar esse hiato); oferecer *feedback* aos estudantes; e envolver estudantes no processo de avaliação e aprendizagem (Fonte: CCSSO, 2008)". IBE Glossário de Terminologia Curricular. UNESCO, 2016.

de 2016 – Espírito Santo (trimestral) e Tocantins (semestral). Como toda avaliação envolve gastos e demanda significativo esforço, tanto dos alunos quanto de toda a equipe da secretaria de educação envolvida, esta só deve ser realizada quando existem condições para que os seus resultados sejam disponibilizados com a celeridade necessária e apresentados de forma adequada, uma vez que apenas nessas circunstâncias uma utilização plena dos resultados é possível de ser alcançada.

A frequência com que avaliações padronizadas e externas devem ser realizadas também deve levar em consideração a complementaridade com as avaliações realizadas pelos professores. Os professores têm formação específica em avaliação e são os responsáveis últimos pela avaliação do grau de proficiência de seus alunos. Eles são, afinal, os membros da comunidade escolar que melhor conhecem os alunos e, portanto, devem ter seu trabalho de avaliação complementado e enriquecido, e não substituído. Embora informações externas sobre a situação do aprendizado de seus alunos possam e devam ser sempre de utilidade para cada professor, um professor atento não deverá, em geral, ficar surpreso com o desempenho individual de cada um de seus alunos numa avaliação externa. Por isso, as avaliações externas nunca poderão substituir a avaliação realizada diretamente pelos professores, devendo sempre ser utilizadas de forma complementar.

Em suma, a tendência geral das unidades da federação é por avaliações anuais próprias, independentes das avaliações federais bianuais. Apenas alguns poucos estados ainda acreditam na integração com as avaliações federais e, em vista disso, planejam avaliações bianuais para serem realizadas nos anos pares e, dessa forma, serem complementares às federais. Quanto às avaliações formativas ao longo do ano letivo, poucas são as unidades da federação que têm alguma experiência com esse tipo de avaliação, e poucas têm a intenção de implementá-las no futuro próximo.

Conforme já mencionado, a definição de quando avaliar tem dois componentes: a periodicidade e o momento ao longo do ano letivo em que a avaliação é realizada. No caso de uma avaliação com periodicidade anual, o momento ideal para sua realização é o início do ano letivo ou o seu final. Nesse caso há certa incompatibilidade ente os usos da avaliação. Quando o uso prioritário da avaliação é a prestação de contas, estabelecimento de metas e outros fins relacionados à avaliação do desempenho, é natural que o final do ano letivo seja o momento mais apropriado. Afinal, o que se deseja avaliar é como evoluiu o aprendizado dos alunos ao longo do ano letivo ou de um ciclo educacional. De fato, toda avaliação somativa[6] tipicamente

[6] "Determinação do desempenho do aluno ao final de um período, estágio, curso ou programa, usualmente – mas não necessariamente – envolvendo provas ou exames formais. Avaliações somativas são mais comumente usadas para classificar, dar notas e/ou aprovar estudantes, assim como para propósitos de certificação." IBE Glossário de Terminologia Curricular. UNESCO, 2016.

ocorre ao final do ano letivo. Em praticamente todos os sistemas estaduais em funcionamento, a avaliação ocorre ao final do ano letivo. Em 2016, apenas o estado do Mato Grosso realizou avaliação no início do ano letivo.

Se a função da avaliação é mais de origem pedagógica, então, ela se aproxima muito mais de uma avaliação diagnóstica[7] ou formativa. Nesse caso, o momento natural para sua realização é o início do ano letivo, desde que os resultados sejam analisados e disponibilizados com a devida celeridade, de forma que possam ser utilizados para adequar as necessidades do corpo docente e discente às práticas a serem adotadas e a formação a ser realizada ao longo do ano letivo.

Quando a avaliação é realizada no final do ano letivo, seu uso pedagógico depende da portabilidade dos resultados, isto é, requer que o resultado da avaliação de cada aluno seja registrado e associado a seu número no sistema educacional, permitindo-se conhecer a sua proficiência em cada nova turma no ano escolar seguinte. Essa portabilidade é particularmente desafiadora quando a avaliação é realizada ao final dos ciclos educacionais, momento em que muitos alunos trocam não só de escola como de rede (saindo de uma escola municipal para uma estadual, por exemplo). Assim, para que avaliações realizadas ao final do ano letivo tenham a devida utilidade pedagógica, é fundamental que os alunos de todas as redes possam ser identificados e que os resultados da avaliação estejam devidamente associados a essa identificação. É também essencial que esses resultados sejam analisados com rapidez e que possam ser apresentados já no início do ano letivo subsequente.

A portabilidade das informações é também importante quando a avaliação ocorre no início do ano letivo. Nesse caso ela serve ao propósito inverso, permitindo identificar qual a proficiência que os alunos tinham ao final do ano letivo anterior, tipicamente ao final da série anterior em que foram avaliados. Dessa forma, uma avaliação diagnóstica no início de um dado ano letivo serve como somativa para o ano letivo anterior. Para isso, os alunos avaliados no início de um ano letivo precisam ser "reenturmados" da forma como estavam ao final do ano letivo anterior e, portanto, sua identificação é indispensável. Essa identificação precisa ser única e invariante no tempo, válida para todas as escolas e redes. Para cada aluno avaliado no início de um ano letivo é necessário determinar a turma, a escola e a rede cursada ao final do ano letivo anterior.

[7] "Avaliação que visa a identificar os pontos fortes e fracos de um aluno, com vistas a tomar as ações necessárias para potencializar a aprendizagem. Também usada antes do processo de ensino e aprendizagem, a fim de aferir o nível de prontidão ou de desempenho do aluno." IBE Glossário de Terminologia Curricular. UNESCO, 2016.

3. QUEM DEVE SER AVALIADO?

Conforme já ressaltado a arquitetura de um sistema de avaliação requer a definição tanto de ***quem*** será avaliado quanto de ***quando*** será realizada a avaliação. Na seção anterior tratamos do ***quando***; nesta iremos tratar de ***quem*** deve ser avaliado. A questão de ***quem*** será alvo da avaliação possui uma variedade de dimensões. Em primeiro lugar temos a macroquestão de quais redes devem ser incorporadas à avaliação. Como o direito à educação é de todas as crianças, adolescentes e jovens, independe se a escola que frequentam é pública ou privada, ou se pertence à rede municipal, estadual ou federal.

Ter esse direito assegurado e devidamente monitorado para cada criança, adolescente e jovem em seu território é uma das funções de cada secretaria estadual de educação, na medida em que é a responsável última pela coordenação do sistema educacional em seu território. Assim, é desejável que os sistemas de avaliação estaduais englobem escolas de todas as redes. No sistema federal (SAEB/Prova Brasil), a participação é optativa para as escolas privadas, mas obrigatória para todas as escolas públicas. Os sistemas estaduais enfrentam um triplo desafio quando buscam avaliações que cubram todas as crianças, adolescentes e jovens em seu território, quais sejam: *i)* conseguir a adesão das redes municipais e das escolas privadas; *ii)* realizar a aplicação (logística) dos testes de forma centralizada, garantindo a integridade, sigilo e segurança da aplicação (e, consequentemente, a confiabilidade dos resultados) e dos ganhos de escala envolvidos; e *iii)* equacionar a questão do financiamento da avaliação, em particular, de como obter e operacionalizar o cofinanciamento da avaliação das redes municipais e, principalmente, das escolas privadas participantes[8].

Por esses motivos, a despeito de terem que arcar normalmente com todo o financiamento da avaliação, mais da metade dos sistemas de avaliação estaduais cobrem as escolas municipais e quase todos que não incluem atualmente as redes municipais gostariam que as escolas dessas redes fossem incluídas na avaliação. No entanto, poucos sistemas estaduais incluem hoje escolas privadas em sua avaliação e apenas quatro planejam incluir essas escolas no desenho futuro de seu sistema. Em suma, a depender da disponibilidade de recursos, a tendência geral da arquitetura dos sistemas de avaliação estaduais é a inclusão das escolas públicas; já a inclusão das escolas privadas é uma questão controversa, com alguns

[8] Os representantes das unidades da federação designados para compor o Grupo de Trabalho de Avaliação do Consed manifestaram preferência por mecanismo de financiamento que onere mais diretamente as escolas particulares.

sistemas caminhando na direção da inclusão[9], enquanto outros caminham na direção de um sistema exclusivamente de escolas públicas.

Uma vez definidas as redes que farão parte do sistema de avaliação, é preciso decidir quais escolas devem participar da avaliação. Em geral, dois aspectos dessa escolha se ressaltam. Em primeiro lugar, tem-se a questão do tamanho das escolas. O sistema deve avaliar o desempenho dos alunos, qualquer que seja o tamanho de sua escola, ou por uma questão de eficiência e contenção de custos deve avaliar apenas os alunos em escolas com um tamanho mínimo? Se um tamanho mínimo deve ser estabelecido, qual deve ser esse tamanho mínimo? Em nível federal, e na vasta maioria dos sistemas estaduais atuais, apenas as escolas com mais de 20 alunos matriculados nos anos/séries avaliados participam da avaliação, com exceção da avaliação da alfabetização (ANA) no 3º ano, quando esse mínimo é reduzido para 10 alunos por escola no 3º ano. A tendência geral é manter esses valores mínimos estabelecidos em nível federal, embora exista uma disposição a reduzir o mínimo nas demais séries para 10 alunos por série, qualquer que seja a série avaliada.

Outra questão fundamental é sobre a cobertura ser universal ou apenas para uma amostra representativa. Em apenas duas unidades da federação que realizaram avaliação em 2016 (Tocantins e Mato Grosso) a cobertura não foi universal (todos os alunos). Parece haver consenso de que a cobertura universal é preferível, a despeito de seus custos bem mais elevados. De fato, a partir de uma cobertura universal, é possível *i)* estabelecer e acompanhar metas por escola, *ii)* fazer uso pedagógico da avaliação levando em consideração as especificidades de cada escola, e *iii)* informar e instruir cada comunidade escolar na busca das práticas pedagógicas mais adequadas à situação do corpo discente e das ações de formação mais adequadas às necessidades do corpo docente.

Uma alternativa à cobertura universal que permite que informações sobre o desempenho educacional em nível de escola sejam obtidas é incluir na avaliação todas as escolas, mas dentro de cada escola avaliar apenas uma amostra de turmas ou uma amostra de alunos em cada turma. A estratégia pode também requerer que todos os alunos façam uma parte da avaliação, com uns alunos ou turmas, por exemplo, realizando a avaliação de Língua Portuguesa, enquanto outros alunos ou turmas fazem a avaliação de Matemática. A vantagem desse procedimento é a possibilidade de obter medidas de desempenho por escola com uma redução de custo e um menor desgaste para cada aluno, que segundo esse desenho só será avaliado em uma das várias disciplinas consideradas na avaliação. A desvantagem dessas abordagens por amostragem de turmas ou alunos em relação a uma avaliação universal é uma menor precisão das estimativas do desempenho por escola. Além disso, nos casos em que para

[9] Em 2016, quatro estados (Espírito Santo, São Paulo, Rio Grande do Sul e Goiás) avaliaram escolas privadas em alguma série do Ensino Fundamental ou Médio.

cada aluno na avaliação se mede a proficiência em apenas uma disciplina, é impossível investigar a associação entre a proficiência nas diversas disciplinas.

Uma vez definidas as escolas a serem avaliadas, o próximo passo no desenho da arquitetura da avaliação é definir quais os anos/séries devem ser incluídos. O PISA se limita a avaliar a proficiência de jovens de 15 anos, independentemente da série que frequentam. Já o sistema de avaliação federal, que inclui a ANA, a Prova Brasil e o SAEB, cobre as séries terminais dos principais ciclos educacionais, isto é, o final do ciclo de alfabetização, 3º ano; o final da 1ª etapa do Ensino Fundamental, 5º ano; o final da 2ª etapa do Ensino Fundamental, 9º ano; e o final do Ensino Médio, 3ª série.

