

AVALIAÇÃO DE IMPACTO EM EDUCAÇÃO

A experiência exitosa do programa
Jovem de Futuro em parceria com o
poder público

Ricardo Henriques
Mirela de Carvalho
Ricardo Paes de Barros



AVALIAÇÃO DE IMPACTO EM EDUCAÇÃO

**A experiência exitosa do programa
Jovem de Futuro em parceria com
o poder público**

**Ricardo Henriques
Mirela de Carvalho
Ricardo Paes de Barros**



REALIZAÇÃO

Instituto Unibanco

CONSELHO DE ADMINISTRAÇÃO

Presidente

Pedro Moreira Salles

Vice-Presidente

Pedro Sampaio Malan

Conselheiros

Antonio Jacinto Matias

Claudia Costin

Cláudio de Moura Castro

Cláudio Luiz da Silva Haddad

Marcelo Luis Orticelli

Marcos de Barros Lisboa

Ricardo Paes de Barros

Rodolfo Villela Marino

Diretoria

Cláudio José Coutinho Arromatte

Jânio Francisco Ferrugem Gomes

Leila Cristiane Barboza Braga de Melo

Marcelo Luis Orticelli

Moises João do Nascimento

Paulo Sergio Miron

Valéria Aparecida Marretto

EQUIPE TÉCNICA

Superintendente Executivo

Ricardo Henriques

Gerentes

César Nunes

Maria Julia Azevedo

Mirela de Carvalho

Tiago Borba

Vinicius da Silva Scarpì

ELABORAÇÃO DO MATERIAL

Autores

Ricardo Henriques

Mirela de Carvalho

Ricardo Paes de Barros

Produção de conteúdo

Beatriz Silva Garcia

Fabiana Silva Bento

Gabriel Guimaraes Leite

Laura Muller Machado

Mariana Bittar

Paula Penko Teixeira

Samuel Franco

Edição de conteúdo

Antonio Gois

PRODUÇÃO EDITORIAL

Coordenação

José Jacinto Amaral

Revisão

Demetrio Pires Weber Candiota da Rosa

Érico Melo

Projeto Gráfico e Diagramação

Fernanda Aoki

Ilustrações

Mario Kanno

Dados Internacionais de Catalogação na Publicação (CIP)
(eDOC BRASIL, Belo Horizonte/MG)

H519a Henriques, Ricardo.
 Avaliação de impacto em educação: a experiência exitosa do programa
 Jovem de Futuro em parceria com o poder público / Ricardo Henriques,
 Mirela de Carvalho, Ricardo Paes de Barros. – São Paulo, SP: Instituto
 Unibanco, 2020.
 56 p. : il. color. ; 16,5 x 24 cm

Inclui bibliografia
ISBN 978-65-86069-53-2

1. Avaliação educacional. 2. Educação e Estado. 3. Programa Jovem de
Futuro. I. Carvalho, Mirela de. II. Barros, Ricardo Paes de. III. Título.

CDD 378.81

ÍNDICE

INTRODUÇÃO	5
1. O QUE É O PROGRAMA?	11
1.1. As três gerações do programa	14
2. PREPARANDO A AVALIAÇÃO DE IMPACTO DO JOVEM DE FUTURO	21
3. O PROGRAMA TEM IMPACTO?	27
4. MAGNITUDE DO IMPACTO	31
4.1. Relevância do impacto no contexto do histórico brasileiro	33
4.2. Comparações internacionais	38
5. O QUE MUDOU NO IMPACTO AO LONGO DAS GERAÇÕES?	41
6. NOVAS PERGUNTAS E O DEBATE CIENTÍFICO EM EDUCAÇÃO	47
CONCLUSÃO	51
BIBLIOGRAFIA	53

INTRODUÇÃO

As ações de liderança promovidas pelos gestores escolares têm papel fundamental no aprendizado dos alunos. A afirmação pode parecer óbvia, mas nem sempre a evidência científica confirma o que parece intuitivo. Neste caso, porém, há sólido respaldo empírico para esta conclusão. A literatura acadêmica internacional comprova que, entre os fatores intraescolares, apenas a ação dos professores supera a dos gestores em termos de impacto no aprendizado dos estudantes¹.

Além de identificar a importância de uma boa gestão escolar, alguns estudos também têm aprofundado a investigação sobre como diretores e suas equipes impactam o aprendizado dos alunos. Num dos mais citados artigos internacionais em revistas acadêmicas sobre o tema, a neozelandesa Viviane Robinson destacou a importância de o gestor atuar na gestão pedagógica e no desenvolvimento profissional de seus professores².

Por meio do projeto *What Works Clearinghouse*, destinado à realização de revisões independentes de pesquisas sobre o que funciona em educação, o Instituto de Ciências da Educação (IES), braço de estatísticas, pesquisa e avaliação do Departamento de Educação dos EUA, constatou que certas práticas de gestão escolar estão altamente associadas à melhoria dos resultados acadêmicos dos estudantes³. Acompanhando escolas de baixa performance que conseguiram grandes melhoras em curto espaço de tempo, o IES aponta quatro práticas decisivas: (1) liderança capaz de sinalizar a magnitude e a urgência da mudança; (2) capacidade de manter o foco na melhoria da aprendizagem em todas as etapas do processo, usando dados para definir metas de melhoria institucional e

1 Ver Leithwood et al., 2004.

2 Ver Robinson, Lloyd e Rowe, 2008.

3 Ver Herman et al., 2008.

reavaliando continuamente seus resultados com o objetivo de afetar a aprendizagem; (3) capacidade de tornar os avanços visíveis para todos os atores, demonstrando que é possível alcançar a meta abrangente da escola e que o caos não é intransponível; e (4) capacidade de desenvolver e estimular o comprometimento da equipe escolar.

Essa maior atenção ao trabalho do diretor e de sua equipe na escola veio acompanhada do aumento de expectativas sobre seu trabalho. No livro *A Decade of Research on School Principals* [Uma Década de Pesquisas sobre Diretores de Escolas]⁴, os autores Helene Årlestig, Christopher Day e Olaf Johansson, ao compilarem estudos de 24 países diferentes, destacam que um tema comum em todas as nações analisadas “é que os diretores de escola em todos os países estão sujeitos a cada vez mais escrutínio público e a prestar contas aos governos pelos resultados acadêmicos e pela equidade das oportunidades de aprendizado para todos os seus alunos”.

Mesmo com toda a evidência do impacto da gestão no aprendizado, essa dimensão ainda tem recebido pouca atenção dos formuladores de políticas públicas no Brasil e na América Latina. Ao analisarem as políticas de liderança escolar num estudo publicado em 2016, os pesquisadores José Weinstein e Macarena Hernández concluíram que elas ainda estão em estágio inicial na região, “com diversos problemas, contradições e falta de coerência interna e externa”⁵.

Em contraposição ao estágio prematuro do tema na América Latina, sistemas educacionais de alto desempenho têm intensificado sua atuação em políticas de formação, suporte e apoio aos diretores e suas equipes.

Não há dúvidas de que esse corpo de evidências que relaciona liderança escolar aos resultados dos estudantes foi percebido pelos gestores públicos e associações profissionais. Ele informou o desenvolvimento dos padrões de liderança em educação nos Estados Unidos (Council of Chief State School Officers, 2008), o trabalho do National College of School Leadership and Children's Services na Inglaterra (Leithwood; Day; Sammons; Harris; Hopkins, 2006) e o desenvolvimento dos arcabouços de liderança na Austrália e na Nova Zelândia (New Zealand Ministry of Education, 2008). (Robinson, 2011, tradução livre)⁶.

Outro exemplo ocidental emblemático é o da província canadense de Ontário. Michael Fullan, um dos principais arquitetos da reforma que elevou o aprendizado e a taxa de conclusão nas escolas locais, destaca o papel central da gestão nesse processo: “Se vemos uma melhoria em nossas práticas – e os estudos de caso mostram que o aprendizado dos

.....

4 Årlestig; Day; Johansson et al., 2016.

5 Ver Weinstein e Hernández, 2016.

6 Robinson, 2011.

estudantes tem avançado no longo prazo de forma consistente –, então podemos ter certeza da diferença que a liderança está fazendo em Ontario. É inegável. Podemos rastrear, sentir e observar isso”⁷.

ESCOLHAS INICIAIS DO JOVEM DE FUTURO

O Jovem de Futuro é uma parceria público-privada voltada ao aprimoramento das práticas de gestão e liderança nas escolas, regionais de ensino e secretarias de educação. O objetivo final é que essas transformações aumentem a permanência dos jovens na escola para que eles concluam a educação básica com nível adequado de aprendizagem. A parceria é feita com redes públicas estaduais, com foco nas escolas de ensino médio.

O contexto educacional brasileiro da primeira metade da década passada ajuda a entender o porquê da escolha pelo ensino médio. Em 2005, constatou-se que apenas 5% dos jovens que estavam na 3ª série do ensino médio tinham aprendido adequado⁸ em matemática nas escolas das redes públicas estaduais. Além disso, o Censo Escolar mostrava que a taxa de abandono entre estudantes da rede estadual nesse nível de ensino chegou a 18% naquele ano, enquanto o percentual de reprovados foi de 11%. Isso significa que apenas sete em cada dez alunos eram aprovados. Diante desses indicadores, a sociedade aumentava sua conscientização a respeito dos graves problemas dessa etapa de ensino.

Vale lembrar que, a partir de 2007, essas duas dimensões (o aprendizado medido em testes e a taxa de aprovação) passaram a ser sintetizadas num único indicador: o Ideb (Índice de Desenvolvimento da Educação Básica), até hoje o principal instrumento elaborado pelo Ministério da Educação para avaliar a aprendizagem em todo o país.

A escolha da gestão como forma de incidir nos resultados de aprendizagem dos alunos aconteceu num contexto em que a literatura acadêmica nacional e internacional destacava justamente a importância de os gestores atuarem mais no pedagógico, ou seja, menos como burocratas que simplesmente gerenciam recursos humanos e financeiros, e mais como lideranças que mobilizam todos os atores escolares na busca de seu principal objetivo: o aprendizado e o bem-estar dos jovens.

.....
⁷ Conferência no Congresso de Liderança de Ontário, em abril de 2012. Disponível em: <<http://www.edu.gov.on.ca/eng/policyfunding/leadership/OLSQuickFacts.pdf>>. Acesso em: 8 maio 2020.

⁸ Considerando a escala criada em 2006 pela comissão técnica do movimento Todos Pela Educação (2008).

AVALIAÇÃO DE IMPACTO DESDE A ORIGEM

Além do foco no ensino médio e na gestão, outro elemento presente em todas as gerações do Jovem do Futuro foi a preocupação em avaliar, da maneira mais precisa possível, o impacto do programa nos jovens por ele atendidos, tema principal deste livro. São raros no Brasil os entes públicos ou privados que trabalham com o método mais rigoroso de mensuração do real impacto de um programa social: a avaliação experimental. Nesse método, dois grupos são divididos aleatoriamente, de modo que a única diferença observável entre eles seja a participação ou não em determinado programa. Recentemente, essa abordagem ganhou mais visibilidade para o grande público por causa do prêmio Nobel de economia de 2019, concedido aos economistas Esther Duflo, Abhijit Banerjee e Michael Kremer, por suas pesquisas em saúde e educação em populações em situação de pobreza.

No entanto, o uso desse tipo de avaliação para detectar o impacto não é uma novidade no campo científico. Conhecido em inglês como *randomized control trials* (testes controlados aleatórios, em tradução literal), esse modelo é considerado o “padrão ouro” das avaliações, por ser o que melhor consegue isolar os efeitos de uma variável, seja ela um novo medicamento, um programa ou uma política pública. A medicina é uma das áreas em que essa abordagem é mais comum. O primeiro estudo médico com essa metodologia publicado em revista científica data de 1948, quando o pesquisador britânico Austin Bradford Hill divulgou suas conclusões a respeito da eficácia do uso de um antibiótico no tratamento de tuberculose⁹.

Para testar um novo medicamento, é comum que um grupo de pessoas escolhido aleatoriamente receba o tratamento, enquanto outro grupo recebe um placebo. Várias recomendações de organizações de saúde, que salvaram milhões de vidas, surgiram de estudos feitos com essa metodologia. Desde o século passado, seu uso tem sido ampliado para outras áreas do conhecimento.

A opção de avaliar o impacto do Jovem de Futuro por meio desse tipo de experimento, devidamente adaptado ao contexto educacional, acabou sendo viável porque o Instituto Unibanco e os estados parceiros construíram uma relação pautada em confiança e transparência quanto aos resultados gerados. Todos precisavam estar convencidos de que esse era o melhor modelo a ser adotado, justamente por captar o efeito do programa com mais precisão. O fato de inicialmente apenas um grupo de escolas ser beneficiado poderia gerar o entendimento equivocado de que as demais acabariam prejudicadas. No entanto, o que se buscou foi justamente ter certeza de que o programa traria resultados positivos em cada contexto institucional e social antes de expandi-lo para toda a rede.