A vasta maioria dos sistemas de avaliação estaduais centra sua atenção nesses mesmos anos/séries: 3º, 5º, e 9º ano do Ensino Fundamental e 3ª série do Ensino Médio. Em quatro sistemas estaduais, no entanto, todas as séries do Ensino Médio são incluídas na avaliação – Pará, Piauí, Alagoas e Espírito Santo. Além disso, cerca da metade das unidades da federação pretende estender a avaliação para além desses quatro anos/séries. Vale ressaltar também que existem experiências estaduais com a avaliação da proficiência em Matemática e Língua Portuguesa em todos os anos do Ensino Fundamental e séries do Ensino Médio. A exemplo do que ocorre com o sistema Chileno, é possível avaliar as séries finais de maneira universal, enquanto as séries intermediárias, de maneira amostral; nas avaliações recentes, apenas o estado de SP adota essa prática. As séries intermediárias mais comumente avaliadas pelos sistemas estaduais são a 7º série do Ensino Fundamental e 1ª série do Ensino Médio, onde tipicamente as taxas de reprovação e de abandono são mais elevadas. A essas séries deve-se também incluir o 6º ano do Ensino Fundamental, que, junto com a 1º série do Ensino Médio, representam os anos de transição entre os ciclos e são marcados por baixas taxas de aprovação.

Em suma, existe considerável consenso de que avaliações em todos os anos/séries do Ensino Fundamental e Médio não são necessárias quando o objetivo é a prestação de contas, o diagnóstico e o planejamento educacional, a definição e o acompanhamento de metas, a correção e os ajustes na forma de gestão e a avaliação de impacto de ações, programas e projetos. A avaliação em cada ano/série pode, contudo, ser de grande valia pedagógica. Mas, dada a experiência atual da maioria das redes com o uso pedagógico das avaliações, o volume de informações pertinentes que uma avaliação em cada ano/série pode gerar para este tipo de uso ainda é superior à capacidade de absorção por parte das escolas. Assim, no contexto atual, a avaliação em todos os anos/séries é percebida como um desperdício de recursos e deve ser alcançada por meio de uma expansão gradual dos anos/séries a serem incluídos na avaliação.

Por fim, uma vez definidas as escolas e anos/séries a serem incluídos na avaliação, resta observar que os sistemas de avaliação devem ser inclusivos, requerendo para isso a adaptação

dos meios, instrumentos de coleta e testes e a oferta de profissionais especializados, de forma que alunos deficientes ou pertencentes a comunidades tradicionais (indígenas, quilombolas etc.) possam também participar dessas avaliações.

4. RESUMO E CONCLUSÕES

A arquitetura de um sistema de avaliação envolve determinar quem deve ser avaliado e quando essa avaliação deve ocorrer. Definir a melhor e mais adequada arquitetura que um sistema de avaliação deve ter depende dos usos que se pretenda dar a esse sistema. Neste capítulo descrevemos a arquitetura típica dos sistemas estaduais vigentes e qual deveria ser a sua arquitetura ideal no futuro, considerando que cada sistema deve ser capaz de atender a usos verdadeiramente múltiplos.

Vimos que os resultados das avaliações precisam estar disponíveis ao menos com periodicidade anual. Assim, os sistemas estaduais podem optar ou por terem, eles mesmos, periodicidade anual ou coordenarem esforços com o sistema bianual federal (anos ímpares) e serem também bianuais, complementando o sistema federal nos anos pares. Embora a avaliação possa, em princípio, ser realizada ou no início ou ao final do ano letivo, tanto o sistema federal como os estaduais têm dado preferência pela coleta de informações ao final do ano letivo, e essa deve ser a tendência no futuro. Essa escolha privilegia os usos da avaliação para planejamento, prestação de contas, gestão e avaliação de impacto, enquanto dificulta o uso pedagógico. Quando a avaliação ocorre ao final do ano letivo, o seu uso para fins pedagógicos requer a garantia de portabilidade dos resultados, isto é, a identificação do aluno que participou da avaliação, de maneira que possa ser utilizada como avaliação diagnóstica no início do ano letivo subsequente.

Quanto a quem deve participar da avaliação, o mais comum é incluir tanto as escolas da rede estadual como aquelas da rede municipal e federal. Tradicionalmente a avaliação é feita cobrindo o universo das escolas públicas, com eventual participação das privadas por amostragem ou por adesão, e essa deve ser também a tendência futura. Em cada escola, tipicamente avaliam-se todos os alunos em todas as turmas nas séries selecionadas, que normalmente envolvem as séries finais dos ciclos educacionais (3º, 5º e 9º anos do Ensino Fundamental e 3ª série do Ensino Médio), mas há uma forte demanda para a inclusão também das séries inicias dos ciclos, em função de suas elevadas taxas de abandono e reprovação (6º ano do Ensino Fundamental e 1ª série do Ensino Médio). A importância da inclusão na avaliação

dos alunos da 1ª série do Ensino Médio parece ser uma demanda consensual, a despeito da possibilidade de se utilizar a avaliação ao final do 9º ano do Ensino Fundamental como diagnóstica para o início da 1ª série do Ensino Médio. Por fim, é importante ressaltar a importância de a avaliação ser inclusiva, permitindo a participação de deficientes e de membros de comunidades tradicionais.

CAPÍTULO 6

ECONOMIAS DE ESCALA E OPORTUNIDADES PARA A REDUÇÃO NOS CUSTOS DAS AVALIAÇÕES ESTADUAIS DA EDUCAÇÃO BÁSICA

ECONOMIAS DE ESCALA E OPORTUNIDADES PARA A REDUÇÃO NOS CUSTOS DAS AVALIAÇÕES ESTADUAIS DA EDUCAÇÃO BÁSICA

INTRODUÇÃO

Num país federado como o Brasil, em que cada unidade da federação tem considerável autonomia para conduzir sua própria política educacional, existe uma tendência para que cada uma desenvolva e utilize um sistema de avaliação próprio adequado e especializado às suas necessidades e prioridades específicas. Afinal, não só as prioridades diferem entre unidades da federação, também os currículos exibem diferenças significativas. É natural, portanto, que os sistemas de avaliação estaduais, ao levarem essas dessemelhanças em consideração, acabem também marcados por relevantes diferenças.

Mas, exceto pela maior adequação às especificidades locais, sistemas de avaliação localmente diferenciados apenas acumulam desvantagens. A estadualização dos sistemas de avaliação impede que alguns importantes fatores redutores de custo sejam aproveitados. Dois se destacam: por um lado, a estadualização do sistema reduz o poder de barganha dos governos estaduais no momento em que contratam serviços de avaliação e tende a elevar os custos, na medida em que inviabiliza o aproveitamento de economias de escala; por outro, a adequação das avaliações estaduais às necessidades idiossincráticas de cada estado dificulta a padronização e, consequentemente, a comparabilidade dessas avaliações.

Assim, em princípio, a criação de um consórcio interestadual voltado à realização conjunta de avaliações pode, sem perda de alguma dose de customização estadual, viabilizar reduções de custo, via um maior poder de barganha na contratação dos serviços ou um melhor aproveitamento das economias de escala, além de gerar ganhos de comparabilidade interestadual dos resultados da avaliação decorrentes de uma maior padronização do conteúdo e da metodologia a ser utilizada na avaliação. Ademais dessas duas vantagens, um consórcio deve também facilitar e promover a troca de experiência em fazer, mas, principalmente, em utilizar as avaliações. Por fim, uma estreita coordenação entre unidades da federação poderá facilitar a organização do financiamento, em particular, dando ao Consed melhores condições para negociar um maior volume de recursos federais para a realização de avaliações.

Em suma, essa estreita coordenação das avaliações estaduais pode, sem perda da capacidade de ajustes às especificidades locais, trazer quatro grandes vantagens: *i)* reduzir os custos; *ii)* facilitar o financiamento; *iii)* conferir maior comparabilidade entre os resultados das avaliações; e *iv)* viabilizar maior troca de experiências, não apenas em relação a como fazer, mas, principalmente, em relação a como usar as avaliações. Neste capítulo, centramos a atenção na primeira das quatro vantagens de cooperação entre unidades da federação com relação a avaliações: a redução de custos.

No entanto, para podermos tratar dessa questão, é necessário antes considerarmos uma questão mais ampla: qual a magnitude e qual a estrutura do custo de uma avaliação. Como a importância das economias de escala varia substancialmente em relação ao tipo de custo, o conhecimento sobre a estrutura dos custos é imprescindível para se investigar a magnitude das economias de escala e, então, a possibilidade de reduções de custo via a coordenação e cooperação interestadual. Nesse sentido, desdobramos a questão da redução de custos via economias de escala em quatro questões: *i)* qual a magnitude do custo da avaliação e qual a estrutura desse custo; *ii)* em que componentes existem economias de escala significativas; *iii)* qual o grau dessas economias de escala; e, por conseguinte, *iv)* como o custo da avaliação para cada unidade da federação pode diminuir com a cooperação entre as unidades.

Economias de escala, no entanto, não são a única forma como os custos de uma avaliação podem ser reduzidos. Aspectos da prática na condução dessas avaliações podem também tornar a sua realização mais eficiente e menos custosa. Essa possibilidade de redução de custos via ganhos de eficiência também é tratada ao final do capítulo, onde particular atenção é dada à melhor forma de divisão do trabalho entre as equipes internas às Secretarias e equipes especializadas e contratadas externamente para esse fim.

1. MAGNITUDE DOS CUSTOS

Antes de passarmos a uma análise detalhada dos custos de uma avaliação é importante termos uma ideia da magnitude dos recursos necessários. A magnitude anual dos custos da avaliação pode ser estimada combinando-se informações sobre o número de estudantes a serem avaliados a cada ano com uma estimativa do custo médio da avaliação por aluno por ano.

No Brasil, temos hoje cerca de 36 milhões de alunos matriculados nos Ensinos Fundamental e Médio. Assumindo que apenas os alunos nas séries terminais de cada ciclo serão avaliados (2º ou 3º, 5º e 9º Anos do Ensino Fundamental e 3ª Série do Ensino Médio), chegamos

a uma previsão de que cerca de 1/3 dos alunos devem ser avaliados a cada ano, o que nos leva a uma necessidade de avaliar cerca de 12 milhões de alunos[1].

Nossas estimativas, apresentadas nas seções seguintes, para o custo unitário anual por aluno avaliado variam de R$25 a R$30 (considerando que a avaliação de cada aluno requer apenas um dia letivo). Dado esse custo unitário e o número esperado de alunos a serem avaliados a cada ano, encontramos um custo para o conjunto das avaliações entre R$300 e R$360 milhões por ano.

Trata-se, portanto, de um custo expressivo, levando a que qualquer avaliação mereça toda a atenção, tanto no que se refere à qualidade e à utilidade das informações que oferece, quanto à eficiência com que é conduzida. No entanto, frente ao gasto total com os Ensinos Fundamental e Médio, da ordem de R$200 bilhões por ano[2], o gasto com a avaliação representa uma parcela ínfima, inferior a 0,2% do gasto total, e é certamente um preço adequado a ser pago para verificarmos se todo esse investimento em educação de fato levou a uma efetiva garantia do direito à educação para todos.

2. ECONOMIAS DE ESCALA

A cooperação e coordenação das unidades da federação na condução de suas avaliações devem trazer, conforme já ressaltado, reduções de custo por uma variedade de razões. Além das possibilidades oferecidas pelos ganhos de escala, a cooperação eleva o poder de negociação do conjunto dos estados com os provedores dos serviços necessários em comparação ao que cada estado teria, caso fizesse a negociação em separado. Além disso, a cooperação pode intensificar a troca de experiências e melhores práticas, tornando, assim, o processo de realização das avaliações mais eficiente.

No que se refere às economias de escala[3], essas se apresentam de duas formas bem distintas que vale ressaltar. Por um lado, as etapas de desenvolvimento dos instrumentos a serem utilizados na avaliação, de definição da metodologia a ser usada para o planejamento da sua execução e de desenho da devolutiva são comuns a todas as avaliações. Dessa forma, representam um custo fixo que, no caso de uma estreita cooperação entre estados, precisam ser incorridos uma única vez e, em boa medida, independem da escala (número de alunos a serem avaliados) e da periodicidade com que as avaliações devem ser realizadas (anuais ou

[1] Fonte: Censo Escolar/INEP, 2016.