.....
⁹ Bothwell et al., 2016.

Métodos menos rigorosos de avaliação, como a simples comparação de indicadores antes e depois da intervenção, podem até gerar resultados mais vistosos para a opinião pública. Mas esses métodos são frágeis, pois não asseguram que o avanço – quando existente – foi causado especificamente pelo programa, já que uma série de outros fatores pode ter contribuído para a diferença observada.

Uma avaliação de impacto rigorosa é justamente aquela que consegue distinguir causa de correlação. Para entender a diferença entre essas duas dimensões, vale a pena recorrer a um exemplo hipotético. Suponhamos que uma prefeitura decida trocar todos os livros didáticos de suas escolas e que, quatro anos depois, os resultados dos alunos de toda a rede apresentem melhoria em testes de aprendizagem. Esses dois fatos – a melhoria do aprendizado e a introdução de novo material didático – estão certamente correlacionados, pois ocorreram ao mesmo tempo. A simples constatação dessa correlação, porém, não prova que uma seja a causa da outra.

Ao mesmo tempo em que os novos livros didáticos foram introduzidos, outras ações podem ter acontecido paralelamente e influenciado os resultados de aprendizagem medidos em testes. Por exemplo, pode ter havido diminuição do número de alunos por turma, migração de crianças da rede privada para a pública (ou vice-versa) ou mudança nos critérios de escolha de diretores ou de professores. Enfim, qualquer outra alteração simultânea pode ter impactado aquele resultado. Os novos livros didáticos podem inclusive ser piores do que os antigos, mas, devido a todos os demais fatores, mesmo assim os alunos melhoraram suas notas.

Nesse exemplo hipotético, como avaliar o impacto da introdução do novo material didático nas notas dos alunos? A melhor maneira de chegar o mais próximo possível dessa resposta é comparar dois grupos separados aleatoriamente. A separação aleatória garante que os grupos tenham perfil semelhante e se diferenciem apenas pelo fato de um deles ter recebido os livros novos. Se o grupo de estudantes que recebeu o novo material didático registrou melhoria nas notas enquanto o outro teve piora, estagnação ou mesmo avanço em menor escala, pode-se dizer que a introdução dos novos livros teve um impacto positivo naqueles alunos¹⁰.

Ou seja, mais do que uma simples correlação, a avaliação indica para o formulador de políticas públicas se aquela ação específica teve impacto positivo (ou não) no desempenho dos alunos, e também sinaliza a magnitude desse efeito.

.....

¹⁰ A constatação de impacto positivo pode ocorrer até mesmo num cenário de queda no desempenho. Se ambos os grupos comparados registram piora, mas o grupo beneficiado por aquela política específica tem queda menos intensa, uma avaliação rigorosa pode dizer que aquela política foi positiva, pois impediu que a piora fosse ainda mais intensa.

É isso que o Instituto Unibanco e os estados parceiros fazem ao avaliar rigorosamente o impacto do Jovem de Futuro. A explicação mais detalhada de como foi conduzida a avaliação experimental do Jovem de Futuro será abordada mais à frente. Antes, porém, será feita uma breve explicação de como funciona o programa e de seu histórico.

O QUE É O PROGRAMA?

O Jovem de Futuro teve início em 2008, e foi constantemente aperfeiçoado até chegar ao formato atual. Porém, permaneceram alguns pontos que sempre foram percebidos como importantes. Ao longo de mais de uma década de implementação, tem sido um programa voltado para fortalecer a competência de liderança e gestão de diretores de escola e coordenadores pedagógicos, apoiando-os com dados, indicadores, metas, processos, formação, assessoria e materiais variados. O objetivo sempre foi aumentar a permanência de todos os estudantes na escola e a taxa de conclusão do ensino médio, com níveis mais elevados de aprendizagem.

Antes de detalhar as mudanças pelas quais o programa passou, descreveremos o desenho atual do Jovem de Futuro. O programa parte do princípio de que os educadores e gestores em escolas públicas possuem enorme potencial. No entanto, muitas boas ideias não são disseminadas ou sequer chegam a virar realidade por falta de atenção em sua implementação. Como consequência, limita-se a capacidade de aprender com a prática, a partir de erros e acertos, e de disseminar esse conhecimento. O aprendizado com a experiência requer uma nova atitude analítica e mais colaboração entre pares e entre as instâncias envolvidas no processo – secretarias, regionais e escolas.

No apoio aos estados parceiros, e para viabilizar a transformação pretendida, são oferecidas ações, sempre gratuitas, de formação, assessoria técnica, governança, mobilização e avaliação.

Para estimular os gestores a refletirem sobre a sua própria prática e a organização de rotinas, o Jovem de Futuro trabalha com um método chamado Circuito de Gestão, inspirado, em sua origem, em outro método, conhecido como PDCA, sigla em inglês para planejamento, execu-

ção, monitoramento e replanejamento das ações¹¹.

O Circuito prevê uma sequência de ciclos compostos por seis etapas: pactuação de metas, planejamento, execução, avaliação de resultados, compartilhamento de práticas e correção de rota (replanejamento). Para gerar, entre outros resultados, alinhamento e coerência nas ações das três principais instâncias de uma rede estadual – a secretaria, suas regionais, e as escolas –, o Circuito roda de forma integrada nessas três instâncias.

1 Alinhamento dos 3 níveis administrativos

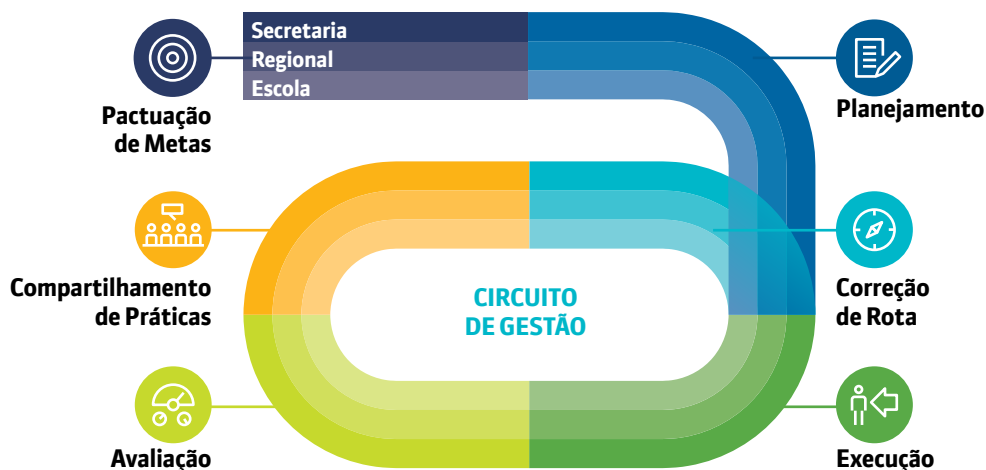


Figura 1 – Funcionamento do Circuito de Gestão nas redes estaduais

A primeira etapa prevista no Circuito de Gestão é a pactuação de metas. Esse é o momento em que a Secretaria de Educação, com assessoria do Instituto Unibanco, define a meta de melhoria educacional a ser atingida num ciclo de quatro anos, com base no Ideb, indicador elaborado pelo MEC a partir das médias dos alunos em testes de matemática e língua portuguesa e de suas taxas de aprovação. A meta para esse período de quatro anos – que corresponde ao ciclo político de uma gestão – é desdobrada em metas anuais, para garantir que a rede caminhe para aquele objetivo ao longo do processo.

Essa pactuação gera um efeito cascata: cada escola passa a ter compromisso com a sua própria meta. Aquelas que estão em situação de pior desempenho são consideradas prioritárias, o que significa que receberão maior atenção das regionais e da secretaria, para que consigam avançar

.....
¹¹ O PDCA (Plan, Do, Check, and Act) se tornou popular no pós-guerra, a partir da obra de W. Edwards Deming.

em ritmo mais rápido e assim diminuem a distância que as separa do restante da rede. Cada regional também tem sua própria meta, calculada a partir da média das metas das escolas sob sua coordenação.

Estabelecido um compromisso coletivo em relação a esses objetivos, todas as instâncias – secretaria, regionais e escolas – partem para a segunda etapa, que prevê o planejamento das ações que serão realizadas para alcançar os resultados propostos. Nesse momento, são elaborados planos de ações, documentos produzidos pelos gestores com o detalhamento das atividades que serão executadas, o período, os resultados esperados e os responsáveis por elas.

Nessas e em todas as demais etapas do Circuito, as equipes de regionais e secretarias recebem assistência técnica e formação do Instituto Unibanco e são guiadas por protocolos, documentos que detalham passo a passo os indicadores e as atividades que devem ser monitoradas. O uso de protocolos, porém, não significa que as soluções para os problemas sejam impostas às escolas. Pelo contrário. Como o objetivo é justamente gerar capacidade de aprender a partir da prática, são as equipes gestoras que, em cada unidade, e em diálogo com a comunidade escolar, fazem o seu diagnóstico das causas e das ações que serão executadas para combatê-los.

Numa determinada escola, por exemplo, a equipe gestora e a comunidade escolar podem concluir que um dos principais entraves para a melhoria da qualidade do ensino é o alto número de faltas dos estudantes. Diante disso, são pensadas, de forma coletiva, as ações para atacar aquele problema. Outras escolas fazem o mesmo exercício de forma simultânea, mas podem chegar a diagnósticos e soluções distintas.

Em redes de alta complexidade e com grande número de escolas, um risco a ser considerado é o de que as ações sejam feitas de forma descoordenada, sem acompanhamento ou apoio das instâncias acima das escolas. No Jovem de Futuro, esse apoio acontece principalmente na figura dos supervisores, que são técnicos lotados nas regionais de ensino, designados para visitar escolas com frequência quinzenal ou semanal, no caso daquelas consideradas prioritárias.

Tendo conhecimento dos problemas e das ações previstas em cada escola para melhorar a qualidade do ensino, as regionais também elaboram seus planos de ação, prevendo as atividades que serão executadas para dar apoio às unidades de ensino. Por fim, o mesmo é feito no âmbito da secretaria. É por isso que o Circuito de Gestão induz a rede a trabalhar de forma coerente, a partir do princípio da corresponsabilização, ou seja, da ideia de que, para alcançar os objetivos propostos, as escolas precisam de apoio das instâncias superiores.

Após a etapa de planejamento, o próximo passo é a execução das ações propostas. Ainda que exista uma etapa específica para a avaliação dos

resultados, isso não significa que as escolas deixam de ser acompanhadas nessa fase. Pelo contrário, a figura do supervisor continua sendo essencial, pois ele ajuda as escolas no monitoramento das ações e na identificação de pequenos ajustes que possam ser feitos de forma imediata. As próprias regionais, além da secretaria, também estão mobilizadas nesse momento para a execução das ações de sua responsabilidade.

A etapa seguinte do Circuito é a de avaliação, quando todo o sistema faz uma pausa para o balanço coletivo do que foi feito e dos resultados preliminares encontrados. A dinâmica é de cascata invertida, ou seja, a avaliação acontece primeiro nas escolas, depois nas regionais e, por último, nas secretarias. Por exemplo, uma escola que tenha identificado que o principal problema a ser atacado é o alto número de faltas dos estudantes verificará se o problema diminuiu e se já é possível ver algum avanço na aprendizagem devido à ação implementada. Essa avaliação é compartilhada com a regional, que, por sua vez, compartilha a avaliação geral das ações sob sua alçada com a secretaria.

Uma vez que um dos principais objetivos do Circuito de Gestão é gerar aprendizado a partir da experiência, uma etapa igualmente importante é a de compartilhamento de práticas. São momentos em que as equipes gestoras de todas as escolas se reúnem, na respectiva regional, para trocarem experiências a partir de seus erros e acertos na execução de suas ações. Novamente recorrendo ao exemplo hipotético da escola que identificou como problema o alto número de faltas dos estudantes: caso sua ação tenha sido bem-sucedida, ela será compartilhada com as demais nesse momento. Caso os resultados não tenham sido satisfatórios, será uma oportunidade de escutar de outras escolas que enfrentaram desafios semelhantes quais ações foram bem-sucedidas ou como foram implementadas em seus contextos. O mesmo é feito entre as regionais, com o objetivo principal de inspirar e ser inspirado por seus pares.

Por fim, o Circuito de Gestão prevê a correção de rotas, quando ocorre o replanejamento das ações a partir do aprendizado com os erros e acertos ao longo do processo. A partir da correção de rotas, o Circuito se reinicia, dando origem a um novo ciclo de execução, avaliação e compartilhamento de práticas, de modo que o avanço buscado seja contínuo, que as boas práticas não se percam por dificuldades de implementação e que sejam compartilhadas com um grupo maior, aumentando o repertório de ações de toda a rede em busca dos objetivos comuns.

AS TRÊS GERAÇÕES DO PROGRAMA

Da mesma forma que o Jovem de Futuro propõe às escolas a constante análise de suas ações em busca de aprendizado a partir da prática, o programa também foi aperfeiçoado ao longo do tempo até chegar ao

atual desenho. As mudanças mais significativas ocorreram em três momentos, chamados nos registros históricos do programa de “gerações”. Em uma década, foram colocadas em funcionamento três gerações: a primeira funcionou entre 2008 e 2015; a segunda entre 2012 e 2015; e a terceira a partir de 2015. Nesse período, houve alguns momentos de interseção, em que uma geração do programa conviveu com a seguinte, enquanto o processo de transição se fazia de forma planejada.

Na primeira geração, o programa teve formato piloto, com poucas escolas atendidas e muitas ações de apoio implementadas por colaboradores do próprio Instituto Unibanco. Participaram dessa geração 197 escolas em quatro estados: Minas Gerais, Rio Grande do Sul, Rio de Janeiro e São Paulo. Apesar da parceria firmada com a Secretaria Estadual de Educação, o trabalho era desenvolvido em contato direto com as escolas, que recebiam formação e assessoria técnica para a elaboração, a execução e o monitoramento de um plano estratégico de melhoria. Eram pactuadas metas ambiciosas para o aumento da aprendizagem e a redução dos percentuais de abandono e de alunos em padrão crítico de aprendizagem¹², acompanhadas por meio de avaliações externas e por dados coletados diretamente na escola. Essas metas eram desdobradas em sete resultados¹³, organizados em três eixos – estudantes, professores e gestão escolar – que deveriam orientar a elaboração do plano.

Para apoiar esse processo, colaboradores do Instituto faziam visitas semanais às escolas. Cada uma contava com a ajuda de um estagiário de pedagogia. Além disso, eram oferecidos recursos didáticos aos professores e ferramentas de fomento à participação comunitária, conhecidos, respectivamente, como metodologias pedagógicas e de mobilização. Para incentivar a adesão das escolas ao novo processo de gestão, havia transferência direta de recursos financeiros do Instituto para as escolas¹⁴. Esses recursos poderiam ser utilizados para premiar professores e alunos, melhorar a infraestrutura, capacitar profissionais, criar fundos de apoio a projetos pedagógicos ou atividades promovidas pelos estudantes. O sucesso dessa geração, comprovado por avaliação experimental, abriu portas para o ganho de escala e a busca por sustentabilidade das transformações, imprimindo nova marca à segunda geração do Jovem de Futuro.

.....

12 Os pontos de corte que definem os padrões ou níveis de aprendizagem foram definidos pelos próprios estados, de acordo com o desenho curricular e as expectativas locais de aprendizagem para cada série e etapa de ensino. Para a primeira geração, o banco de dados do Jovem de Futuro não conta com informações sobre as escolhas locais e, por essa razão, foram adotados os pontos de corte da escala criada em 2006 pela comissão técnica do movimento Todos pela Educação e pelo pesquisador José Francisco Soares (UFMG), ex-presidente do Inep.

13 Os sete resultados eram: aprendizagem de estudantes avaliados em testes de língua portuguesa e matemática; alunos com alto índice de frequência; desenvolvimento de habilidades socio-econômico-ambientais; diminuição do número de faltas dos professores; novas práticas pedagógicas; melhoria da infraestrutura da escola; e gestão orientada para resultados.

14 Cada unidade participante recebia uma transferência anual no valor de R\$ 100 por aluno matriculado no ensino médio regular.

2 A Evolução do Jovem de Futuro

OBJETIVOS

Garantir aprendizagem, permanência e conclusão do Ensino Médio



COMO FUNCIONA

Programa de gestão

Ciclos de mudança, execução e monitoramento dos planos, com replanejamento das ações

Foco no pedagógico

Direciona esforços para melhoria do ensino e aprendizagem

Metas de desempenho

Melhora da aprendizagem e da aprovação, com redução de desigualdades

Grupo gestor

Atuação coletiva envolvendo diretor, coordenador pedagógico, professores e estudantes

Supervisores

Atuam no assessoramento, formação e apoio do grupo gestor, além de monitorar ações

Ações e recursos

De mobilização, formação, assessoria técnica, governança* e sistemas de informação



A PRIMEIRA GERAÇÃO

Instituto Unibanco atua diretamente nas escolas. O objetivo, mantido em todas as gerações do programa, é tornar a gestão mais focada na aprendizagem e na permanência dos alunos de ensino médio, com redução das desigualdades.

Um dos mecanismos propostos para incentivar a adesão era a transferência direta de recursos financeiros do Instituto Unibanco para cada unidade escolar.

O **ciclo de mudança**, implementado na escola, era de **três anos**.

197 escolas

em

4 estados

(MG, RS, RJ e SP)





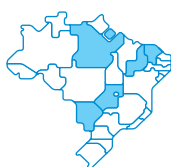
A SEGUNDA GERAÇÃO

Marca a mudança de um projeto em escolas para uma política da rede de ensino, implementada pela Secretaria de Educação, visando escala e sustentabilidade.

O Estado passa a ter um papel central, disponibilizando supervisores e técnicos da Secretaria. O Instituto Unibanco forma, apoia e monitora a implementação. O incentivo financeiro agora é garantido pelo programa federal Ensino Médio Inovador (ProEMI).

O **ciclo de mudança**, implementado na escola, passa a ser **anual**.

2.166 escolas
em
5 estados
(CE, GO, MS, PA e PI)



A TERCEIRA GERAÇÃO

Consolida a transformação do Programa numa política da rede de ensino.

É introduzida a **gestão para o avanço contínuo em educação**, que reforça o foco no estudante, na gestão pedagógica e o processo de aprendizagem pela prática. A atuação passa a ser sistêmica e a transferência financeira deixa de existir.

O **ciclo de mudança** implementado em escolas, regionais e no órgão central é reforçado e passa a ser **trimestral**.

3.549 escolas
em
7 estados
(ES, PI, GO, PA, CE, RN e MG)



Figura 2 – Evolução do Jovem de Futuro

A segunda geração nasceu com o desafio de transformar o Jovem de Futuro em uma política da rede pública parceira. Ao ser implementado por técnicos das secretarias de educação, o programa poderia ganhar maior escala, penetração no território e sustentabilidade no tempo. Assim, técnicos de apoio à gestão nas secretarias e supervisores escolares nas regionais de ensino passaram a receber formação e assessoria. Cada supervisor escolar ficou responsável por visitar mensalmente dez escolas em média, apoiando o ciclo de mudança. Esses atores se tornaram fundamentais, fazendo a ponte entre as escolas e as regionais de ensino, facilitando o fluxo de informações e aproximando as escolas da estratégia de gestão da rede. Em consequência, o número de escolas atendidas aumentou para 2.166, em cinco estados: Ceará, Goiás, Mato Grosso do Sul, Pará e Piauí (87% das matrículas públicas totais de ensino médio nesses estados).

O plano estratégico de melhoria feito pelas escolas passou por duas grandes transformações. A primeira delas veio da associação ao Programa Ensino Médio Inovador (ProEMI), do Ministério da Educação, que buscava inovações bem documentadas que pudessem ser oferecidas às escolas do Brasil. Havia proximidade entre a visão de gestão escolar por trás do Jovem de Futuro e a que o ProEMI propagava, de modo que a associação foi um passo coerente, que tornou o Jovem de Futuro parte de uma política pública federal. O ProEMI foi implantado em todo o território nacional entre 2009 e 2018. Oferecia incentivos financeiros às escolas¹⁵ e diretrizes para a reforma curricular do ensino médio¹⁶. A associação do Jovem de Futuro ao ProEMI priorizou os estados parceiros para a inclusão de escolas no programa federal. A aposta na conjugação de “melhoria na gestão” e “reforma curricular” se respaldou em evidências de impacto positivo e relevante na primeira geração, que atuou em ambas dimensões.

Na segunda geração, também foram introduzidos aprimoramentos no método de gestão levado às escolas. Foi nessa geração que apareceu a primeira versão do Circuito de Gestão, que lançou as bases da ideia de avanço contínuo, princípio que passou a ser perseguido mais explicitamente a partir da terceira geração.

Ainda havia muito a ser aprimorado no desenho do Jovem de Futuro. Era preciso aumentar o engajamento das escolas na transformação de suas práticas. Os processos burocráticos presentes na segunda geração (com a sobreposição de prestações de contas do Jovem de Futuro

.....

15 Era repassado um valor médio de R\$ 70 por aluno matriculado no ensino médio, de modo a viabilizar a implementação do *projeto de redesenho curricular (PRC)*.

16 O novo currículo deveria ser elaborado com a participação efetiva da comunidade escolar, estar fortemente alinhado com o projeto político-pedagógico da escola e respeitar os seguintes parâmetros: ampliação da carga horária mínima anual para 3 mil horas; garantia de ações interdisciplinares e elaboradas a partir das quatro áreas de conhecimento; e estímulo à participação dos estudantes no Exame Nacional do Ensino Médio (Enem). Além disso, a escola deveria respeitar os campos de ação pedagógica e curricular propostos pelo MEC, denominados macrocampos.

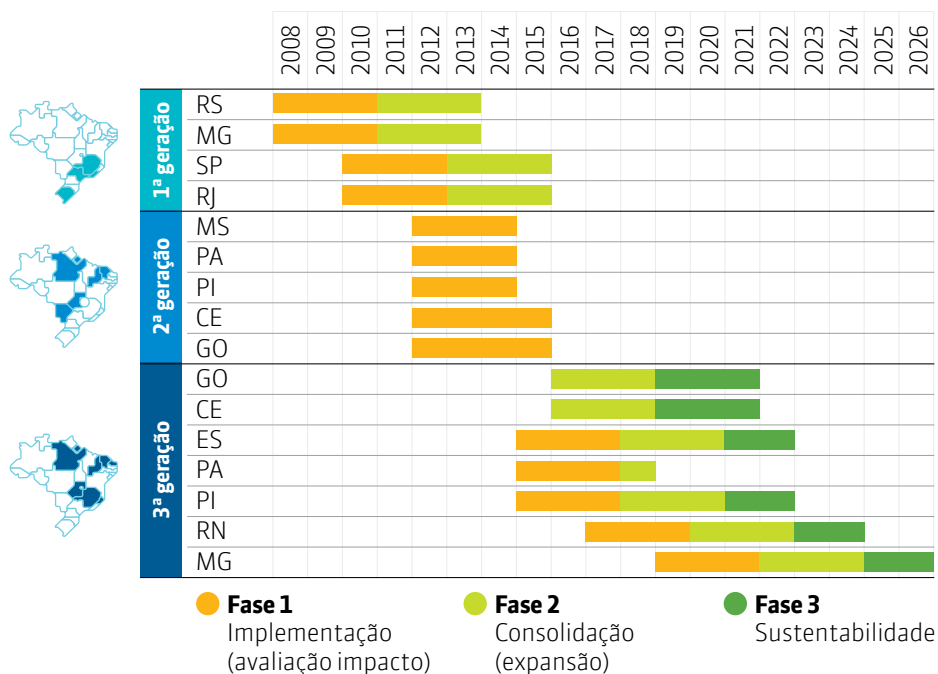
ro e do ProEMI) desperdiçavam energia das escolas e retiravam o foco central dos estudantes. Atrasos no repasse de recursos financeiros do ProEMI às escolas demonstravam que, apesar de esses recursos serem importantes, era possível pensar em uma solução que não dependesse deles.

A terceira geração nasceu, portanto, com o desafio de expandir o Circuito de Gestão para regionais e secretaria, alinhando os esforços de todos em planos integrados e coerentes com as reais necessidades das escolas. A estrutura administrativa da educação brasileira e a divisão de responsabilidades entre as instâncias tornam esse movimento ainda mais estratégico. Uma vez que as escolas têm autonomia relativamente limitada para tomar decisões ligadas à contratação e à formação de pessoas ou à melhoria de sua infraestrutura, a corresponsabilização é imprescindível.

A terceira geração acabou enfrentando os mesmos desafios de escala e sustentabilidade da segunda, buscando tornar o Jovem de Futuro uma política pública, absorvida pela estratégia de gestão da rede. Sem se apoiar em um programa federal, o Jovem de Futuro se concentrou em gestão e liderança, incluindo a dimensão pedagógica da gestão. A reforma curricular deixou de ser tratada pelas ferramentas do Programa. Essa especialização buscou alcançar patamares de qualidade em gestão educacional não observados nas gerações anteriores.

Por fim, é preciso considerar o fato de que, em educação, processos sustentáveis de mudança levam tempo para se consolidar, pois mexem com culturas e comportamentos arraigados. Por isso, o processo de mudança desencadeado pelo Jovem de Futuro e implementado em parceria com as redes considera um intervalo de oito anos de duração. Nos três anos iniciais, o foco é a disseminação e a experimentação por todos do novo método de gestão. Do quarto ao sexto ano de parceria, intensifica-se a transferência de conhecimento e de tecnologias, já iniciada na fase anterior, de modo que, ao final do período, a própria rede possa implementar as ações. Na última fase, que inclui os dois últimos anos de parceria, a função do Instituto Unibanco passa a ser monitorar a mudança cultural e apoiar processos de inovação em gestão propostos pelo parceiro.