[2] Fonte: Indicadores Financeiros Educacionais/INEP, 2014.

[3] Dizemos que há economias de escala na produção quando o aumento no número de unidades produzidas diminui o custo unitário de produção.

bianuais). Por outro lado, existem os custos que dependem da escala, mas que crescem de forma não linear, menos que linearmente. Ou seja, há tipos de custos em que o acréscimo de unidades adicionais de avaliação não aumenta o custo na mesma proporção, mas numa proporção inferior[4], de forma que o custo unitário se torna decrescente com a escala. Nessa situação, a estreita cooperação entre unidades da federação gera economias de escala para avaliação que reduzem o custo unitário. São exemplos disso os eventuais ganhos que se possa ter na impressão dos instrumentos a serem utilizados na avaliação e no processamento das informações, além dos ganhos na impressão e na publicação dos resultados a serem utilizados nas devolutivas.

3. FONTE DE INFORMAÇÕES

Conforme já ressaltado, para investigarmos a capacidade de redução de custos proporcionada por uma cooperação mais estreita entre unidades da federação no desenho e na condução das avaliações, precisamos analisar a estrutura dos custos da avaliação e as possibilidades de economias de escala em cada um deles. Existem duas opções polares para a condução dessa análise. A primeira opção consiste em conduzir a análise com base numa planilha teórica de custos obtida a partir de um planejamento detalhado de uma avaliação. A segunda consiste em realizar tal análise com base na experiência concreta das unidades da federação que vêm realizando essas avaliações.

Neste capítulo, optamos por essa segunda abordagem. Para isso, obtivemos informações detalhadas, para 15 dos 18 estados que tiveram experiência com a realização de avaliações em escala nos últimos dois anos, sobre o custo e a estrutura de custos e também sobre o número de alunos avaliados. Em 12 desses 15 estados para os quais temos informações – ou seja, em 80% dos casos – o custo unitário ficou entre R$20 e R$35 por aluno por ano, a despeito de todas as diferenças no desenho das avaliações, nas estratégias de implantação e no tamanho e na densidade populacional desses estados. Em duas unidades da federação, o custo unitário ficou bem abaixo de R$20, indicando, possivelmente, que alguns componentes do custo não tenham sido devidamente computados.

[4] Uma boa maneira de ilustrar as diferenças entre os dois casos de economia de escala é imaginar uma situação onde dobramos a escala da avaliação, isto é, dobramos o número de alunos avaliados. No caso de economias de escala por existência de custos fixos, dobrar a escala não acarreta qualquer incremento de custo e o custo total permanece igual ao original (antes de dobrarmos a escala da avaliação). No caso de economias de escala advindas de um crescimento não linear de custos, quando dobramos a escala, o custo total aumenta, mas não dobra (menos que dobra).

4. COMPONENTES E ESTRUTURA DO CUSTO

Para efeito dessa análise, conforme apresentado no Quadro 1, dividimos os custos da avaliação em cinco componentes. O primeiro componente trata do **desenvolvimento dos instrumentos a serem utilizados na avaliação e do planejamento de toda a avaliação**, em particular, de todo o trabalho de campo. Estima-se, com base na experiência das 15 unidades da federação para as quais temos informação sobre o custo da avaliação, que esse componente fica em torno de R$716 mil, o que representa 11% do total do custo médio da avaliação (cerca de R$2,24 por aluno avaliado por ano). Esse é um custo fixo que independe da escala e pode ser quase que perfeitamente compartilhado por todos as unidades da federação, levando a que, se compartilhado, possa representar, em vez de 11%, apenas 0,4% do custo total previsto da avaliação em cada estado. Sem dúvida, esse é o componente onde maiores ganhos podem ser derivados de uma estreita cooperação entre os estados.

O segundo componente trata de um conjunto de atividades que depende sobremaneira da escala da avaliação (número de alunos a serem avaliados). São atividades que envolvem a formação de pessoal e a impressão e distribuição dos instrumentos a serem utilizados na avaliação, ou seja, de **atividades preparatórias para a avaliação**. Nesse caso, o custo depende da escala e encontramos fortes evidências de economias de escala. O custo médio dessa fase da avaliação é próximo a R$8 por aluno avaliado por ano, representando 32% do custo da avaliação. Esse custo, no entanto, é bem sensível à escala, podendo variar de R$6 por aluno, quando cerca de um milhão de alunos serão avaliados, a mais de R$11 por aluno, quando apenas 100 mil serão avaliados. Esse é um componente que poderia, em princípio, trazer economias de custo se uma estreita cooperação entre estados fosse alcançada. Caso as avaliações de todas as unidades da federação fossem processadas e analisadas em conjunto, o custo da avaliação desse componente poderia ser reduzido a quase 60% de seu valor atual. Dessa forma, esse componente, que representa 32% do custo, poderia ser reduzido a 13%. Na prática, porém, as vantagens dessa cooperação podem ser difíceis, porque dependem de uma perfeita coordenação dos tempos em que as avaliações são realizadas e de uma elevada padronização dos instrumentos de avaliação e a utilização das mesmas normas.

O terceiro componente do custo da avaliação se refere à **realização da avaliação** propriamente, isto é, à aplicação dos instrumentos, ao recolhimento desse material e ao envio para o processamento. Os custos nessa etapa dependem sobremaneira da escala da avaliação (número de alunos a serem avaliados) e não existem muitas oportunidades para economias de escala nessas atividades. Estimamos que esse componente deva custar cerca de R$8 por aluno avaliado por ano, representando 33% do custo total da avaliação por estado médio. A evidência que examinamos corrobora que não existe possibilidade de economia de escala nesse componente; o custo cresce *pari passu* ao número de alunos avaliados.

O quarto componente trata do **processamento dos dados e da análise estatística dos resultados**. Nesse caso, o custo também depende da escala, porém com fortes evidências de economias de escala. O custo médio dessa fase da avaliação é R$3,7 por aluno avaliado por ano, representando 15% do custo da avaliação. Assim como o custo do segundo componente, esse custo também possui sensibilidade à escala, variando de R$2,8 por aluno, quando cerca de 1 milhão de alunos serão avaliados, a R$5,2 por aluno, quando apenas 100 mil serão avaliados. Sendo assim, esse é um componente que poderia, em princípio, trazer economias de custo caso uma estreita cooperação entre estados fosse alcançada. O processamento e a análise das avaliações de todas as unidades da federação poderiam ter seu custo reduzido a, talvez, metade de seu valor atual, caso as avaliações estaduais fossem realizadas em conjunto. Dessa forma, esse componente, que representa 15% do custo, poderia ser reduzido a 7%. No entanto, ressalta-se novamente a dificuldade de se aproveitar essas vantagens, uma vez que é necessária uma perfeita coordenação dos tempos em que as avaliações são realizadas e uma elevada padronização dos instrumentos de avaliação e a utilização das mesmas normas. Ao final, um processamento único das informações exige que todas as unidades envolvidas na colaboração adotem exatamente os mesmos instrumentos (portanto a mesma matriz de referência com os mesmos descritores) e unifiquem a forma de correção dos testes.

O quinto e último componente do custo de uma avaliação é formado por **análise, impressão e disseminação dos resultados da avaliação nas escolas**. Esse componente é, assim como o terceiro, extremamente sensível à escala da avaliação (número de alunos avaliados). Como no caso do terceiro componente, aqui também as possibilidades de economias de escala são mínimas. A evidência indica que, independentemente da escala, o custo médio dessa etapa é R$2,5 por aluno por ano. Essa etapa representa, portanto, 10% do custo total da avaliação e não se constitui em uma etapa que deva contribuir muito para a redução de custos, mesmo na presença de estreita cooperação entre as unidades da federação na avaliação.

Em resumo, o custo da avaliação pode ser decomposto em cinco componentes. Dois desses componentes (terceiro e quinto) dependem de forma quase linear da escala da avaliação e, portanto, não dão origem a qualquer economia de escala significativa. Como, em conjunto, esses componentes representam 43% do custo da avaliação, fica evidente que o uso das economias de escala para a redução de custo pode ser considerável. Dos três componentes restantes, o primeiro representa 11% do custo total. É um custo fixo que poderia ser perfeitamente compartilhado entre os estados, possibilitando, dessa forma, uma redução de custos para todos os estados de 10% do total. Já o segundo componente, como representa 32% do custo total, permite um maior ganho de escala. Se a avaliação for realizada em cooperação entre as unidades da federação, pode-se alcançar a partir desse componente uma redução de quase 20% do custo total. No entanto, ressalta-se a difícil coordenação dos tempos de avaliação, que requer nesse caso elevado grau de sincronismo.

Quanto ao quarto componente, que representa 15% do custo total, há possibilidade de algum ganho de escala e redução de custo. No entanto, também requer, para que essa redução

de custo seja alcançada, um alto grau de sincronismo e padronização dos instrumentos e normas para sua correção. De qualquer forma, se plenamente explorado, esse componente permitiria que alcançássemos uma redução de 8% no custo total.

No conjunto, estimamos que uma estreita colaboração entre unidades da federação poderia reduzir os custos da avaliação em 37%. Assim, em vez de um custo agregado de quase R$190 milhões de um sistema não cooperado, poderíamos, na situação de um sistema de cooperação, passar a um custo de R$115 milhões, ou seja, uma economia de R$75 milhões por ano.

Quadro 1: Custos estimados das avaliações estaduais

Estado	Custo total estimado	Custo estimado por grupo de atividades					Número de alunos avaliados	Custo estimado por aluno
		Componente 1: Desenvolvimento e Planejamento	Componente 2: Organização da Avaliação	Componente 3: Aplicação da Avaliação	Componente 4: Processamento de Dados e Análise Estatística	Componente 5: Análise Pedagógica e Disseminação dos Resultados		
RR	**1.548.557**	715.777	406.025	189.192	179.400	58.163	23.360	66
AP	**1.894.884**	715.777	552.286	291.256	246.026	89.540	35.962	53
TO	**2.629.863**	715.777	843.864	527.796	380.168	162.259	65.168	40
RO	**2.736.764**	715.777	884.657	563.921	399.045	173.365	69.628	39
SE	**3.136.711**	715.777	1.034.374	702.173	468.520	215.867	86.698	36
AL	**3.412.008**	715.777	1.135.063	799.870	515.396	245.902	98.761	35
MA	**3.439.988**	715.777	1.145.199	809.905	520.121	248.987	100.000	34
DF	**3.484.350**	715.777	1.161.233	825.853	527.598	253.890	101.969	34
AC	**3.487.881**	715.777	1.162.508	827.124	528.192	254.280	102.126	34
RN	**4.005.775**	715.777	1.346.665	1.016.557	614.258	312.517	125.516	32
MS	**4.059.086**	715.777	1.365.330	1.036.371	622.999	318.609	127.962	32
PI	**4.153.774**	715.777	1.398.356	1.071.698	638.474	329.469	132.324	31
PB	**4.423.480**	715.777	1.491.585	1.173.226	682.209	360.682	144.860	31
AM	**4.769.260**	715.777	1.609.409	1.305.227	737.586	401.262	161.158	30
MT	**4.812.521**	715.777	1.624.023	1.321.878	744.462	406.381	163.214	29
GO	**4.899.645**	715.777	1.653.372	1.355.502	758.276	416.718	167.365	29
SC	**5.981.974**	715.777	2.009.655	1.782.187	926.463	547.893	220.049	27
ES	**6.447.452**	715.777	2.158.675	1.970.241	997.053	605.706	243.268	27
PR	**8.301.996**	715.777	2.732.075	2.741.516	1.269.811	842.816	338.498	25
RS	**8.510.512**	715.777	2.794.776	2.830.156	1.299.736	870.067	349.443	24
PE	**10.387.380**	715.777	3.345.780	3.642.564	1.563.435	1.119.823	449.752	23
CE	**11.154.724**	715.777	3.564.883	3.981.437	1.668.626	1.224.002	491.593	23
RJ	**12.022.405**	715.777	3.808.853	4.368.747	1.785.957	1.343.072	539.415	22
BA	**12.130.679**	715.777	3.839.031	4.417.367	1.800.485	1.358.019	545.418	22
PA	**12.327.893**	715.777	3.893.854	4.506.085	1.826.884	1.385.293	556.372	22
MG	**14.908.276**	715.777	4.595.268	5.684.267	2.165.466	1.747.498	701.844	21
SP	**22.920.314**	715.777	6.628.298	9.501.264	3.154.028	2.920.947	1.173.133	20

Notas: 1) Os dados disponíveis para Tocantins referem-se apenas aos componentes 2 e 3.
2) Os dados disponíveis para o Espírito Santo não contemplavam o componente 5.
3) As Unidades Federativas com o número de alunos destacados em vermelho corresponde a uma estimativa.