3 Duração das 3 gerações do Jovem de Futuro



Fonte: Instituto Unibanco

Figura 3 – Duração das 3 gerações do Jovem de Futuro

PREPARANDO A AVALIAÇÃO DE IMPACTO DO JOVEM DE FUTURO

Com mais de dez anos de atuação em onze estados brasileiros, o Jovem de Futuro jamais perdeu o foco na gestão escolar. Perseguir metas, usar indicadores para tomar melhores decisões e seguir um ciclo bem definido de mudança são atividades continuamente aperfeiçoadas em colaboração com as redes parceiras. No entanto, as três gerações aconteceram em contextos e tempos distintos, e cada uma foi marcada por ações e formas de implementação próprias. Assim, é importante que a avaliação de impacto não apenas ofereça estimativas do efeito acumulado em uma década de existência do Jovem de Futuro, mas também indique mudanças relevantes no padrão de impacto entre as gerações¹⁷.

A escolha das medidas de resultado utilizadas na avaliação de impacto reflete os objetivos finais pretendidos: afetar a aprendizagem dos estudantes, sua permanência na escola e sua aprovação ao longo do ensino médio. Testes padronizados de língua portuguesa e matemática, organizados pelos próprios estados como parte dos sistemas internos de monitoramento, são utilizados como fontes de informação para medir os níveis de aprendizagem dos estudantes nessas disciplinas ao final do ensino médio (3ª série).

Além das médias em língua portuguesa e matemática, também é avaliado o impacto do programa na redução no percentual de alunos com notas muito baixas, que estão no chamado padrão crítico de aprendizagem. Afinal, as escolas precisam melhorar a aprendizagem de todos, sem deixar para trás aqueles com mais dificuldades. Já a permanência na escola e a conclusão dos estudos são inferidos a partir da taxa de aprovação em todas as séries do ensino médio. Se a aprovação cresce,

.....

¹⁷ Nesta publicação optou-se por analisar as gerações do programa em vez dos estados, uma vez que amostras maiores oferecem estimativas mais precisas. As flutuações no impacto, quando se comparam estados, são explicadas sobretudo por flutuações estatísticas. Não são, portanto, variações reais.

trata-se de um indicativo de que mais jovens estão seguindo adiante e concluindo seus estudos. A taxa de cada escola é disponibilizada anualmente ao público por meio do censo escolar, organizado pelo Inep, o instituto de avaliação e estatística do Ministério da Educação, em parceria com as redes de ensino.

Em termos conceituais, a avaliação de impacto do Jovem de Futuro não mede os efeitos do programa sobre uma coorte específica de estudantes (mesmo grupo acompanhado ao longo do tempo). Essa ressalva é importante na interpretação dos resultados, pois nas escolas é comum que as turmas que ingressam no 1º ano do ensino médio não sejam compostas exatamente pelos mesmos alunos dois anos depois, no 3º ano. Isso acontece porque pode haver, por exemplo, migração de alunos entre escolas, reprovações ou abandonos ao longo da trajetória dos estudantes. A evolução da escola, portanto, é acompanhada a partir da média dos alunos que, em cada ano da avaliação, estavam nela matriculados.

Vale lembrar ainda que, em geral, avaliações de impacto experimentais de programas compostos por múltiplas ações – caso do Jovem de Futuro – não permitem apontar qual ação foi mais relevante, pois isso envolveria um desenho mais sofisticado, com múltiplos tipos de tratamento, e amostra maior.

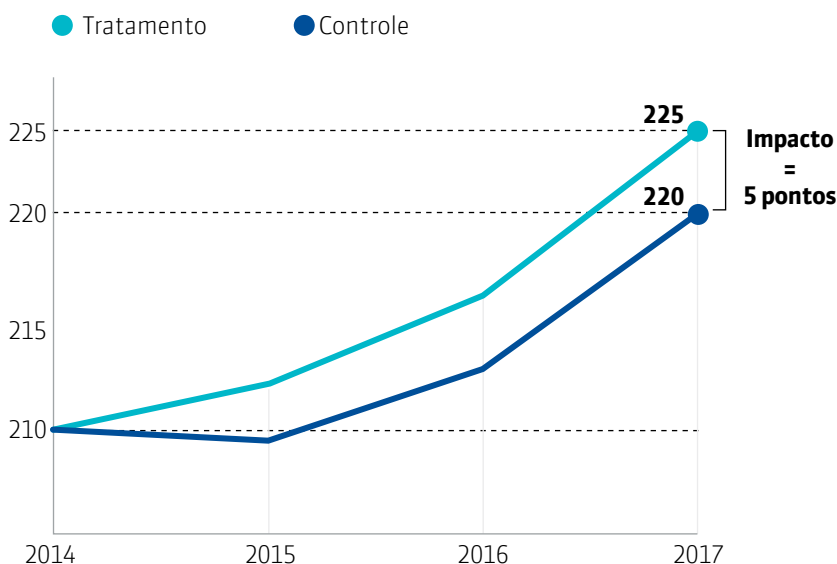
No caso do Jovem de Futuro, considerando a experiência de uma década de programa, a complexidade de sua implementação e o número crescente de escolas atendidas nos estados, há segurança para afirmar que a melhoria da gestão educacional teve impactos positivos. Mas não foi possível, até o momento, “abrir a caixa preta”, ou seja, saber qual das ações, entre as várias implementadas, teve maior contribuição para o resultado.

Mais recentemente, com a terceira geração, uma dificuldade adicional passou a influenciar a avaliação: o fato de que a atuação do Jovem de Futuro tornou-se sistêmica, integrando as ações de escolas, regionais e secretaria. Uma vez que o efeito final é sentido nas escolas, o que a avaliação realmente capta é a diferença entre ter ou não o Circuito de Gestão implementado. Mas o fato é que a presença do Circuito nas regionais e secretarias altera a política educacional como um todo, pois traz mais evidências ao planejamento e ferramentas para a boa gestão da administração central, beneficiando todas as escolas. Esse efeito adicional não é capturado pela atual avaliação, que não foi calibrada em seu desenho para incluir, além das escolas, efeitos das ações de amplo alcance das demais instâncias – secretaria e regionais. É possível, portanto, que os impactos medidos até este momento estejam subestimados. Veremos mais adiante que a entrada de Minas Gerais no programa, a partir de 2019, trouxe uma nova oportunidade para a geração de conhecimento científico sobre gestão educacional. A experiência nesse estado pode ajudar a abrir a caixa preta dos resultados atuais e a responder novas perguntas.

Além de mais confiável, outra vantagem das avaliações experimentais é simplificar consideravelmente o cálculo do impacto, tal como mostra o Gráfico 1. Voltemos ao exemplo hipotético citado na introdução deste livro: uma rede que introduziu novos livros didáticos em seu sistema. Suponhamos que o grupo de tratamento, aquele que recebeu os novos livros, tenha melhorado 15 pontos na nota de matemática segundo um teste padronizado¹⁸, passando de 210 para 225 pontos num intervalo de três anos. Essa melhora não necessariamente representa o impacto do programa. É preciso descontar desses pontos a evolução do controle, grupo de escolas com as mesmas características que não recebeu os livros. Digamos que o controle tenha melhorado 10 pontos. O impacto da política é medido pela diferença entre a melhora dos grupos de tratamento e a do controle, que neste exemplo é de 5 pontos.

Exemplo conceitual de impacto medido por avaliação experimental

Proficiência em matemática nas escolas tratadas (pontos na escala Saeb)



Fonte: Exemplo hipotético. Elaboração própria.

Gráfico 1 – Evolução das notas em matemática no caso hipotético

¹⁸ Os testes padronizados estaduais utilizam comumente a mesma escala do Sistema de Avaliação da Educação Básica (Saeb), que reúne um conjunto de avaliações externas em larga escala elaborado pelo Inep, com o intuito de oferecer um diagnóstico da educação básica brasileira. Para mais informações, consulte <<http://portal.inep.gov.br/educacao-basica/saeb>>. O sistema foi criado de forma a permitir a comparação das notas dos estudantes da educação básica no 5º e no 9º anos do fundamental e na 3ª série do médio, indicando assim a evolução da qualidade. Especificamente para o ensino médio, o intervalo de notas de matemática varia de 225 a 475 pontos. Esse intervalo, por sua vez, é dividido em nove faixas de proficiência, de aproximadamente 25 pontos cada.

Desde que os grupos de tratamento e controle se diferenciem apenas pelo fato de que um acessou o programa e o outro não, podemos estar seguros de que o impacto gerado é de 5 pontos. Conforme vimos, para criar grupos comparáveis, é preciso que as escolas sejam separadas através de um sorteio. No Jovem de Futuro, o sorteio é público e transparente. Todos recebem informações sobre o programa e a avaliação. Um protocolo é cumprido para evitar falhas. Considerando o período completo da parceria, nenhuma escola, nem mesmo as que estão no grupo de controle, é deixada fora do programa. Todas ingressam em algum momento. As escolas de tratamento começam logo no primeiro ano, enquanto as do controle ingressam no início do quarto ano.

Para o sorteio, as escolas são previamente reunidas em estratos, predominantemente em pares e trios, segundo critérios de semelhança¹⁹. Esse processo é importante para garantir que as escolas dos dois grupos (tratamento e controle) tenham características similares. Sem agrupar as escolas, o sorteio seria feito a partir do conjunto total, situação em que o acaso poderia, por exemplo, gerar um grupo tratado com um número proporcionalmente maior de escolas urbanas de grande porte, enquanto no controle poderia haver maior proporção de colégios rurais de pequeno porte. Fazendo a divisão por estratos, é possível garantir que, entre tratados e controles, haverá tanto escolas urbanas de grande porte como colégios rurais pequenos.

De fato, há três justificativas para fazer um sorteio para cada estrato de escolas. A primeira, como mencionado, é gerar balanceamento de escolas com diferentes perfis entre os grupos de tratados e de controle. Como consequência, há maior homogeneidade dentro dos estratos – ou seja, dentro do estrato as escolas têm perfil semelhante – e maior variância entre eles – entre os estratos, as escolas têm perfis diferentes, já que as redes de ensino têm grande diversidade. Esse procedimento aumenta a precisão das medidas de impacto.

Em segundo lugar, formar estratos de escolas semelhantes promove maior igualdade de oportunidades no ato do sorteio. Os estratos dão aos perfis escolares minoritários a oportunidade de serem tratados logo no primeiro ano de programa.

Por fim, essa estratégia ameniza alguns problemas comuns em avaliações de impacto, ligados ao fato de que, com o passar do tempo, é natural que a amostra perca algumas escolas. Depois do sorteio, a escola pode ser fechada devido ao planejamento físico das redes, ou deixar de participar dos testes estaduais que permitem apurar sua melhora interanual.

.....
¹⁹ Das 1.160 escolas que fazem parte da avaliação de impacto do Jovem de Futuro, cerca de 60% foram agrupadas em pares ou trios. Para maiores informações, ver Barros e Franco, 2018.

Esses são alguns dos exemplos mais frequentes de fatores geradores de perdas na amostra. Situações como essas são conhecidas pelos estatísticos como geradoras de atrito na amostra. Preocupam porque podem desequilibrar o perfil dos grupos de tratamento e controle. Tal desbalanço de perfil pode explicar a melhor performance das escolas tratadas. Desde que o motivo que tenha levado as escolas a sair da amostra não esteja ligado ao programa (ou seja, a saída aconteceria mesmo se o programa não estivesse na rede), é possível excluir da análise apenas as escolas que foram afetadas, junto com as demais que fazem parte de seu estrato. Com isso, a integridade do sorteio permanece válida²⁰.

Até o final de 2019, a avaliação de impacto do Jovem de Futuro considerou 380 estratos, abarcando 1.161 escolas em nove estados. Com um sorteio para cada estrato, foram realizados 380 sorteios.

20 Originalmente, foram formados 433 estratos para a avaliação de impacto – 84 na primeira geração, 101 na segunda e 248 na terceira. Na primeira geração, foram perdidos 31 estratos (37%), dos quais 25 eram do Rio Grande do Sul, deixados de fora da avaliação de impacto pelo fato de o estado não possuir avaliações próprias de língua portuguesa e matemática, fazendo com que provas aplicadas pelo próprio Instituto Unibanco fossem utilizadas como fonte. Os impactos encontrados se mostraram extremamente elevados. Algumas hipóteses indicavam viés nas escolas tratadas, que tinham motivação diferenciada para fazer a prova por temor de perder o programa, ou mesmo a possível disseminação dos itens que compunham as provas, pois também eram utilizados em provas diagnósticas que faziam parte do tratamento oferecido pelo programa. Assim, a equipe de avaliação resolveu ser conservadora e excluir esse superimpacto das estimativas acumuladas. Na primeira geração, havia um pacto com as escolas de que a falta de adesão ao programa geraria ruptura da parceria. Portanto, os outros seis estratos perdidos ocorreram por encerramento da parceria. Um dos casos se deveu ao fechamento de escola por decisão da rede. Na segunda geração, foram perdidos catorze estratos (14%), e na terceira apenas oito (3%), devido ao fechamento das escolas ou a sua não participação na avaliação estadual.

O PROGRAMA TEM IMPACTO²¹?

A maneira usualmente utilizada pelos avaliadores para responder à pergunta acima passa pela realização de um teste estatístico. O teste consiste em usar todos os estratos para investigar em quantos deles o desempenho das escolas tratadas (as que tiveram acesso ao programa logo no primeiro ano de implementação) foi melhor do que o das escolas de controle (incluídas no programa a partir do quarto ano de implementação). Mesmo que o programa tenha impacto, não necessariamente todas as escolas tratadas registrarão desempenho melhor do que as do grupo de controle. A questão é: em quantos estratos, no mínimo, é preciso observar que o grupo de tratamento se saiu melhor para que se afirme, com alguma segurança, que há impacto positivo²²?

Na avaliação do Jovem de Futuro, existem 380 estratos²³. Se o programa não tiver nenhum impacto, então a melhora do grupo de tratamento em relação ao de controle será um evento aleatório, isto é, em metade

.....

21 Este capítulo foi elaborado a partir de resultados e documentos produzidos por Ricardo Paes de Barros e sua equipe desde 2007. Uma síntese acadêmica do trabalho foi publicada no repositório de artigos do Banco Mundial, em 2018, e será publicada num artigo em fase de finalização. Para saber mais, veja Barros et al. (2018) e Barros et al. (no prelo).

22 Na realidade, esse raciocínio, típico da estatística, constrói um teste que avalia a hipótese nula, isto é, a de que o programa não tem impacto. O ideal é que as evidências disponíveis permitam rejeitar a hipótese nula. Em termos lógicos, rejeitar a hipótese de que o programa não tem impacto equivale a afirmar que ele tem impacto.

23 Embora Pará, Piauí, Goiás e Ceará tenham participado da segunda e da terceira gerações, os dados das escolas dos dois últimos estados foram utilizados exclusivamente para estimar o impacto da segunda geração. Na terceira geração, o Programa passou a ser implementado nesses estados com cobertura integral. No Piauí, optou-se por interromper a segunda geração antes de encerrada a avaliação de impacto. Por esse motivo, fez-se novo sorteio na terceira geração e a avaliação foi reiniciada. Portanto, os dados das escolas do Piauí são utilizados apenas para estimar o impacto da terceira geração. No Pará, os estratos da segunda geração seguiram até o final na implementação, e por isso foi possível utilizá-los na avaliação correspondente. No entanto, novo sorteio foi feito nesse estado na terceira geração. Por isso, o Pará aparece nas estimativas da segunda e da terceira gerações com estratos diferentes.

dos estratos (190), escolas de tratamento melhorarão mais do que as de controle, enquanto na outra metade ocorrerá o inverso. Desequilíbrios nessa partição “meio a meio” podem indicar impacto positivo ou negativo. Quantos casos acima ou abaixo de 190 estratos nos autorizarão a afirmar, com segurança, que o impacto é positivo ou negativo?

A resposta dependerá da margem de segurança (chamada pelos estatísticos de nível de significância) do teste estatístico definido pelo avaliador. Na avaliação do Jovem de Futuro, é testada a hipótese de que o programa não tem impacto e o resultado do teste indica se há evidências suficientes para aceitá-la ou rejeitá-la. O nível de significância definido indica quão confiante o avaliador está para aceitar ou rejeitar a hipótese em questão. Para um nível de significância de 10%, assume-se que a vantagem não terá vindo do programa em somente 10% das vezes em que for observada uma situação em que o número de estratos tratados se sai melhor. Em outras palavras, o resultado do teste é bastante confiável porque em apenas 10% dos casos se assumiria que o programa tem impacto, quando, na realidade, ele não tem. Com esse desenho de teste, para uma amostra de 380 estratos, é preciso que em mais de 207 deles as escolas tratadas tenham se saído melhor para que se possa afirmar que há impacto positivo. Da mesma forma, se as escolas tratadas se saírem melhor em menos de 174 estratos, então o programa tem impacto negativo²⁴.

O Gráfico 2 apresenta o balanço de estratos do grupo de tratamento com desempenho melhor do que os de controle, considerando os indicadores de resultados escolhidos para a aferição do impacto: notas médias em língua portuguesa e matemática ao final do ensino médio; percentual de alunos no padrão crítico de notas de língua portuguesa e matemática ao final do ensino médio; e taxa de aprovação em todas as séries do ensino médio.

.....

²⁴ Esses limites foram definidos a partir da hipótese de que, quando o programa não tem impacto, cada um dos 380 estratos equivale à situação de um lançamento de moeda, na qual há probabilidade de 50% de chances de a escola tratada ter desempenho melhor que o da escola de controle e vice-versa, dando à distribuição um formato binomial.

Onde o Jovem de Futuro teve maior impacto

Distribuição do número de estratos em que o tratamento evoluiu mais do que o controle das escolas no Programa Jovem de Futuro - 1ª, 2ª, 3ª Geração

Cauda inferior

(estratos 0 a 174)

Maioria de estratos de escolas nessas áreas indicaria impacto negativo do programa

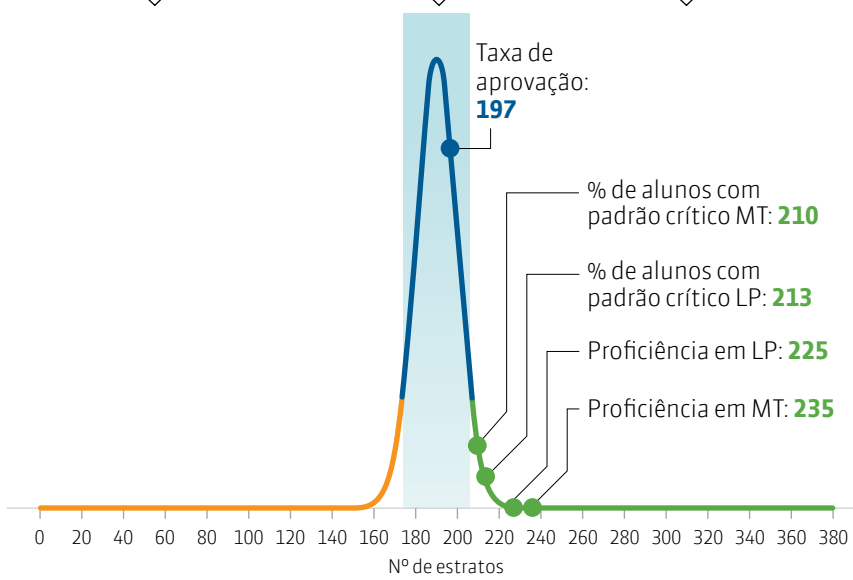
Área de incerteza de resultados

(estratos 175 a 206)

Cauda superior

(estratos 207 a 381)

O fato de haver mais estratos de escolas nessa área indica que o programa tem impacto



No caso da **Taxa de Aprovação**, por exemplo, o resultado está dentro da área de **incerteza** estatística



Já no caso dos aprendizados em **língua portuguesa e matemática e no percentual de alunos no padrão crítico**, podemos ter certeza que há **impacto positivo**

Fonte: elaborado a partir do artigo de Barros et al. (2018)

Gráfico 2 – Distribuição do número de estratos nas três gerações do Programa

Para as notas de língua portuguesa e matemática (proficiências médias), o número de estratos em que as escolas tratadas tiveram desempenho melhor do que as de controle esteve sempre acima do corte de 207, tal como exigido pelo nível de significância estatística do teste. Isso também ocorre para o percentual de alunos nos padrões críticos de aprendizagem (ambas as disciplinas). Já para a taxa de aprovação, detectou-se impacto positivo em um número menor de estratos, abaixo do ponto que marca a zona de certeza.

Considerando a área verde do Gráfico 2, é baixa a chance de se observarem tantos estratos com impacto positivo não decorrente do programa. Por exemplo, para o percentual de alunos no padrão crítico de aprendizagem em matemática, essa chance é aproximadamente de 1 a cada 2 mil casos²⁵. Afirmar que o programa tem impacto, porém, é apenas parte da resposta. Resta saber o tamanho dessa contribuição.

.....

25 P-valores são parâmetros estatísticos para certificar que os resultados encontrados nos testes não são aleatórios. Indicam a probabilidade de se observar uma situação que parece ser de impacto positivo sem que o efeito medido tenha sido gerado pelo programa. Quanto mais baixo o p-valor, menor a probabilidade de esse “engano” ocorrer. Na avaliação do Jovem de Futuro, escolheu-se um corte de 10% para o p-valor. Então, as estimativas com p-valor superior a 10% não serão consideradas significativas. Para os testes que buscam comprovar a existência do impacto, os p-valores referentes às estimativas para a nota de língua portuguesa, a nota de matemática e o percentual de alunos no padrão crítico de aprendizagem em matemática (estimativas na zona verde do Gráfico 2) são de, respectivamente, 0,8%, 0,4% e 0,0%. A probabilidade de se observar melhor performance das escolas tratadas e isso não ser um impacto do programa é de, respectivamente, 1 em 128 casos, 1 em 282 casos e 1 em 2.141 casos.

MAGNITUDE DO IMPACTO

A avaliação experimental simplifica consideravelmente o cálculo da magnitude do impacto. De volta aos 380 estratos: para cada um deles, basta tomar a diferença entre a evolução do grupo de tratamento e a de seu respectivo controle. A evolução do tratamento é a diferença entre os resultados de “antes e depois” do início do programa, da qual é preciso subtrair a diferença entre “antes e depois” do controle para que se chegue ao impacto. Esse método é conhecido como “diferença em diferença”. O impacto final é dado pela simples média das 380 “diferenças em diferenças”.

A Tabela 1 (na página seguinte) mostra as estimativas dos impactos e sua margem de erro, bem como o p-valor, a medida estatística usada para saber se os resultados obtidos resultam de flutuações aleatórias. Quanto mais o p-valor se aproxima de zero, menores são as chances de atribuir impacto ao Programa indevidamente.

Praticamente todas as estimativas de impacto do Jovem de Futuro têm p-valores muito baixos, a maioria perto de zero, repetindo o que já vimos na seção anterior: seria muito difícil observar uma situação em que as escolas tratadas melhoraram sem influência do Programa. A chance de ver efeitos dessa magnitude ou ainda maiores sem que a vantagem venha do Jovem de Futuro é de 1 em 1.000 ou 2.000 para as notas de língua portuguesa e matemática. Para a taxa de aprovação, observam-se com mais frequência situações que parecem resultantes do Programa, mas não são (1 a cada 25 casos).

Nas escolas atendidas pelo Jovem de Futuro, o impacto na nota média de língua portuguesa foi de 4,4 pontos, o que equivale a 31% da melhora global das escolas tratadas. Em matemática, o impacto foi de 4,8 pontos, respondendo por 87% da evolução global das escolas tratadas. A margem de erro é de até 1,6 ponto para cima ou para baixo.

Estimativas de impacto do Jovem de Futuro – 1ª, 2ª e 3ª gerações

Variáveis		Impacto ao final de 3 anos de JF		
		Impacto (pontos na escala)	Margem de erro	P-valor (%)
Proficiência	Língua Portuguesa	4,4	1,6	0,0
	Matemática	4,8	1,3	0,0
Taxa de Aprovação		1,4	1,1	4,1
% alunos no padrão crítico de proficiência	Língua Portuguesa	-1,0	0,3	0,0
	Matemática	-4,2	1,1	0,0
% alunos nos padrões adequado e avançado de proficiência	Língua Portuguesa	2,9	0,8	0,0
	Matemática	0,9	0,5	1,0

Fonte: elaborado a partir do artigo de Barros et al. (2018)

Tabela 1

Com relação ao fluxo escolar, as escolas tratadas aumentaram a taxa de aprovação no ensino médio em 1,4 ponto percentual em relação às escolas de controle. Contudo, a margem de erro é alta, quase da mesma magnitude que a estimativa de impacto – 1,1 ponto para cima ou para baixo –, deixando dúvidas sobre o real efeito do programa quando se considera o conjunto das três gerações.

Como já mencionado, além do aprendizado dos alunos (medido pelos testes de português e matemática) e das taxas de aprovação (calculada a partir dos registros do Censo Escolar), o Jovem de Futuro procura incidir também na diminuição do percentual de alunos com aprendizado em nível crítico. A avaliação do programa mostra que são significativos os efeitos sobre a redução dessa percentagem. Em matemática, as escolas tratadas reduziram a fração de alunos nesse padrão em 4,2 pontos em relação às escolas de controle, o que corresponde a treze vezes a redução experimentada pelas escolas tratadas. A margem de erro é de apenas 1,1 ponto para cima ou para baixo. Além disso, houve impacto positivo entre os melhores alunos das escolas

(com aprendizagens em nível adequado ou avançado). Em língua portuguesa, mais uma vez, os efeitos foram positivos, embora mais modestos²⁶.