5. ESTIMANDO O CUSTO UNITÁRIO E TOTAL E SUA DISTRIBUIÇÃO ENTRE UNIDADES DA FEDERAÇÃO

Na seção anterior, tratamos de como uma estreita colaboração entre unidades da federação pode reduzir custos por meio do aproveitamento das economias de escala. Para isso, identificamos as possibilidades de economias de escala, investigando a estrutura do custo de uma avaliação e procurando desvendar quais componentes são caracterizados por custos fixos e quais possuem custo unitário decrescente com a escala. Em seguida, apuramos qual seria o custo agregado da avaliação, bem como o custo unitário médio por aluno avaliado, na presença dessas economias de escala e contrastamos com o custo observado na sua ausência.

Nesta seção, tomamos uma abordagem complementar que pretende estimar qual seria o custo da avaliação em cada unidade da federação. O desafio nesse caso é estimar como diferenças em escala, mas também na geografia e densidade populacional, influenciam o custo por aluno avaliado por ano.

O resultado apresentando na seção anterior aponta para uma situação em que, no agregado – e uma vez considerados todos os componentes do custo, existe evidência da existência de considerável economia de escala. Existe, ainda, evidência de que o custo por aluno avaliado por ano tenda a ser maior para os estados onde a densidade populacional é menor. Assim, enquanto o custo unitário médio da avaliação é estimado como sendo da ordem de R$25 por aluno avaliado por ano, esse custo pode ser inferior a R$20 por aluno avaliado por ano em estados com alta densidade populacional, como o Estado de São Paulo, enquanto que esse mesmo custo pode passar de R$30 em unidades da federação com baixa densidade, como o Estado do Acre.

Supondo que 1/3 dos alunos matriculados nos Ensinos Fundamental e Médio devam ser avaliados a cada ano, teríamos uma população de 12 milhões de alunos a serem avaliados. Estima-se um custo unitário de R$22, isso levaria a que a avaliação tivesse um custo total de R$265 milhões por ano. Uma vez que unidades da federação com um maior volume de alunos tendem a ter maior densidade populacional e, dessa forma, um menor custo unitário, o custo total de R$265 milhões encontra-se distribuído entre os estados de forma um pouco menos que proporcional ao número de alunos a serem avaliados. Com isso, São Paulo teria 20% dos alunos, mas apenas 16% do custo, enquanto o Pará teria 5% dos alunos e 5% do custo.

A estimativa do número de alunos avaliados usada acima, 12 milhões, considera todos os matriculados, sejam aqueles matriculados na rede federal, na estadual, nas municipais ou em escolas particulares. Evidentemente o total de matrículas nas redes estaduais é bem menor. Apenas 40% das matrículas nos Ensinos Fundamental e Médio são nas redes estaduais, assim, o custo total estimado referente à avaliação de alunos nas redes estaduais é próximo de R$106

milhões, com os restantes R$159 milhões representando o custo da avaliação de alunos matriculados nas redes municipais e nas escolas federais e particulares. Desses R$159 milhões, como esperado, estima-se que R$116 milhões decorra da avaliação dos alunos matriculados nas redes municipais. Os restantes R$43 milhões são necessários para a avaliação dos alunos matriculados nas escolas federais e, principalmente, nas escolas particulares.

Em última instância, a missão dos governos estaduais e distrital é a garantia do pleno direito à educação de todas as crianças, adolescentes e jovens residentes em seus estados, independentemente da rede de ensino a que a escola pertença. Assim, é parte de suas atribuições avaliar e monitorar a efetiva garantia desse direito a todas as crianças, adolescentes e jovens residentes e não apenas daquelas matriculadas em escolas da rede estadual. É evidente, no entanto, que essa responsabilidade é compartilhada com toda a sociedade: famílias, Conselhos Estaduais e Municipais de Educação e Tutelares, Ministério Públicos, governos municipais e federal, além das escolas particulares, que devem encontrar na garantia do direito à educação o ponto de partida para suas atividades.

Por ser uma responsabilidade compartilhada, não deveria ser necessariamente obrigação única dos governos estaduais e distritais o financiamento das avaliações e, por conseguinte, o monitoramento da garantia do direto à educação para todos. Como os demais gastos com educação, o financiamento da avaliação deveria, também e naturalmente, ser compartilhado entre os diversos entes federados e também com as escolas particulares. Em princípio, o governo federal deveria ter recursos reservados não só para coordenar ou apoiar a coordenação do sistema de avaliações locais, como também para apoiar tecnicamente e financeiramente as avaliações locais. Em cada unidade da federação, o governo estadual deveria coordenar as ações de avaliação e organizar o seu financiamento com a participação e apoio técnico e financeiro dos governos municipais. Na medida em que as escolas particulares participam do processo de avaliação, seria desejável contar com seu apoio técnico e financeiro nesse processo.

6. AÇÕES DIRIGIDAS PARA A REDUÇÃO DOS CUSTOS DA AVALIAÇÃO

Como vimos, uma estreita cooperação entre os estados no desenho e na implementação da avaliação pode levar a reduções nos custos em cerca de 40%. Essa forma de cooperação, no entanto, não é a única oportunidade para que as unidades da federação contribuam para a redução dos custos de um sistema próprio de avaliação. A cooperação entre elas pode ser importante para a redução dos custos, mesmo quando cada uma tem seu próprio sistema de avaliação com desenho e implementação próprios. Uma forma como essa cooperação pode ser vital é via a troca de melhores práticas: tal como o Brasil pode aprender muito, reduzir custos e

elevar a eficácia de seu sistema de avaliação, via intercâmbio técnico com países que têm experiências exitosas, como o Chile, cada unidade da federação pode aperfeiçoar seu sistema próprio de avaliação, reduzindo custos e elevando a efetividade, via a troca de experiências com as demais.

Embora a troca de experiências entre unidades da federação possa ocorrer nas mais variadas dimensões, talvez seja na forma de condução da avaliação que os estados e o Distrito Federal tenham mais o que compartilhar. A seguir, serão detalhados três aspectos da condução de avaliações locais que têm grande interferência sobre o custo e a efetividade da avaliação. Nos três casos, a experiência concreta dos estados mereceria documentação e intercâmbio.

O primeiro aspecto é a divisão de tarefas e funções entre o que idealmente deveria ser realizado pela equipe local de avaliação e o que deveria ser contratado externamente. Entre outras questões, seria importante determinar, a partir das experiências existentes, em que medida uma participação maior das equipes locais aumenta ou reduz o custo da avaliação e em que medida essa participação de fato aumenta a eficácia e melhora a assimilação dos resultados da avaliação pelas comunidades escolares. Vale também determinar não apenas qual a extensão ideal da participação interna e externa, mas também qual a melhor divisão de funções entre as duas equipes. Em particular, seria importante corroborar a percepção geral de que as devolutivas e a disseminação de resultados, garantida a necessária disponibilidade de corpo técnico, deveria ser uma atribuição da equipe interna da Secretaria. Já a elaboração dos instrumentos de avaliação e sua análise estatística, por serem atividades altamente especializadas, deveria ser uma atribuição de uma equipe externa.

Em segundo lugar, seria também importante determinar, com base na experiência existente, qual a ideal composição da equipe interna, tanto em termos de quantidade como de formação dessa equipe e qual deveria ser o papel de parceiros externos na formação e na transferência progressiva de tecnologia à equipe interna das Secretarias.

Por fim, é importante resgatar, das experiências das unidades da federação, em que medida um maior envolvimento da equipe interna da Secretaria com a realização avaliação é saudável. Afinal, esse maior envolvimento pode levar, por um lado, a uma perda de foco em outras atividades mais finalísticas da Secretaria, e por outro, a uma avaliação menos independente e isenta. Vale, porém, reforçar que, no que se refere ao uso da avaliação, o recomendado é que não apenas as secretarias, como toda a rede, tenham um maior envolvimento, incluindo professores e alunos. O capítulo sobre uso das avaliações aborda esse ponto.

CAPÍTULO 7

RESUMO DAS REUNIÕES TÉCNICAS DO GRUPO DE TRABALHO – AVALIAÇÃO

RESUMO DAS REUNIÕES TÉCNICAS DO GRUPO DE TRABALHO – AVALIAÇÃO

1. SÍNTESE DAS DISCUSSÕES

Do início de 2016 a meados de 2017, representantes das Secretarias de Educação de todas as 27 unidades federativas do Brasil, sob a coordenação do Consed, participaram de reuniões presenciais e virtuais para a elaboração conjunta de uma proposta para as avaliações estaduais da educação básica no País. Compondo juntamente com os parceiros institucionais envolvidos, o Grupo de Trabalho – Avaliação, os representantes estaduais puderam contribuir com o conhecimento e a experiência acumulados em suas unidades da federação sobre o tema. Ao longo de quatro reuniões técnicas realizadas em Brasília, num total de 5 dias de trabalho conjunto, foram realizadas discussões temáticas e apresentações de especialistas em diversos aspectos das avaliações educacionais, de acordo com as decisões e priorizações estabelecidas pelos membros do grupo de trabalho. A seguir, as principais discussões realizadas nas quatro reuniões técnicas serão sintetizadas.

Custo de implementar um sistema de avaliação. Os custos, identificados como o principal desafio para implementar um sistema de avaliação, incluem custos financeiros e custos políticos. O custo financeiro é conhecido pelos estados no que se refere aos serviços para a realização das provas. O custo político envolve situações que já acontecem em diversas secretarias, como a descontinuidade de processos avaliativos. Além disso, a rotatividade da equipe, muitas vezes em decorrência das mudanças na gestão, também impacta na implementação de um sistema, já que inviabilizam a composição de uma equipe e aprofundamento de sua formação para trabalhar com os resultados.

Frequência das avaliações. Quanto à frequência da avaliação estadual, observou-se que, se for muito espaçada, perde-se o tempo de intervenção. Por outro lado, uma periodicidade muito curta faz com que não haja tempo para a assimilação de um grande volume de dados, nem tempo para se observar mudanças. Sugeriu-se pensar em uma periodicidade mista: algumas avaliações com periodicidade anual e outras menos frequentes, como no caso do Chile. Foi ressaltado que, para se definir a periodicidade, tem-se que definir primeiramente o uso, já que para certos usos é necessária uma curta periodicidade, enquanto para outros, uma periodicidade mais longa é desejável. A grande quantidade de dados gerados por cada avaliação exige tempo hábil para a elaboração de indicadores e para a apropriação desses por parte da

rede de ensino. Outra preocupação foi referente à existência de várias avaliações durante o ano letivo, o que traz risco de sobrecarregar o aluno e tirar o papel fundamental do professor na avaliação formativa.