RELEVÂNCIA DO IMPACTO NO CONTEXTO DO HISTÓRICO BRASILEIRO

A qualidade do ensino médio estadual brasileiro está estagnada. Durante os três anos obrigatórios do curso, aprende-se pouco em comparação ao que já se sabia no final do ensino fundamental. Tipicamente, um estudante que concluiu o ensino médio estadual em 2017 expandiu seu conhecimento de língua portuguesa em 17 pontos e em 13 pontos no de matemática, na escala Saeb²⁷.

Nesse cenário, os quase 5 pontos de impacto do Jovem de Futuro correspondem a 26% do que um estudante aprende tipicamente em língua portuguesa durante todo o ensino médio e a 37% do que aprende em matemática. Portanto, integrar o Programa faz com que se aprenda mais matemática tanto quanto se mais uma série fosse adicionada ao ensino médio no atual cenário, um resultado bastante significativo.

A baixa performance do ensino médio brasileiro persiste há anos. De acordo com o Gráfico 3, desde 2013, a nota média de matemática é aproximadamente 260 pontos, a mesma de 2005. Em língua portuguesa, a evolução chega a pouco mais de 10 pontos, menos da metade do que o ensino fundamental evoluiu. Ano após ano, o ensino fundamental melhora, mas o ensino médio não, o que torna o impacto do Jovem de Futuro ainda mais relevante.

²⁶ Em um artigo que analisa o impacto da adoção de boas práticas de gestão em escolas públicas norte-americanas, o economista Roland Fryer, de Harvard, destaca que outros estudos que medem o impacto de diversas intervenções voltadas à melhoria da aprendizagem também encontram resultado superior em matemática. O autor cita duas hipóteses para esse fenômeno. Em uma delas, de acordo com pesquisas em psicologia do desenvolvimento, o período crítico para o desenvolvimento da linguagem ocorreria na infância. Para funções cognitivas mais complexas, o período se estenderia até a adolescência. Na outra, o aprendizado em leitura seria influenciado pelo contexto externo à sala de aula – em especial o familiar e comunitário –, o que poderia explicar porque estudantes que não falam inglês formal (*nonstandard English*) em casa ou na sua comunidade têm desempenho inferior. Ambas explicariam a estimativa de impacto maior para matemática do que para linguagem. Ver Fryer, 2014.

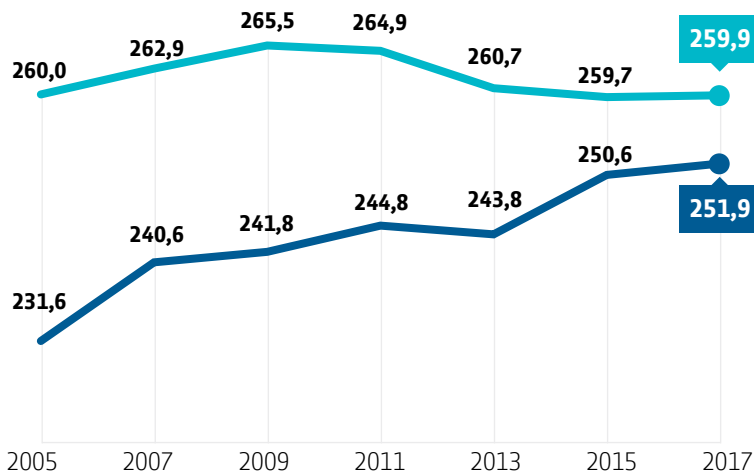
²⁷ Considerou-se que aqueles que concluíram a 3ª série do ensino médio estadual em 2017 estavam cursando o 9º ano do ensino fundamental em 2014. Uma vez que as provas do Saeb são aplicadas em anos ímpares, tomamos a média dos resultados do 9º ano público em 2013 e 2015.

Evolução da proficiência da rede estadual de ensino médio e ensino fundamental (anos finais) público – Brasil – escala Saeb

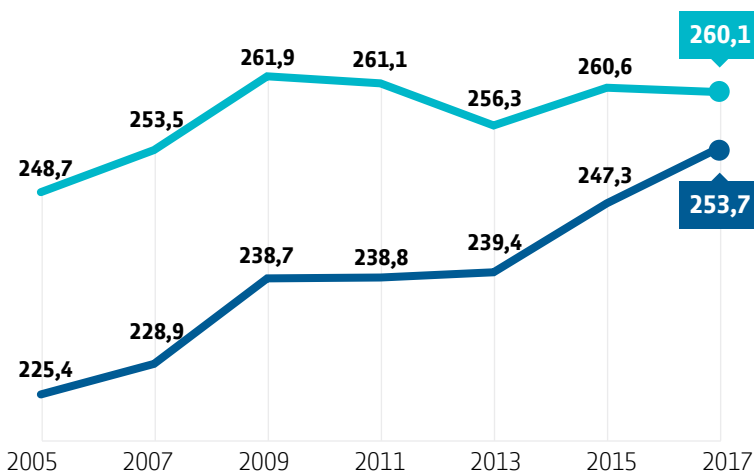
Brasil - escala Saeb

● Ensino Fundamental (anos finais) ● Ensino Médio

1² Matemática



A^B Língua Portuguesa



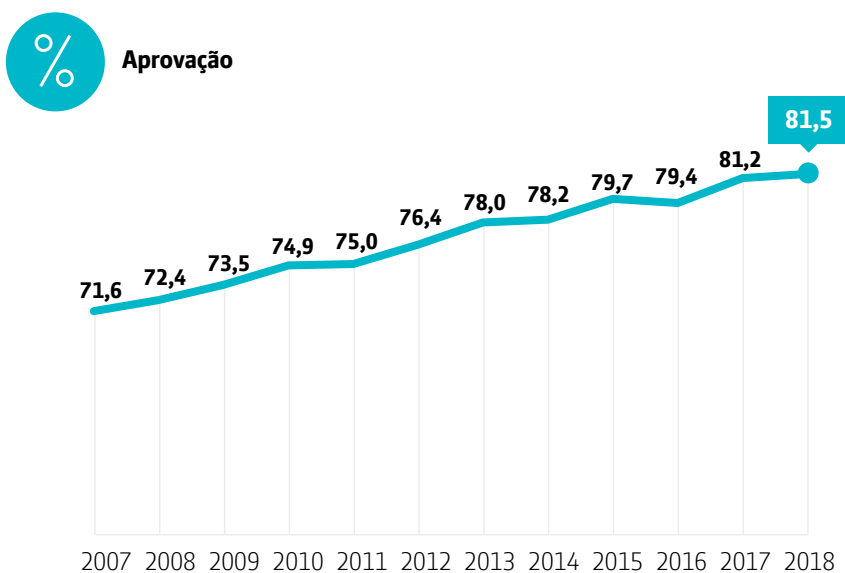
Fonte: Saeb

Gráfico 3 – As notas no Saeb e sua evolução em 2005-17

Com relação à taxa de aprovação, desde 2007 há evolução bastante significativa nos resultados do ensino médio estadual. O país melhorou quase 10 pontos no período ou 0,9 ponto ao ano. Como o impacto do Jovem de Futuro considera três anos de intervenção, três anos de melhora “natural” nessa década gerariam 2,7 pontos de aumento na taxa de aprovação. Assim, com 1,4 ponto adicional de impacto (que foi o verificado no Jovem de Futuro) na taxa de aprovação brasileira, o ritmo de melhoria dessa taxa aumentaria em 50% no período.

■ Evolução da taxa de aprovação da rede estadual de ensino médio

Em %



Fonte: INEP

Gráfico 4 – Aprovação do ensino médio em 2007-18

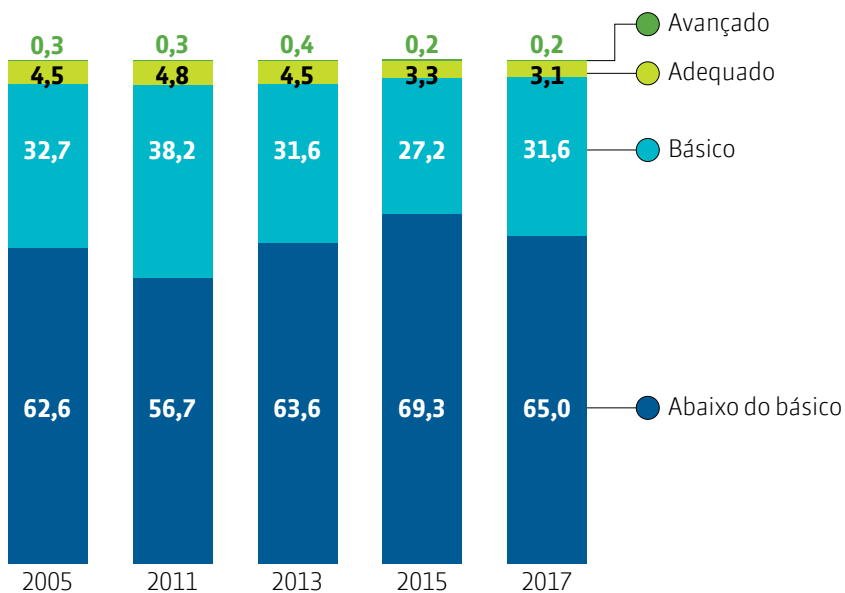
Por fim, é muito grave o histórico brasileiro da concentração de alunos no padrão mais crítico²⁸ de aprendizagem. De 2005 a 2017, o Brasil piorou nesse quesito. O percentual de alunos com nota de matemá-

²⁸ Por nível mais crítico, entende-se o percentual dos alunos que estão com aprendizado abaixo de 275 pontos em matemática e de 250 pontos em língua portuguesa, na escala do Saeb. Esse recorte é o mesmo utilizado pelo pesquisador Francisco Soares, ex-presidente do Inep, formulador do Saresp (Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo), e um dos formuladores da nota técnica do Movimento Todos Pela Educação que estabeleceu os níveis considerados adequados e inadequados para a aprendizagem.

tica nesse padrão crítico foi de 62,6% para 65,0%. Ou seja, enquanto o cenário nacional preocupa por conta do patamar de aprendizagem dos estudantes, baixo e desigual, o programa Jovem de Futuro mostrou ser capaz de ir na direção contrária, ou seja, diminuir o percentual de jovens nos patamares mais baixos de desempenho.

■ Evolução da porcentagem de alunos por padrão de desempenho em matemática

Em %

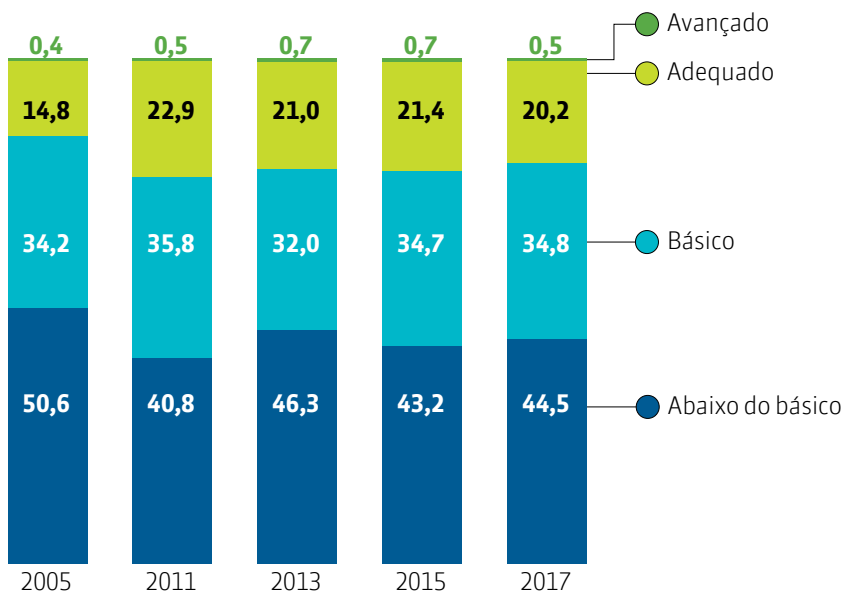


Fonte: elaboração própria a partir dos resultados do Saeb. Os 4 níveis foram estabelecidos com base na interpretação da escala realizada por Francisco Soares.

Gráfico 5 – Distribuição dos padrões de desempenho em matemática, 2005-17

■ Evolução da porcentagem de alunos por padrão de desempenho em língua portuguesa

Em %



Fonte: elaboração própria a partir dos resultados do Saeb. Os 4 níveis foram estabelecidos com base na interpretação da escala realizada por Francisco Soares.