Cobertura de anos e séries. Sobre quais anos e séries devem ser avaliados, se somente séries ou anos críticos (como anos de passagem com altas taxas de reprovação, ou anos iniciais e finais de um ciclo), ressaltou-se a atenção devida ao Ensino Médio, sem que isso gere descuido com a avaliação dos anos iniciais. Sobre os anos em que não são realizadas avaliações pelo SAEB, houve entendimento que esses apresentam oportunidade para a construção de uma complementariedade entre as avaliações nacionais e as estaduais. Foi mencionado que a avaliações de todos os anos pode significar um custo incompatível com os usos possíveis, sendo preferível uma otimização de recursos. Em relação à prática de espaçar avaliações por quatro séries, como ocorre comumente na atualidade, argumentou-se que implica em dificuldades para um efetivo uso pedagógico, por isso a sugestão de incluir, além do 3º e 5º ano do Ensino Fundamental, a avaliação anual no 7º ano do ensino fundamental e, para o Ensino Médio, o primeiro ano.

Disciplinas e áreas de conhecimento na avaliação dos alunos. Houve consenso de que a avaliação de estudante deve incluir aspectos cognitivos, não-cognitivos e competências de todas as quatro áreas de conhecimento. As opções preferíveis foram a inclusão de todas as disciplinas ou de todas as quatro áreas de conhecimento, que incluem ciências naturais e humanas, matemática e linguagens (sendo incluído, nessa área, português com avaliação a leitura e não redação). Foi recomendado que, dentre as competências acadêmicas dos alunos, todas as áreas (com focos variados a cada ano – a exemplo do PISA) presentes no currículo deveriam ser incluídas, medidas em escalas nacionalmente definidas a partir da BNCC.

Habilidades socioemocionais nas avaliações de alunos. Foi acordado entre os membros do grupo, que, num cenário ideal, seriam incluídas habilidades socioemocionais, atitudes, interesses, engajamento dos alunos nas atividades escolares; a exposição às atividades escolares; o progresso e retenção escolar; a condição socioeconômica da família; as competências socioemocionais bem como o interesse dos pais nas atividades escolares dos seus filhos. Os dados de contexto considerados relevantes para inclusão na avaliação foram características étnicas e sócio demográficas, rural/urbano, entorno socioeconômico/ demográfico da escola. A inter-relação entre competências socioemocionais e a percepção dos alunos sobre o clima e o ambiente escolar foi discutida, já que podem ser confundidas em avaliações. Foi destacado que as competências socioemocionais podem ser avaliadas de formas indiretas, como feito pelos estudos PISA.

Cobertura de escolas e turmas. Foi recomendada que a cobertura de escolas e turmas seja universal, respeitando o mínimo 10 alunos no ano/série a ser avaliado. Sobre a participação de escolas municipais, federais e privadas, recomenda-se a cobertura censitária para as escolas públicas (sejam elas municipais, estaduais ou federais) e amostral para escolas privadas. O estado deve contemplar toda a logística, pois a outra possibilidade, de concessão da prova para aplicação de terceiros, gera perda de confiabilidade do sigilo e dos resultados.

Momento das avaliações. Sobre se a avaliação deve ser aplicada no início ou no final do ano, ponderou-se que é preciso saber em que momento queremos que os indicadores produzam resultados. A preferência recaiu sobre o final do ano letivo, em parte porque aa avaliação ao longo do ano letivo é de aprendizagem, fora do escopo do sistema de avaliação externa.

Definições para os alunos com deficiência. Recomendou-se a adaptação de provas e oferta de profissionais especializados para alunos com deficiência no Ensino Regular. Foi discutido pelo grupo uma classificação existente para as avaliações nacionais sobre quem pode ser atendido, que atualmente são cegos e surdos, dentro de um leque em que é possível atender por adaptação de meios.

Definições para as escolas rurais, indígenas, quilombolas e multisseriadas. Para escolas rurais, indígenas, quilombolas, recomenda-se cobertura censitária mas por adesão, ou seja, sem obrigatoriedade. As escolas multisseriadas não participariam da avaliação, segundo a recomendação do grupo.

Avaliações de profissionais de educação. Foram apontadas as dificuldades referentes à avaliação dos diferentes profissionais e níveis de gestão escolar e ao monitoramento da formação dos professores. Destacou-se que não se pode separar o saber docente das condições de trabalho dos docentes. Para além da comunidade escolar, os funcionários da secretaria de educação foram mencionados como possível alvo de avaliação, devido à falta de mecanismos para acompanhamento da qualificação e empenho desses profissionais. Argumentou-se que essas avaliações poderiam contribuir para a formação e para o aperfeiçoamento dos profissionais.

Avaliação institucional. Houve consenso que a avaliação institucional deve ser qualitativa, sobre o funcionamento dos equipamentos, que já aspectos quantitativos são levantados no censo educacional. Os aspectos relacionados ao clima escolar seriam incluídos nessa avalia ção.

Foi ressaltado pelos membros do grupo a necessidade de não reduzir o escopo da avaliação institucional, que precisa cobrir uma variedade de temas e atores para cumprir sua função. Mencionou-se também a importância de integrar a avaliação institucional com o planejamento e a definição de metas. A avaliação institucional deve ser necessariamente um processo de auto avaliação da escola, com protagonismo de seus atores. Foi recomendada a inclusão de vários níveis de análise para a avaliação institucional, já que essa só faria sentido se for feita como auto avaliação relevante para toda a comunidade escolar, abrangendo todos os temas sobre os quais os diferentes atores podem prover informações, o que contribuiria para que essa mesma comunidade se apropriasse dos resultados posteriormente.

Fontes de recursos para financiar as avaliações. Foram discutidas fontes de financiamento públicas e privadas, com a sugestão de que o governo federal deveria ter uma rubrica de avaliação para financiar e repassar recursos, de forma vinculada, para o Sistema Nacional de Avaliação, através de um fundo federal de avaliação. Há possibilidade também de que parceiros privados, como bancos, destinassem recursos para o Sistema Nacional de Avaliação, investindo parte do lucro como investimento social. Adicionalmente, a contribuição das escolas particulares para que as avaliações sejam realizadas em suas redes foi vista como desejável e possível.

Ações para reduzir os custos da avaliação. Uma maior participação da equipe estadual reduz os custos significativamente, mas é necessário ter uma equipe qualificada e empoderada internamente. Em relação às empresas contratadas pelos estados para a realização das avaliações, foi proposto analisar cada contrato estadual individualmente, com objetivo de identificar o que os contratos têm em comum.

Consórcio entre estados. A ideia de um consórcio entre os estados, para o grupo, poderia aproveitar sinergias e economias de escala e escopo para reduzir custos. Um exemplo seria a elaboração de itens, que poderiam ser trocados e compartilhados entre os estados. Outros aspectos em que a criação de um consórcio seria benéfica seria na contratação de gráficas, na logística, na formação dos técnicos e no treinamento de aplicadores.

2. DETALHAMENTO DAS DISCUSSÕES

2.1. 1ª REUNIÃO TÉCNICA DO GRUPO DE TRABALHO DE AVALIAÇÃO DO CONSED, EM 22 DE JUNHO DE 2016

Conforme divisão definida anteriormente em plenária, os integrantes do GT de Avaliação passaram a compor os seguintes subgrupos temáticos: Conteúdo, Usos e Arquitetura bem como Custos e Financiamento. Em seguida, será resumida a trajetória de discussões de cada um desses subgrupos.

2.1.1. SUBGRUPO DE CONTEÚDO

Agradecemos o intenso e produtivo trabalho dos membros do Subgrupo de Conteúdo, que foram: coordenadora Amanda Midori Amano (Distrito Federal),coordenadora adjunta Jonyele Cheyenne Moura Cruz (Sergipe), Cheila F. B. S. Vasconcelos (Alagoas), Weyne Maria Magalhães Carneiro (Goiás), Luis Alex Lopes Correia (Maranhão), Marta de Medeiros Correia (Paraíba), Hilda Maria Freire Montysuma (Roraima), Salete Doria Albuquerque (Rio Grande do Sul), Cyntia Lemes da Silva G. de Fonseca (São Paulo) e Emerson Azevedo Soares (Tocantins).

No tocante às discussões substantivas transcorridas durante o trabalho de subgrupo, essas foram iniciadas por uma proposta de trabalho apresentada pela facilitadora, Diana Coutinho. Essa proposta incluía a identificação de uma estrutura de aspectos que são ou deveriam ser incluídos no conteúdo das avaliações da educação básica. Após a constatação que nenhum dos estados representados no subgrupo conduz avaliações do ensino profissionalizante, enfocou-se nos âmbitos das avaliações do ensino regular, que incluíam: 1. Conteúdos, competências e habilidades avaliados, representando as áreas de conhecimento (matemática, linguagens, ciências da natureza, ciências humanas) e o escopo avaliado em cada uma delas (relativamente a linguagens, por exemplo, avalia-se, geralmente, português e, muitas vezes, apenas no referente à capacidade de leitura); 2. Contexto (características étnicas e sócio demográficas, rural/urbano, entorno socioeconômico/demográfico da escola); 3. Recursos (infraestrutura, formação de professores e outros profissionais, condições de trabalho) e 4. Gestão escolar e avaliação institucional da secretaria.

Dentro desses âmbitos, algumas áreas cuja inclusão ou não nas avaliações foi amplamente discutida foram as competências socioemocionais, habilidades não-cognitivas, clima escolar e avaliação dos profissionais de educação. A inter-relação entre competências socioemocionais e a percepção dos alunos sobre o clima e o ambiente escolar foi discutida, já

que podem ser confundidas em avaliações. Adicionalmente, uma discussão bastante detalhada foi referente à avaliação dos diferentes profissionais e níveis de gestão escolar. Foram apontadas as dificuldades em monitorar a formação dos professores, e destacou-se que não se pode separar o saber docente das condições de trabalho dos docentes. Nos níveis de coordenador e diretor, foram discutidas dificuldades de avaliação. A experiência de outros países na avaliação docente, em todos os níveis relevantes, foi mencionada como insumo para delinear cenários para o futuro. Para além da comunidade escolar, os funcionários da secretaria de educação foram mencionados como possível alvo de avaliação, devido à falta de mecanismos para acompanhamento da qualificação e empenho desses profissionais. Argumentou-se que essas avaliações poderiam contribuir para a formação e para o aperfeiçoamento dos profissionais.

No que toca à avaliação de desempenho dos alunos nas quatro áreas de conhecimento delimitadas na Base Nacional Comum Curricular (linguagens, matemática, ciências da natureza e ciências humanas), argumentou-se que a avaliação de estudante deve incluir aspectos cognitivos, não-cognitivos e competências de todas as quatro áreas, não apenas português e matemática. Dentro de linguagem, avalia-se atualmente língua portuguesa na maioria dos estados, em especial no âmbito da leitura, mesmo que alguns estados avaliam outros idiomas e produção de texto. Dentro de matemática, avalia-se a solução de problemas.

Dentro dessa discussão, foram abordados também aspectos referentes à utilização, para fins pedagógicos, das avaliações dos alunos, o que deveria, segundo os participantes do subgrupo, estar fortemente refletido na arquitetura das avaliações, em especial na escolha das séries a serem avaliadas. A integração entre avaliações estaduais e nacionais foi um tema amplamente discutido, já que os estados contemplam avaliação estadual, principalmente, como complementar à avaliação nacional. Ao mesmo tempo, demoras nas devolutivas e outras deficiências das avaliações nacionais foram destacadas como motivos para que os estados avaliam as mesmas séries e as mesmas áreas de conhecimento. Ainda no tocante à arquitetura, foram discutidas maneiras utilizadas pelas avaliações para inclusão de escolas com menos de 20 alunos.

Finalmente, houve um debate sobre como as escolas urbanas, rurais e indígenas são avaliadas da mesma maneira, mesmo que haja objeções quanto à comparação entre esses diferentes tipos de escola. Uma sugestão foi separar os resultados das avaliações do ensino urbano e do ensino rural ou indígena. Subjacente a essa discussão, foram mencionadas as peculiaridades do ensino para alunos com necessidades especiais/deficientes, para os quais foi sugerido um acompanhamento dos resultados da avaliação que inclua o nível do aluno, e ensino para alunos em privação

2.1.2. SUBGRUPO DE USOS E ARQUITETURA

Agradecemos o intenso e produtivo trabalho dos membros do Subgrupo de Usos e Arquitetura, que foram: coordenadora Fátima Cristina D. Medeiros (Bahia), coordenador adjunto Afonso Gomes Ferreira Filho (Rio Grande do Norte), Aparecida Meireles de Souza e Souza (Rondônia), Augusto Cezar Barbosa Brito (Amapá), Carmilva Souza Flores (Ceará), Evandro dos Santos Paiva (Pará), Irene de Souza Costa (Mato Grosso), Sandra Mara Cardoso (Santa Catarina) e Vania Machado Oliveira (Rio de Janeiro).