Gráfico 6 – Distribuição dos padrões de desempenho em língua portuguesa, 2005-17

COMPARAÇÕES INTERNACIONAIS

Para saber se o impacto do Jovem de Futuro está em linha com o que foi encontrado em avaliações sobre gestão escolar em outros países, é preciso garantir que as medidas de resultados sejam comparáveis. Um aumento de 5 pontos no aprendizado em matemática na escala do Saeb não necessariamente representa o mesmo numa escala internacional. A saída está em converter os valores em porcentagem do desvio padrão, medida utilizada pelos estatísticos para mensurar a dispersão de uma distribuição em relação a sua média.

As notas de língua portuguesa e matemática utilizadas para avaliar o Jovem de Futuro estão na mesma escala de proficiência do Saeb, que é o exame nacional brasileiro. Por essa razão, os impactos do programa são expressos como porcentagem do desvio padrão da distribuição nacional dada pelo Saeb. Adotamos 2017 como ano de referência, por ser o dado mais recente disponível. Os desvios-padrão obtidos em relação à referência do Brasil são de 50 pontos²⁹. Consequentemente, o impacto do Jovem de Futuro, expresso como porcentagem do desvio padrão, é da ordem de 10%.

Os exemplos citados tratam principalmente de avaliações experimentais que analisaram iniciativas no âmbito da gestão ou de outras variáveis internas à escola com reconhecido impacto no aprendizado dos estudantes. Uma pesquisa mais aprofundada sobre outros estudos rigorosos feitos com metodologia semelhante está em curso. Uma limitação deste exercício é que os artigos científicos trazem poucas informações sobre o desenho das intervenções avaliadas, dificultando a interpretação da comparabilidade dos efeitos.

Com relação ao impacto da boa gestão escolar, medido a partir de avaliação experimental, os trabalhos de Roland Fryer, do Departamento de Economia de Harvard, são amplamente reconhecidos. Em 2017, o economista demonstrou que os estudantes de escolas cujo diretor passou por um bom programa de formação em gestão tiveram aprendizado maior do que os das escolas de controle em diversas disciplinas, como inglês, artes, matemática, ciências sociais e ciências naturais, tendo esse efeito chegado a 10% do desvio padrão³⁰. O programa em questão oferecia ao gestor escolar 300 horas de treinamento em planejamento escolar, durante dois anos, com instrução baseada em dados, observação de aulas e coaching. Num estudo experimental anterior, de

.....
²⁹ Foi considerada a distribuição de notas de língua portuguesa e matemática para o ensino médio, na rede pública.

³⁰ A avaliação experimental foi conduzida em escolas públicas de Houston, nos Estados Unidos. Os efeitos reportados utilizando *high-stakes test scores* (provas com consequências de *accountability*) são de 10% do desvio padrão, e os efeitos reportados utilizando *low-stakes test scores* (provas sem consequências de *accountability*) são de quase 20% a mais no desempenho. Para maiores detalhes, ver Fryer, 2017.

2014, Fryer comparou o desempenho de um grupo de escolas públicas escolhido aleatoriamente para receber assistência e adotar boas práticas em gestão, já difundidas em *charter schools*³¹, tais como tutoria, aumento de tempo em sala de aula dedicado ao ensino e disseminação de altas expectativas. O impacto foi de 15% do desvio padrão dos aprendizados de matemática, um efeito estatisticamente significativo. Já em leitura, o efeito foi estatisticamente nulo³².

Outra referência importante é a meta-análise publicada em 2007 por Carolyn Hill, do Georgetown Public Policy Institute, e coautores, em parceria com a MDRC (Manpower Demonstration Research Corporation). O trabalho reúne avaliações de intervenções educacionais voltadas à melhoria do aprendizado de jovens em diversas frentes. Considerando apenas os experimentos que envolvem o ensino secundário, o impacto médio encontrado em 43 avaliações foi de 27% do desvio padrão. Embora a publicação não entre no mérito sobre o tipo de intervenção que teve maior impacto na aprendizagem dos estudantes, há evidências de que ações voltadas aos professores são as mais efetivas, tal como amplamente estudado por Eric Hanushek, da Universidade de Stanford. Segundo ele, a qualidade do professor é o fator com maior capacidade de impactar a aprendizagem, com efeitos da ordem de 10 a 20% do desvio padrão ao ano. Ou seja, considerando que o impacto do Jovem de Futuro ao longo de três anos é da ordem de 10%, o impacto do professor (10% do desvio padrão a cada ano) equivaleria a três vezes o impacto do Jovem de Futuro³³.

De maneira geral, mais de um terço das avaliações experimentais incluídas na meta-análise organizada por Hill et al. (2007) apresentou impactos inferiores ao do Jovem de Futuro. De acordo com a meta-análise liderada pelos economistas brasileiros Paes de Barros e Portela³⁴, intervenções com efeitos de magnitude semelhante à do Jovem de Futuro correspondem a reduzir o tamanho da sala de aula ou a substituir os professores por outros mais experientes.

Portanto, o impacto do Jovem de Futuro tem magnitude similar à observada em programas internacionais focados na gestão: um efeito considerável, porém mais baixo do que o trabalho direto com os professores.

.....

31 *Charter schools* são escolas privadas ou comunitárias que recebem recursos públicos e atendem gratuitamente os estudantes. O universo avaliado contrapõe escolas públicas de Houston que adotaram práticas difundidas em *charter schools* a escolas públicas que não adotaram essas práticas. Nenhuma escola no experimento se tornou *charter*.

32 Esse estudo também compreende escolas públicas de Houston, nos Estados Unidos, com universo diferente do anterior. Para mais detalhes, ver Fryer, 2014.

33 Ver, por exemplo, Hanushek e Rivkin (2006, 2010), que mostram efeitos de diversos estudos sobre o tema.

34 Caminhos para melhorar o aprendizado. Disponível em: <www.paramelhoraroaprendizado.org.br>. Acesso em: 8 maio 2020.

Avaliações não experimentais (feitas com metodologia diferente da do Jovem de Futuro) oferecem resultados não comparáveis aos do Programa, mas é importante ver a tendência que elas indicam. Em 2015, o professor Nicholas Bloom, atualmente na Universidade de Stanford, investigou com outros coautores a associação entre as boas práticas de gestão dos diretores escolares e o aprendizado em matemática dos estudantes. Seu universo de análise incluiu mais de 1,8 mil escolas de ensino médio em oito países do mundo, inclusive o Brasil³⁵. O aumento do desvio padrão na qualidade das práticas de gestão dos diretores associou-se a um aumento de 24% do desvio padrão no aprendizado dos estudantes. Especificamente no Brasil, entretanto, a relação é menos intensa: o aumento do desvio padrão na qualidade das práticas de gestão dos diretores está associado ao crescimento de 10% do desvio padrão no aprendizado dos estudantes.




Em suma, a gestão é importante para transformar a educação. E o Jovem de Futuro trouxe uma relevante contribuição para o cenário brasileiro, não só em termos de resultados, mas também para a geração de conhecimento científico em educação. O impacto maior de determinada iniciativa não significa, contudo, que apenas um tipo de intervenção deva ser considerado. Em geral, reformas mais amplas que resultaram em melhorias sustentáveis procuraram melhorar as condições de ensino e aprendizagem em múltiplas frentes. Cada política tem a sua importância.

.....
³⁵ Os países integrantes do estudo são Reino Unido, Suécia, Canadá, Estados Unidos, Alemanha, Itália, Brasil e Índia. Ver Bloom et al., 2015.

O QUE MUDOU NO IMPACTO AO LONGO DAS GERAÇÕES?

Como vimos, a primeira geração do Jovem de Futuro também é chamada de “piloto”. Sua pequena cobertura foi suficiente para aperfeiçoar a implementação e gerar estimativas de impacto capazes de influenciar a decisão sobre a expansão do Programa. Na segunda geração, o Jovem de Futuro deixou de ser uma experiência piloto e passou a ser concebido como uma estratégia da rede de ensino, com protagonismo do estado na implementação e ampla cobertura, conforme mostra a Tabela 2. A cobertura saiu de menos de 10% das escolas e da mesma porcentagem de matrículas, na primeira geração, para patamares entre 70% e 94% entre os estados parceiros na segunda geração. Na terceira, o atendimento foi praticamente universalizado.

Cobertura máxima atingida nos estados parceiros do Jovem de Futuro que concluíram a avaliação de impacto até 2018

	Estados parceiros	Número de escolas	Percentual de escolas	Percentual de matrículas
Primeira Geração 	Minas Gerais	44	2	6
	Rio Grande do Sul	46	4	8
	Rio de Janeiro	30	3	8
	São Paulo	77	2	3
	Total*	197	2	5
Segunda Geração 	Ceará	439	69	76
	Goiás	580	93	93
	Mato G. do Sul	271	88	88
	Pará	458	86	93
	Piauí	418	88	90
	Total*	2.166	84	87
Terceira Geração 	Espírito Santo	237	91	95
	Piauí	451	91	90
	Goiás	590	100	96
	Pará	203	33	33
	Ceará	640	99	99
	Total*	2.121	68	70

* O percentual de escolas e matrículas atendidas diz respeito, respectivamente, ao total de escolas e matrículas atendidas pelo JF sobre o total de escolas e matrículas de EM dos estados atendidos.

Nota: na terceira geração, não foram incluídos os estados de Minas Gerais e Rio Grande do Norte porque o primeiro só iniciou a parceria em 2019 e o RN estava no período da janela de avaliação.

Fonte: elaboração própria

Tabela 2

Apesar da extraordinária ampliação da taxa de cobertura entre a primeira e a segunda geração, e da transferência para as secretarias de educação de uma parte relevante da implementação, os impactos se mantiveram mais ou menos no mesmo padrão, conforme indica a Tabela 3. Já na terceira geração, o efeito sobre as notas médias de língua portuguesa e matemática se reduziu um pouco, enquanto, pela primeira vez, o programa passou a ter impacto na taxa de aprovação do ensino médio.

X Impacto nas proficiências e taxa de aprovação por geração

Impacto ao final de 3 anos de JF em:

Gerações do JF	AB			12			%		
	Língua portuguesa			Matemática			Taxa de aprovação		
	Impacto	P-valor (%)	Margem de erro entre gerações	Impacto	P-valor (%)	Margem de erro entre gerações	Impacto	P-valor (%)	Margem de erro entre gerações
Média	4,4	0,0	-	4,8	0,0	-	1,4	4,1	-
1ª Geração	5,5	3,8		5,3	2,7		1,1	33,5	
2ª Geração	5,6	1,3	4,4	5,8	0,7	3,6	-0,8	25,9	4,1
3ª Geração	3,1	5,5	3,5	3,7	2,9	3,0	2,8	2,9	2,5

Fonte: elaborado a partir do artigo de Barros et al. (2018)

Tabela 3

Existem evidências de que as flutuações do impacto sobre as notas da segunda para a terceira geração não são relevantes, pois estão dentro das margens de erro da estimativa³⁶. Contudo, as diferenças de impacto na taxa de aprovação nos fazem crer que o programa mudou de padrão a partir da terceira geração. O impacto sobre a taxa de aprovação aumentou 3,6 pontos, sendo que a margem de erro da estimativa é de aproximadamente 2,5 pontos para cima ou para baixo.

O fato de a redução do impacto nas notas estar dentro da margem de erro nos coloca em dúvida sobre a existência, de fato, de uma redução. É possível que o aumento do impacto na taxa de aprovação do ensino médio venha acompanhado de redução na nota dos alunos que chegam à 3ª série. Afinal, se a taxa de aprovação em cada uma das séries do ensino médio aumenta, isso significa que os estudantes com perfil mais vulnerável, que antes não costumavam ser aprovados, agora passam de ano. Mesmo considerando que suas aprendizagens foram avaliadas pelas escolas como suficientes para a aprovação, a 3ª série do ensino médio passa a ser mais heterogênea, com alunos de diferentes

³⁶ O erro padrão da diferença de impacto entre a segunda e a terceira geração é dado pela raiz quadrada da soma dos quadrados dos erros da estimativa de impacto de cada geração. Para uma significância estatística de 0,10, esse erro padrão da diferença foi multiplicado por 1,65, em acordo com a tabela normal padrão.

bagagens, e isso também pode reduzir as notas das escolas tratadas na comparação com as notas passadas³⁷. É interessante notar que o maior efeito observado na taxa de aprovação ocorre na 1ª série, que é justamente aquela em que há maior abandono e evasão escolar.

A Tabela 4 mostra que, na terceira geração do Jovem de Futuro, as escolas tratadas melhoram sua taxa de aprovação na 1ª série do ensino médio em 3,6 p.p. em relação às de controle. Na 2ª série, o impacto seguiu relevante, de 2,1 p.p. Os impactos declinam ao longo das séries, e na 3ª série deixam de ser relevantes quando consideramos 10% de significância estatística. Portanto, o novo padrão de impacto a partir da terceira geração parece ir na direção de aumentar a inclusão social, com redução nas desigualdades educacionais.