O percurso de discussão foi seguido de acordo com quatro perguntas trazidas pela facilitadora Carolina Serain, começando pela percepção do subgrupo a respeito do diagnóstico obtido pelo questionário sobre avaliação da educação básica respondido por todos os representantes estaduais. Vários comentários foram feitos quanto à dificuldade de obter informações sobre custo. Outra sugestão apresentada foi que, uma vez que a pesquisa foca nos dois anos mais recentes, poder-se-ia pedir que cada estado traçasse um breve histórico do sistema de avaliação no seu estado. Assim saberíamos há quanto tempo ele existe, teríamos uma série histórica dos dados existentes, saberíamos quem fez e interrompeu avaliações, quando interrompeu e o motivo. Por exemplo, a Bahia afirmou ter interrompido a avaliação por algum tempo no passado, dadas restrições orçamentárias.

Finalmente, parte do subgrupo se manifestou com a crítica de que se está falando de avaliação de uma forma restrita. Muitas vezes o estado tem uma outra forma de avaliação, por exemplo utilizando os dados [agregados] do INEP (ex: taxa de reprovação) em sua avaliação. A essa preocupação foi respondido que uma maneira de ampliar o escopo contemplado seria, ao longo dos próximos passos, falar de monitoramento em um sentido mais amplo, abordando, por exemplo, como os indicadores das avaliações se compõem com outros indicadores.

Sobre outras informações que poderiam ser levantadas a respeito das avaliações estaduais, o subgrupo concordou que seria interessante saber se existe uma avaliação das ações que foram implementadas para sanar as dificuldades ou resultados não favoráveis encontrados nas avaliações. O subgrupo lembrou que muitas vezes acredita-se que o ciclo se encerra na devolutiva da secretaria para as escolas com o resultado da avaliação. Deve-se pensar, no entanto, que o ciclo se encerra com a devolutiva da escola para a secretaria, com um acompanhamento das ações implementadas após o resultado da avaliação. Do contrário, não se sabe o que foi feito.

Em seguida, foram delimitados os âmbitos em que os resultados das avaliações podem ser usados, na seguinte ordem: formação de profissionais de educação, devolutivas para a sociedade e para os alunos, gestão por resultados ou por evidências, promoção de uma cultura de avaliação, sistema de acompanhamento e premiação.

No tocante à formação dos profissionais de educação, que incluem profe ssores, equipe pedagógica e equipe gestora, foi destacado a importância da integração do trabalho das equipes, de avaliação e pedagógica, bem como da definição de papéis nos usos e apropriações de resultados. Foi mencionado que muitas vezes os resultados são apresentados para as regionais. A divulgação direta com os professores muitas vezes não é possível, principalmente se os resultados da avaliação não forem positivos.

Sobre a promoção de uma cultura de avaliação, houve consenso de que existe hoje uma percepção equivocada da avaliação, resistência por parte dos professores, e distorções no que eles acreditam sobre a utiliade da avaliação. Outro tema polêmico é a premiação e bonificação. Por um lado, ela é vista como um incentivo que faz com que se obtenha bons resultados em termos de alcance de metas. Por outro, existe uma tendência a tornar a premiação, e assim a a avaliação, como um fim em si próprio.

Finalmente, o subgrupo elencou aspectos da arquitetura a serem levados em consideração, que incluiam: público-alvo e anos críticos, periodicidade, formato de itens e participação dos municípios na avaliação. Sobre os anos críticos, não houve consenso sobre quais anos e séries devem ser avaliados, mas sim de que avaliações de todos os anos são desperdício de recurso. Quanto à periodicidade, ponderou-se que, se for muito espaçada, perde-se o tempo de intervenção, mas se for muito curta, não há tempo para se observar uma mudança e para utilizar um volume alto de dados. Foi destacado que, para se definir a periodicidade, tem-se que definir primeiramente o uso, e que pode-se pensar em uma periodicidade mista, com algumas avaliações anuais e outras mais espaçadas. Sobre abrangência das redes de ensino, não houve consenso no subgrupo se deve-se focar na rede estadual somente ou nos municípios também. O subgrupo discutiu que a adesão pode ser gradual, à medida em que um município vê outro fazendo uso dos resultados da avaliação estadual, e então decide aderir.

2.1.3. SUBGRUPO DE CUSTOS E ORÇAMENTO

Com pesar destacamos postumamente o empenho da integrante Maria Cristina Guilherme (Acre) e agradecemos o intenso e produtivo trabalho dos membros do Subgrupo de Custos e Financiamento: coordenador Eduardo Malini (Espírito Santo), coordenador adjunto Cosme Rocha (Piauí), Soraya Regina Cruz (Mato Grosso do Sul), Katya Prust (Paraná), Fredson da Costa (Amapá), Marinaldo Alves (Pernambuco) e Geniana Faria (Minas Gerais).

A pauta inicial proposta pelas facilitadoras Gabriela Pluciennik e Juliana Candian enfocou na análise de como são registrados os custos das avaliações estaduais da educação básica, em vista de que em torno de 21 estados contratam serviços de avaliação externa pelo CAED. Foi

proposto que entender as variáveis de custos deste fornecedor pode indicar um caminho para chegar ao entendimento do que seriam custos fixos e variáveis do contrato. Mesmo que existem diferenças, houve o entendimento da importância de um esforço de análise dos custos de avaliação. Foi lembrado que há uma portaria recente que limita gastos com avaliação e que este documento (Decreto 7590, posteriormente compartilhado) também deve ser considerado base para a discussão.

O subgrupo elencou as rubricas em que são divididos os custos das avaliações, incluindo: Oficina de elaboração de itens, pré-testagem do item, desenho e planejamento, produção dos instrumentos, impressão dos instrumentos (e manuseio), capacitação de equipe, distribuição e aplicação, recolhimento, processamento, análise e divulgação (incluindo impressão e divulgação dos cadernos/boletim, boletim online em servidor e oficinas ou seminários). Além desse quadro, o subgrupo identificou outras variáveis que impactam nos cursos dos serviços, como as quantidades de alunos, escolas, regionais e municípios. Ainda, para a elaboração do novo questionário de pesquisa, o subgrupo sugeriu a inclusão de informações sobre fontes de financiamento (que podem ser Tesouro, recursos internos ou externos, como bancos de desenvolvimento), sugerindo também estudos e pesquisas para compreensão dos meios de financiamento em outros países.

Foi sugerido que todos os estados sejam instigados a analisar cada contrato individualmente, com objetivo identificar o que os contratos têm em comum e o que não têm, bem como quais categorias são comparáveis e quais não, como valores de logística e remuneração de equipe, que varia de local a local. No final do trabalho em subgrupo, foi sugerida a criação de um painel de serviços, não com o objetivo de suprimir a autonomia dos modelos estaduais, mas com a possibilidade da realização de consórcios entre os estados.

2.2. 2ª REUNIÃO TÉCNICA DO GRUPO DE TRABALHO DE AVALIAÇÃO DO CONSED, EM 17 E 18 DE AGOSTO DE 2016

O principal foco das discussões dos Subgrupos de trabalho na 2ª Reunião Técnica do GT de Avaliação foi a elaboração de configurações de possíveis desenhos de avaliação em que os principais aspectos fossem contemplados de três maneiras: básica, atendendo às necessidades mínimas, intermediária e avançada. Essas elaborações foram ancoradas em uma série de perguntas para a reflexão de cada um dos Subgrupos, elaboradas pela equipe do Instituto Ayrton Senna. A seguir, serão resumidas as discussões geradas nesse contexto por cada um dos subgrupos.

2.2.1. SUBGRUPO DE CONTEÚDO E ARQUITETURA

Durante a 2ª Reunião Técnica do GT de Avaliação, o Subgrupo de Conteúdo passou a discutir, adicionalmente, o tema de arquitetura das avaliações, anteriormente alocado a outro subgrupo. O trabalho em subgrupo enfocou em responder as perguntas para discussão sobre o conteúdo e arquitetura das avaliações, com destaque para os temas a seguir.

A inclusão de disciplinas do currículo na avaliação foi o primeiro tema. As opções ponderadas foram de considerar todas as disciplias ou todas as quatro áreas de conhecimento (linguagens, matemática, ciências humanas e da natureza, ou apenas língua portuguesa e matemática.

Em seguida, deliberou-se sobre a matriz de referência. A recomendação do subgrupo foi que a matriz de referência seja do SAEB, à luz de eventuais mudanças trazidas pela a Base Nacional Comum Curricular (BNCC), que estabelece 60% de conteúdo comum nos currículos nacionais. Mesmo que o currículo todo não seja contemplado, a matriz de referência tirada dali deverá representar no máximo pequenas mudanças na matriz do SAEB, estabelecida em 2005. Lembrou-se que é importante assegurar a comparabilidade da série histórica, a despeito de mudanças curriculares.

As outras avaliações do aluno também foram abordades. Idealmente, deveriam incluir interesse, exposição e engajamento dos alunos nas atividades escolares; progresso e retenção escolar; condição socioeconômica da família; competências socioemocionais; atitude em relação ao aprendizado; percepção do clima escolar e interesse dos pais nas atividades escolares dos seus filhos. Os três primeiros itens foram considerados indispensáveis num cenário básico, enquanto para um cenário intermediário foi sugerido, além desses, a inclusão do quarto item.

Foi discutido que a avaliação do desempenho dos alunos deveria vir apoiada por informações advindas dos professores, idealmente com uma auto avaliação deles sobre sua motivação e interesse pelas atividades escolares, bem como sobre aspectos institucionais ligados ao desempenho da coordenação pedagógica, às formas de gestão, às necessidades da escola e ao clima escolar. O único item que os participantes julgaram ser adequado apenas a uma avaliação avançada foi a avaliação do clima escolar, devendo os outros itens necessariamente figurarem em avaliações de escopo básico ou intermediário.

De modo semelhante, a visão dos diretores deve ser contemplada, nos aspectos de desempenho e necessidades da escola, sobre o apoio recebido das secretarias e regionais e, em avaliações intermediárias e avançadas, clima escolar. Além dessas, uma avaliação avançada poderia, segundo o subgrupo, abordar a motivação e interesse dos diretores pelas atividades escolares.

Sobre avaliação institucional, foi considerada fundamental em todos os escopos de avaliação, nos âmbitos gestão pedagógica, gestão de resultados educacionais, gestão participativa, gestão de pessoas e gestão de serviços e recursos. Esse mesmo escopo deveria, segundo os participantes, ser aplicado nas regionais e secretarias de educação.

A inclusão de informações sobre os profissionais de educação foi um tema amplamente discutido. Foi discutido pelo subgrupo a necessidade de diferenciar as informações obtidas dos professores e dos outros profissionais envolvidos no sistema educacional, mesmo os que desempenham funções de apoio nas escolas. Idealmente, seriam avaliados suas competências cognitivas e conhecimentos específicos sobre as funções realizadas (disciplinas ministradas, práticas pedagógicas, gestão escolar etc.), atitudes, motivação, competências socioemocionais e hiatos de formação. Em escopos de avaliação básico e intermediário, sugeriu-se a possibilidade de omitir avaliação de competências socioemocionais e hiatos de formação.

Na sequência, foram respondidas pelo subgrupo as perguntas referentes à arquitetura da avaliação. Como modelo ideal ou avançado, foi definida uma avaliação anual, no 3º, 5º, 7º e 9º anos do Ensino Fundamental e no 1ª e 3ª anos do Ensino Médio, censitária para todas as escolas públicas (inclui escolas rurais, indígenas e quilombolas) com ao menos 20 alunos no ano/série avaliado, exceto 3º ano do Fundamental quando o mínimo deve ser 10 alunos, e amostral nas escolas particulares. Em um cenário intermediário, a avaliação seria bianual, nas mesmas séries mencionadas acima e com a mesma abrangência. Em um cenário básico, a avaliação bianual se restringiria aos seguintes anos: 3º, 5º, e 9º do Ensino Fundamental e 3ª do Ensino Médio, com as mesmas especificações adicionais anteriores. As disciplinas incluídas nessas três versões da arquitetura foram discutidas anteriormente dentro do tema de conteúdo.

2.2.2. SUBGRUPO DE USOS

O primeiro tema discutido pelo subgrupo foi sobre quais usos pretendemos fazer do sistema de avaliação a ser construído, no qual foram mencionados o uso para formação (promover o desempenho educacional), diagnóstico, avaliação de políticas, apoio à gestão educacional, priorização de esforços, reconhecimento e certificação (com cuidado para não

transformar em premiação, no intuito de não interferir na cultura da informação), identificação (triagem de boas práticas) e compartilhamento de boas práticas, monitoramento (dos resultados e do uso dos resultados na escola), comparação de práticas e, por fim, colaboração entre pares.

Sobre os usos pedagógicos, o primeiro seria a formação docente, trabalhando os descritores dos itens que mais foram acertados e os que menos acertos tiveram, na formação do professor. Outros usos seriam: intervenção pedagógica (ações diferenciadas para escolas que precisem de atenção especial por terem maior dificuldade) e prática pedagógica.

Em seguida, foram discutidos temas a respeito de como o sistema de avaliação pode influenciar diferentes níveis da prática dos profissionais de educação e do funcionamento da escola. Foi destacada a importância da auto avaliação institucional no sentido de identificar as dimensões da escola que poderiam ser melhoradas, como por exemplo indicadores externos, internos e envolvimento da família.

Nos usos voltas à gestão do sistema escolar, uma possibilidade identificada foi a de priorização das escolas para intervenção e planos de ação claros com responsáveis claros

Outra possibilidade de uso discutida foi a de atrelar uma avaliação de desempenho do professor ao desempenho do aluno faz sentido. Entretanto, mencionou-se o risco de resistência por parte dos professores, o que inviabilizaria esse uso.

Discutiu-se, também, sobre como potencializar os usos, seja através de melhoria do conteúdo e da forma de apresentação dos resultados da avaliação, ou de formação da comunidade escolar para a compreensão das devolutivas disponíveis, ambos considerados importantes. A melhoria das devolutivas foi um tema abordado, com a sugestão de uso de plataformas interativas (ex. Foco em Aprendizagem), subgrupos focais, adequação e melhoria da linguagem (com exemplos, vídeos e depoimentos), vídeos específicos para a família (mais lúdicos e sem caráter publicitário).

Ainda dentro de fatores que potencializam o uso, a formação dos educadores, especificamente para compreender, ler e saber melhor utilizar os resultados das avaliações foi considerada de enorme importância. Duas possibilidades foram contempladas para atingir esse objetivo: formar formadores ou formar diretamente os professores. Houve consenso no

subgrupo de que o ideal seria ir diretamente ao professor, mas na realidade a secretaria pode não ter pessoal suficiente para isso.

Quanto aos profissionais das secretarias, sua formação deve priorizar a capacidade de utilizar resultados das avaliações para o planejamento, em particular o planejamento da formação continuada dos educadores.

Em resumo, o cenário ideal de usos da avaliação deveria contemplar, segundo o subgrupo, as seguintes áreas: Prestação de contas (informar a sociedade dos progressos alcançados e do alcance ou não das metas estabelecidas; identificação de escolas e regionais com alto desempenho para compartilhar boas práticas); Planejamento e gestão (subsidiar o órgão central, as regionais e a comunidade escolar na sua auto avaliação e diagnóstico; planejamento, alocação de recursos, identificação de ações corretivas e mudanças de rumo); Pedagógico (subsidiar escolas na sua busca por melhores estratégias ou planos pedagógicos; dar subsídios aos professores de tal forma que esses possam melhor adequar as suas práticas às necessidades dos alunos) e de Formação (dar subsídios aos professores, escolas e à Secretaria para a formulação de planos de formação continuada e fornecer insumos ao processo de seleção de professores). Em um cenário intermediário, poderiam ser suprimidos os usos para melhoria das práticas dos professores em relação às necessidades dos alunos e o uso dos resultados das avaliações para a seleção de professores.

2.2.3. SUBGRUPO DE CUSTOS E ORÇAMENTO

Dentro da proposta da reunião, foram discutidas quais as dificuldades de se realizar avaliações estaduais, que incluem custos para contratação dos serviços terceirizados, ausência de política e legislação específica que garanta a continuidade das ações mesmo na troca de gestão, rotatividade do corpo técnico que dificulta a formação de uma equipe qualificada, entendimento da equipe de currículo do que é uma avaliação em larga escala, e falta de formação dos professores em avaliação.

Em seguida, foi debatido em que medida os recursos humanos disponíveis e alocados nas Secretarias para a função de monitoramento e avaliação estão adequadamente qualificados e se existem na quantidade necessária. Foram constatadas grandes deficiências nos âmbitos de existência de banco de itens e pessoal qualificado para a manutenção e desenvolvimento desse banco, pessoal qualificado para seleção e construção das avaliações de desempenho, para a construção ou o acompanhamento de consultores na construção das avaliações

socioemocionais, de motivação, estilo de aprendizagem e clima escolar. Apesar do subgrupo considerar que as secretarias têm pessoal qualificado para acompanhar a construção de avaliações institucionais, os quadros disponíveis seriam insuficientes para acompanhar a execução.

Outra área em que a quantidade de profissionais e suas especializações seriam insuficientes seria para compilar e calcular indicadores de resultado a partir das informações brutas obtidas nas avaliações, bem como para interpretar os indicadores de resultados obtidos nas avaliações, bem como consolidar, interpretar e analisar os resultados das avaliações institucionais.

Em relação ao acompanhamento de devolutivas, o subgrupo foi unânime sobre a pouca especialização dos quadros para desenhar devolutivas a partir dos indicadores de resultados obtidos nas avaliações para pais, escolas e secretaria, bem como sua insuficiência numérica para realizar tais devolutivas.

Em relação às restrições operacionais não relacionadas a disponibilidade de recursos humanos, foi destacado que há falta de recursos em diagramação, editoração e impressão dos materiais utilizados na avaliação.

Em seguida, foi discutido pelo grupo como se chegar a uma estimativa de custo por aluno para implementar um sistema de avaliação em um estado. A sugestão foi que esse cálculo fosse feito por disciplina e por aluno, para cada ano de avaliação, excluindo valores gastos com logística.

Sobre o custo adicional de incluir as escolas municipais, inicialmente, o subgrupo entendeu que o aumento de custos é mínimo, pois já se aproveita a rede logística. No andamento da discussão, entenderam que a custo de gráfica também aumentaria. Na conclusão, não houve consenso no subgrupo, alguns consideram que o impacto é significativo, outros entendem que a aumento é mínimo.

Como estratégias de gestão, foi recomendada uma maior participação da equipe estadual nas avaliações, contratando externamente elaboração, aplicação e resultado. Para a secretaria, as responsabilidades seriam a devolutiva e a disseminação de resultados e formação na rede, dependendo do corpo técnico de cada equipe.

Em seguida, debateu-se como um consórcio entre estados poderia aproveitar sinergias e economias de escala e escopo para reduzir custos, principalmente na elaboração de itens, que podem ser trocados e compartilhados entre os estados, na contratação de gráficas, realizando as avaliações no mesmo período, na logística, na formação dos técnicos e treinamento de aplicadores.

Sobre o financiamento, foram destacadas as fontes de recursos públicas e privadas. Para o subgrupo, o governo federal deveria ter uma rubrica Avaliação para financiar e repassar recursos para o Sistema Nacional de Avaliação, um Fundo de Avaliação, com recurso vinculado: Fundo Federal de Avaliação. Mencionou-se também a possibilidade também de parceiros privados, como bancos, destinassem recursos para o Sistema Nacional de Avaliação, investindo parte do lucro como investimento social.

Finalmente, houve debate sobre a exigência de uma contribuição financeira das escolas particulares no sistema de avaliação. Para o subgrupo, é desejável e possível, com base na possibilidade de o Conselho Estadual de Educação autorizar apenas as escolas que participem do Sistema Nacional de Avaliação, assumindo também custos. Parte do subgrupo, no entanto, questionou a viabilidade por forças políticas e na influência dos grupos de educação privada em Conselhos e outras esferas do governo.

2.3. 3ª REUNIÃO TÉCNICA DO GRUPO DE TRABALHO DE AVALIAÇÃO DO CONSED, EM 8 DE NOVEMBRO DE 2016

Na 3ª Reunião Técnica, os integrantes do GT de Avaliação foram divididos em seis grupos temáticos, para refinamento de questões pendentes nas reuniões anteriores. Além disso, três representantes estaduais (Eduardo Malini, do Espírito Santo, do Rio Grande do Sul e do Pará) concentraram-se, juntamente com Ricardo Paes de Barros, nos cálculos de custo médio por aluno das avaliações estaduais da educação básica, segundo as informações recebidas. A seguir, serão apresentados breves resumos das deliberações.

2.3.1. DISCUSSÕES SOBRE AVALIAÇÕES DOS ALUNOS

Agradecemos a participação de Shirlene Noronha Guimarães (Amazonas), Carla Bertânia Souza (Rio de Janeiro), Fátima Cristina D. Medeiros (Bahia) e Sandra Mara Cardoso (Santa Catarina). O foco das discussões foi a maneira de possibilitar que as avaliações censitárias da

educação básica estadual, a serem realizadas anualmente, cubram todas as áreas de conhecimento dos alunos. Uma possibilidade mencionada foi a realização de avaliações com focos variados a cada ano – a exemplo do PISA. Alternativamente à avaliação censitária, uma opção seria sortear amostra dentro das turmas. As avaliações dos conteúdos presentes no currículo, segundo ressaltaram os participantes, devem medidas em escalas nacionalmente definidas a partir da BNCC. Além das avaliações anuais, discutiu-se a inclusão da avaliação formativa, com periodicidade a ser definida.

Além da maneira de inclusão das áreas de conhecimento, outro tema tratado pelos participantes foi a necessidade de garantir a divulgação de resultados de tal maneira que a escola, no início do ano letivo, tenha informação sobre desempenho da turma ao final do ano letivo anterior. Para tanto, destacou-se a importância de garantir que exista um identificador de aluno de tal maneira que a escola no início do ano letivo tenha informação sobre desempenho dos seus alunos ao final do ano letivo anterior.

2.3.2. DISCUSSÕES SOBRE USOS PEDAGÓGICOS

Agradecemos a participação de Carmilva Souza Flores (Ceará), Cyntia Lemes da Silva G. de Fonseca (São Paulo), Marinaldo Alves (Pernambuco) e Hilda Maria Freire Montysuma (Roraima). Os participantes expressaram concordância com os usos pedagógicos das avaliações propostos no cenário mais completo da proposta de avaliação da educação básica em discussão para aplicação estadual. As deliberações centraram-se em três âmbitos dos usos pedagógicos dos resultados das avaliações: formação inicial dos professores, formação continuada dos profissionais de educação e planos pedagógicos.

Quanto ao primeiro tema, concordou-se que os resultados das avaliações podem subsidiar as instituições de ensino superior, influenciando o conteúdo dos cursos de formação inicial. Esse uso, no entanto, foi visto como mais dispendioso do que o uso para formação de professores já em exercício, que deve ser o foco. Foi ressaltada a importância de dar subsídios aos professores, escolas e à Secretaria para a formulação de planos de formação continuada.

Finalmente, foram discutidas possibilidades para a efetivação do uso dos resultados das avaliações para subsidiar escolas na sua busca por melhores estratégias ou planos pedagógicos. Uma observação do grupo foi que esse uso deve ser feito conjuntamente com todos os atores envolvidos, sejam eles a comunidade discente, docente ou familiares dos alunos.

2.3.3. DISCUSSÕES SORE USOS PARA GESTÃO DAS AVALIAÇÕES

Agradecemos a participação de Weyne Maria Magalhães Carneiro (Goiás), Augusto Cezar Barbosa Brito (Amapá), Emerson Azevedo Soares (Tocantins) e Irene de Souza Costa (Mato Grosso). Os participantes dessa rodada de discussão centraram suas discussões sobre usos para prestação de contas, incluindo informar a sociedade dos progressos alcançados e do alcance (ou não) das metas estabelecidas, uso para identificação de escolas e regionais com alto desempenho para compartilhar boas práticas, bem como usos de planejamento e gestão, incluindo subsidiar o órgão central, as regionais e a comunidade escolar na sua auto avaliação e diagnóstico, planejamento, alocação de recursos, identificação de ações corretivas (plano de ação) e mudanças de rumo.

Sobre a definição da política de formação para uso dos dados e definição do modelo para engajamento dos atores envolvidos, foram discutidos modelos de apropriação dos resultados por professores e gestores realizado por meio de seminários para os órgãos centrais e para as regionais. Neste modelo, os representantes das regionais que atendessem o seminário seriam transformados em multiplicadores, que posteriormente disseminariam o resultado nas escolas a eles jurisdicionadas.

Foi ressaltado o uso para definição de melhores práticas. Houve consenso sobre a existência de melhores práticas e acha importante ressaltá-las. O modo de implementação poderia ser individualizado, onde uma escola aprenderia com outra escola que tenha solucionado o mesmo problema.

Sobre a definição de metas, foi recomendado um modelo de metas bianual, com definição da política de formação para uso dos dados e definição do modelo para engajamento para os atores envolvidos. O ideal, segundo os participantes, é que seja constituído um fórum ampliado de discussão para a definição de metas e modelos de mensuração do progresso. Desta forma, seria possível alcançar uma política de estado, em contraponto a não-desejadas políticas de governo, que mudam de prioridades e interesses a cada ciclo eleitoral, já que poderia também ser feita a institucionalização das metas esperadas através de projeto de lei.

Ainda com relação à definição de metas e progresso das escolas, ponderou-se ser mais prudente um atendimento individualizado, onde a situação escola por escola é avaliada, liderada pela secretaria e em conjunto com a regional. Existe consenso de que é necessária uma

definição de metas claras para se resolver problemas específicos da escola (de gestão, por exemplo, caso o baixo desempenho da escola seja devido a uma alta evasão escolar, ou dado o transporte muito longo dos alunos). Muitas vezes, o problema não se trata de redefinição de metas de desempenho escolar, mas metas para se resolver o problema que causa o baixo desempenho. Como encaminhamento, a proposta seria mobilizar todas as instâncias envolvidas e necessárias para a implementação de uma medida corretiva.

2.3.4. DISCUSSÕES SOBRE AVALIAÇÃO DE PROFISSIONAIS DE EDUCAÇÃO E ESCOLA

Agradecemos a participação de Cheila F. B. S. Vasconcelos (Alagoas), Marta de Medeiros Correia (Paraíba), Aparecida Meireles de Souza e Souza (Rondônia) e Soraya Regina Cruz (Mato Grosso do Sul). Os participantes dessa rodada de discussão, com base no modelo de conteúdo e arquitetura desenhado nos encontros anteriores, esclareceram alguns pontos e definiram prioridades em uma possível implementação, a partir de uma visão de que as avalições podem gradativamente incorporar elementos em seu desenho.

Sobre a avaliação as competências e percepção dos profissionais da educação, observou-se que é necessária mais clareza na definição de profissionais de educação. A proposta é trabalhar com professor, coordenador, diretor e os demais profissionais das escolas. Quanto às competências e percepções específicas desses profissionais, a ideia de avaliar hiatos de formação também foi vista como pouco clara. Foi sugerida a avaliação sobre formação acadêmica e formação continuada, assim como o índice de permanência e rotatividade. A seleção de diretores também foi vista como uma como uma variável importante de ser avaliada.

Como modificação ao conteúdo sugerido para um escopo de avaliação básico dos profissionais de educação, houve unanimidade que esse deveria, também, incluir a percepção do clima escolar, avaliação da coordenação pedagógica, das formas de gestão e das necessidades da escola. Também complementando as deliberações anteriores, foi destacado que, independentemente da periodicidade das avaliações dos alunos, a avaliação dos profissionais de educação deveria ser bianual.

2.3.5. DISCUSSÕES SOBRE AVALIAÇÃO INSTITUCIONAL

Agradecemos a participação de Joniely Cheyenne Cruz (Sergipe), Katya Prust (Paraná), Amanda Midori (Distrito Federal) e Afonso Gomes (Rio Grande do Norte). Os participantes centraram a discussão na melhor definição da avaliação institucional, de acordo com a experiência acumulada pelos participantes dessa rodada de discussão. Os níveis de análise para a avaliação institucional foram: na área de condições materiais, mencionou-se infraestrutura, recursos, condições de trabalho. A redação proposta foi infraestrutura e condições de usos. Outro tópico que foi consenso entre os integrantes foi sobre a inclusão de gestão de pessoas e gestão financeira na avaliação institucional. Sobre a dimensão da avaliação democrática, foi ressaltada a importância de inclusão desse quesito devido aos processos de gestão da escola que devem garantir participação de todos os atores.

Também foi da opinião dos representantes que a utilização dos dados das avaliações externas e internas, sejam elas de desempenho do aluno, avaliação docente ou de características como socioemocionais, devem ser usadas pela escola em seu processo pedagógico e de gestão, com ampla participação dos atores envolvidos. A avaliação institucional seria um momento útil para pesquisar se essa utilização está ocorrendo da maneira adequada, de forma inclusiva, compreensiva e sustentável. Por isso foi sugerido incluir a dimensão de utilização de dados de avaliação internas e externas. O ciclo da avaliação, que inclui a utilização dos resultados, passaria assim a ser avaliado e discutido pela escola, em busca de melhoramentos.

Sobre a arquitetura das avaliações institucionais, os participantes ponderaram que não se pensou, nos trabalhos anteriores do GT, sobre a forma como se dará o uso das avaliações institucionais e como seriam feitas as devolutivas das avaliações. Isso teria direta relação com a periodicidade, por isso avaliações institucionais bianuais seriam recomendadas. Sobre a avaliação ser censitária, todas as escolas da rede avaliada devem ser incluídas, mas dentro de cada escola a coleta de dados deve ser feita de modo diferenciada. Para os profissionais de educação, a opção preferida pelos participantes foi de incluir todos os professores, mesmo das séries não avaliadas. Outra opção seria avaliar todos os professores das séries avaliadas e não das séries não avaliadas, mas do conjunto dos os outros profissionais de educação, incluindo gestão da escola, das regionais e da Secretaria. Quanto à abrangência de alunos, a recomendação foi de avaliação censitária, para os estudantes das séries avaliadas a partir do 4º ano e por adesão para os pais., a recomendação seria censitária, para os estudantes das séries avaliadas a partir do 4º ano e por adesão para os pais.

Também foram discutidos temas mais polêmicos, como a avaliação de desempenho docente, que pode ter empecilhos nas regulamentações do emprego docente, e sobre se as escolas particulares deverem ser ou não incluídas.

2.3.6. DISCUSSÕES SOBRE OUTRAS INFORMAÇÕES DOS ALUNOS

Agradecemos a participação de Cosme Rocha (Piauí), Luis Alex Lopes Correia (Maranhão) e Maria Cristina Guilherme (Acre). Os integrantes do grupo iniciaram as discussões com uma reelaboração das outras informações dos alunos a serem levadas em consideração, que são: atitudes, interesses, engajamento dos alunos nas atividades escolares; exposição às atividades escolares; progresso e retenção escolar; condição socioeconômica da família; competências socioemocionais; percepção do clima escolar; práticas pedagógicas e de gestão da sala de aula e interesse dos pais nas atividades escolares dos seus filhos.

Dada a impossibilidade de que todas essas competências, além das disciplinas ou áreas de conhecimento, sejam incorporadas às avaliações dos alunos, foram discutidas possibilidades para viabilizar a avaliação dessas competências. Uma primeira sugestão foi de coletar informações antes da prova (em formulários online), fazendo uso da tecnologia para possibilitar uma mais abrangente avaliação dos alunos. Outra estratégia sugerida foi a substituição de coleta primária por secundária, na medida da existência das informações em registros administrativos e avaliações nacionais (Censo Escolar, SAEB).

Como alterações na arquitetura das avaliações, foi apresentada a ideia de priorizar as séries em que os alunos terão mais condições de responder ao questionário sobre essas competências, que seria do 9º ano do Ensino Fundamental em diante. Também foi discutida a possibilidade de que essas avaliações sejam conduzidas de forma amostral dentre os alunos, preservando a informação censitária ao nível de escola.

Para que seja decidido sobre a adequação ou não desses modos de implementação, sugeriu-se ouvir de especialistas as perdas e os ganhos em relação às alternativas listadas. Outro aspecto para o qual a opinião de especialistas pode trazer um maior aprofundamento quanto a como integrar essas informações nas avaliações é sobre o risco de viés de resposta pelos alunos (em especial, nos quesitos relacionados a atitudes, interesses, socioemocionais e clima), e maneiras de tratar esse risco.

Finalmente, os participantes listaram possíveis fontes de informação que devem ser integradas para que se consiga, com máxima economia de recursos, a maior amplitude e profundidade de informações sobre os alunos.

2.4. 4ª REUNIÃO TÉCNICA DO GRUPO DE TRABALHO DE AVALIAÇÃO DO CONSED, EM 4 DE MAIO DE 2017

Com participação de todos os membros do Grupo de Trabalho, foi realizado um debate sobre os principais pontos da proposta de avaliação, elaborada a partir de opções resultantes das três reuniões do GT e escolhidas pelos Secretários de Educação em votação. Vários representates estaduais relataram sobre os desafios de seus sistemas de avaliação estadual. Algumas ponderações iniciais ressaltaram os empecilhos para estabelecer um novo sistema, dada a variedade de avaliações já existentes e a dificuldade de garantir a vontade política para expandir o leque de avaliações realizadas.

Sobre novos desenvolvimentos que influenciam a proposta do GT, foram mencionadas as mudanças trazidas pela mudança em curso que fará o ciclo da alfabetização terminar no 2º ano, o que traz impactos para o desenho da avaliação. Uma observação foi feita sobre a determinação de que o mínimo de 20 estudantes por série para realização da avaliação seja mudado em consonância com novas diretrizes do SAEB, que agora colocam um mínimo de 10 alunos por série.

Alguns pontos subjacentes à proposta também foram abordados, como dificuldades logísticas trazidas pela prova de dois dias. Foi proposto um modelo que pode ser gradativo para que se chegue ao modelo máximo preferido por muitos Secretários. Também sobre os princípios subjacentes à proposta de avaliação, mencionou-se que, para que se possa pensar em sistema de avaliação, deve-se entender de forma profunda a rede de educação, que, diferentemente da saúde, não tem protocolos pactuados em um sistema único. Nesse sentido, foi lembrado que o que foi combinado com o GT seria pensar em um cenário ideal e um cenário possível, e que as avaliações anuais teriam grandes desafios, por conta das limitações institucionais, para que haja compreensão e uso dos resultados, principalmente devido às condições de trabalho a que estão sujeitos os professores.

Adicionalmente, foram discutidas demais preocupações com o ciclo da avaliação anual, já que, para que o uso dos dados seja realmente feito de forma satisfatória, é necessário o entendimento apropriado dos professores de como aperfeiçoar a prática pedagógica a partir dos resultados. Ressaltou-se que o uso efetivo dos resultados deve ser o foco para a rede de ensino, desde secretaria até professores, antes que a expansão das avaliações faça realmente sentido e possa trazer resultados significativos de melhoria de aprendizado. Para isso, as oficinas de apropriação de resultados devem ser aprimoradas, para que o entendimento das deficiências dos alunos seja acompanhado de um entendimento de como integrar esses resultados nas práticas pedagógicas.