Impacto na taxa de aprovação da 3ª geração por série

Em pontos percentuais

Série	Impacto médio	Margem de erro (em pontos)	P-valor (%)
1ª série do EM	3,6	2,1	4
2ª série do EM	2,1	1,9	8
3ª série do EM	1,0	1,5	18

Fonte: elaborado a partir do artigo de Barros et al. (2018)

Tabela 4

Além das evidências sobre as taxas de aprovação, um olhar cuidadoso para o comportamento do impacto sobre a diminuição do contingente de estudantes no padrão mais crítico de notas revela que também é possível dizer que houve impacto significativo do programa nesse grupo de alunos. Não há evidências, porém, de que tenha havido uma quebra de padrão ao longo das gerações, ou seja, o padrão do impacto foi muito parecido em todas as gerações, oscilando, na maioria das vezes, dentro da margem de erro.

.....
³⁷ O impacto do nível socioeconômico das famílias no aprendizado, verificado em testes, é uma das evidências mais sólidas da avaliação educacional. O primeiro trabalho a identificar esse impacto foi publicado na década de 1960, nos Estados Unidos, pelo sociólogo James Coleman (Coleman et al., 1966).

Impacto na distribuição de notas

Impacto ao final de 3 anos de JF em:

% de alunos no padrão crítico (nível 1)

Gerações do JF	Língua portuguesa			Matemática		
	Impacto	P-valor (%)	Margem de erro entre gerações	Impacto*	P-valor (%)	Margem de erro entre gerações*
Média	-1,0	0,0	-	-4,2	0,0	-
1ª Geração	-5,2	2,6	2,8	-4,5	3,1	2,8
2ª Geração	-1,0	0,3	2,1	-4,3	0,2	2,4
3ª Geração	-2,2	9,2		-4,1	2,2	

% de alunos nos padrões desejáveis (níveis 3 e 4)

Gerações do JF	Língua portuguesa			Matemática		
	Impacto	P-valor (%)	Margem de erro entre gerações	Impacto	P-valor (%)	Margem de erro entre gerações*
Média	2,9	0,0	-	0,9	1,0	-
1ª Geração	3,4	3,8	2,3	0,6	11,5	1,0
2ª Geração	3,3	0,2	1,8	1,9	0,9	1,0
3ª Geração	1,4	11,0		0,3	27,9	

Tabela 5

Na outra ponta da distribuição, o percentual de alunos com notas adequadas ou avançadas também foi impactado pelo programa – de forma mais modesta na terceira geração, embora a mudança de efeitos da segunda para a terceira esteja dentro das margens de erro. É possível que os efeitos mais fortes sobre o grupo com notas mais baixas sejam indicativos de que a terceira geração foi capaz de reduzir a desigualdade entre os estudantes, sem achatar o topo da distribuição de notas, apenas atuando com mais força sobre os menos favorecidos.

NOVAS PERGUNTAS E O DEBATE CIENTÍFICO EM EDUCAÇÃO

O Instituto Unibanco confirmou, ao longo da última década, que a melhoria da gestão escolar é absolutamente fundamental para que mais estudantes completem a educação básica com nível adequado de aprendizagem. As evidências geradas a partir da avaliação experimental de impacto contribuem para o debate científico em educação, comprovando que liderança, uso de evidências e ciclo de melhoria contínua podem ser efetivos em diversos contextos, tendo em vista o sucesso do Jovem de Futuro em estados de norte a sul do país.

Ao longo das três gerações, o programa foi incorporado às estratégias de gestão próprias de cada rede parceira. Esse movimento foi acompanhado pela universalização do atendimento para todas as escolas e demais instâncias administrativas – regionais e secretaria de educação. Tornar o Jovem de Futuro uma política pública foi a base de uma transformação sustentável e inclusiva na educação.

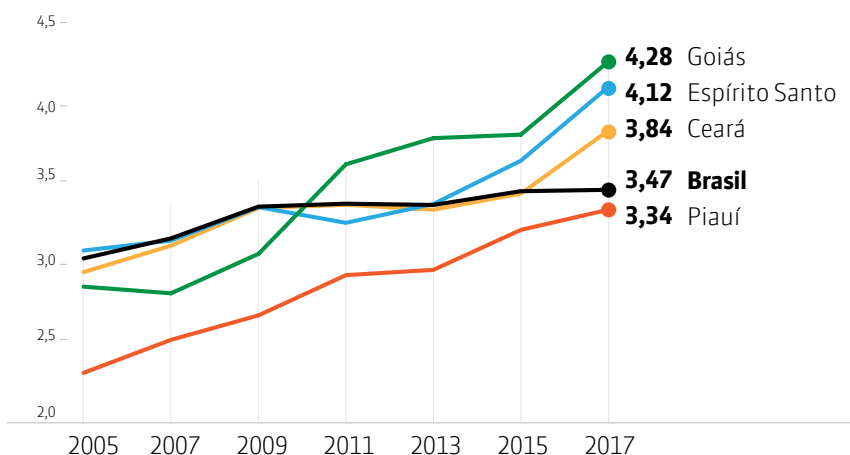
Tendo tais princípios como norte, a implementação avançou e a avaliação de impacto seguiu seu curso. Contudo, na medida em que se avançava rumo a maior escala e sustentabilidade, foi ficando cada vez mais difícil capturar o efeito total do Jovem de Futuro. Até então, ao sortear escolas, a fim de separar aquelas que teriam acesso imediato ao programa das que entrariam posteriormente, era possível identificar o impacto do programa na comparação entre esses dois grupos.

No entanto, a partir do momento em que secretarias e regionais de ensino também começaram a adotar melhores práticas de gestão graças ao programa, esse desenho de avaliação com sorteio de escolas deixou de captar possíveis efeitos do Jovem de Futuro. Com um planejamento mais efetivo e ferramentas apropriadas, assistimos a uma transformação das rotinas centrais e das políticas educacionais em cada estado, beneficiando todo o conjunto de escolas (e não necessariamente apenas aquelas que estavam dentro do programa desde o primeiro momento). Esse efeito adicional,

que beneficia a todas as escolas, não pôde ser capturado pela avaliação experimental, e é possível que os impactos medidos estejam subestimados.

Um fato que pode fornecer algumas pistas – mesmo que não conclusivas – sobre o impacto possivelmente mais amplo do programa é a trajetória dos estados parceiros no Índice de Desenvolvimento da Educação Básica (Ideb), principal indicador do MEC, utilizado no país para monitorar avanços na qualidade do ensino. Ele é formado por uma combinação entre as notas de língua portuguesa e matemática, além da taxa de aprovação, resultados sempre perseguidos pelo Jovem de Futuro e, por essa razão, utilizados na avaliação de impacto. A partir da terceira geração, as metas do programa passaram a ser formuladas em termos do Ideb. É interessante notar que, em todos os estados onde o programa ganhou escala (Ceará, Goiás, Espírito Santo e Piauí), houve melhora sistemática nas respectivas notas do Ideb a partir do início da parceria. Os gestores desses estados consideram a parceria com o Instituto Unibanco um dos elementos que ajudaram a atingir esse resultado³⁸³⁹. No entanto, para ter mais segurança de que não se trata de uma coincidência, o ideal é sanar a dúvida através de uma avaliação experimental de impacto.

Evolução do Ideb dos Estados



Ano de ingresso no Programa	Estado	2ª geração	3ª geração
	Ceará	2012	2016
	Espírito Santo	-	2015
	Goiás	2012	2016
	Piauí	2012	2015

Fonte: elaboração própria com base nos dados do Inep

Gráfico 7 – Evolução do Ideb nos estados onde o JF foi aplicado em grande escala

38 Instituto Unibanco, boletim *Aprendizagem em Foco*, n. 49, mar. 2019. Disponível em: <<https://www.institutounibanco.org.br/aprendizagem-em-foco/49/>>. Acesso em: 8 maio 2020.

39 Ver Henriques e Rocha, 2018.

Em alguns estados, o Instituto Unibanco fez um exercício para tentar estimar a contribuição do programa na melhoria do Ideb. Para isso, foi utilizado não apenas o Ideb (realizado a cada dois anos a partir de testes aplicados pelo Inep/MEC), mas também um indicador análogo, construído a partir das avaliações estaduais, que ficou genericamente conhecido no âmbito do Programa como Ideb* (lê-se “Ideb estrela”). Foi importante utilizar esse indicador, pois o Ideb nacional é disponibilizado somente a cada dois anos. Porém, para a gestão do Jovem de Futuro, é importante contar com metas e resultados anuais. Com o Ideb*, construído à semelhança do Ideb nacional, essa lacuna temporal pôde ser preenchida⁴⁰. Assumindo que Ideb e Ideb* são perfeitamente comparáveis, os impactos medidos pela avaliação experimental da terceira geração indicam que cerca da metade da melhora total no Ideb do Espírito Santo e do Piauí⁴¹ não teria acontecido sem o programa. O fato é que a real contribuição do Jovem de Futuro pode ser ainda maior, dada a melhoria na gestão que ocorre em todo o sistema de educação⁴².

Com a entrada de Minas Gerais no programa, em 2019, novas oportunidades se abriram para aprofundar o conhecimento científico sobre gestão educacional. O estado chegou a participar do programa na fase piloto, que atendeu 44 escolas na ocasião. Dez anos depois, em 2019, foi firmada a parceria com a terceira geração do Jovem de Futuro. Com suas mais de 2,3 mil escolas de ensino médio e 47 regionais de ensino, a nova etapa de avaliação do programa não só dará seguimento à série histórica de impacto, como também investigará duas novas hipóteses.

Até 2018, o número de regionais e escolas em cada estado não era grande o suficiente para viabilizar um sorteio entre regionais – semelhante ao que foi feito com escolas – que gerasse estimativas significativas da contribuição dessas instâncias. As grandes proporções da rede de Minas Gerais permitem que esse tipo de inovação seja incorporado. Assim, será possível conhecer a real importância da gestão nas regionais para a melhoria dos resultados das escolas, e será dado mais um passo para capturar o efeito total do Jovem de Futuro.

O efeito da melhoria da gestão no órgão central é difícil de ser mensurado de forma experimental, entre outras razões porque o número de secretarias estaduais que poderiam vir a se tornar novas parceiras (21

.....

⁴⁰ Ideb* é o nome genérico do indicador anual, análogo ao Ideb, construído para cada estado a partir de informações das avaliações estaduais e da taxa de aprovação, segundo a mesma fórmula de cálculo do indicador nacional.

⁴¹ Embora o Pará tenha participado da avaliação de impacto da terceira geração, a cobertura de escolas se manteve reduzida. Portanto, não é adequado incluí-lo neste exercício de generalização do impacto.

⁴² Vale notar que três estados parceiros do Jovem de Futuro (Goiás, Espírito Santo e Ceará) estavam entre as quatro primeiras posições no ranking de Ideb de ensino médio estadual em 2017. Pernambuco, embora não seja parceiro do programa, tem histórico muito bem consolidado de sucesso em gestão educacional e também aparece nas primeiras posições. Essas evidências somam força ao argumento de que gestão faz diferença para o avanço da educação.

dentre as 27 unidades da federação) é muito pequeno para possibilitar a seleção de grupos de controle e de tratamento. Outras metodologias não experimentais começaram, então, a ser aplicadas.

As grandes proporções de Minas Gerais nos ajudarão ainda a testar uma segunda hipótese, muito relevante no debate atual sobre gestão educacional. Na literatura nacional e internacional, há ampla valorização do que se chama de “liderança instrucional”, relacionada ao acompanhamento do processo pedagógico por parte dos diretores de escola, apoio aos professores e investimento em seu desenvolvimento profissional⁴³. Ainda não há evidências experimentais para essa discussão. Portanto, diferenciar a magnitude dos efeitos das práticas diretamente ligadas ao trabalho com os professores em sala de aula das práticas de gestão e liderança mais gerais, que afetam todo o espaço escolar, é um passo importante para reorientar o desenho.

Assim sendo, todas as ações de gestão pedagógica do Jovem de Futuro foram reunidas em um “pacote pedagógico”⁴⁴ entregue às escolas em distintos momentos. Nenhuma deixará de receber o pacote pedagógico. Um sorteio público foi realizado para definir as escolas que o receberiam logo nos primeiros três anos de implementação e quais seriam incluídas a partir do quarto ano.

Com essas novas avaliações, será possível, portanto, estimar de forma mais precisa o impacto do programa em contextos de implementação em larga escala e entender melhor o peso de algumas ações específicas – no caso, uma ação que busca apoiar gestores a atuarem mais no processo pedagógico – no âmbito das escolas.

.....
⁴³ Alguns dos principais estudos que corroboram essa conclusão são: Fullan, 2001; Leithwood et al., 2004; e Robinson, Lloyd e Rowe, 2008.

⁴⁴ O pacote pedagógico é composto por 7 ações: (1) devolutivas pedagógicas integradas à plataforma Foco Brasil, (2) nova formação para coordenador pedagógico, (3) protocolo direcionado às escolas para a prevenção do abandono escolar, (4) protocolo direcionado às escolas para o acolhimento dos alunos na 1ª série (que chegam do 9º ano com lacunas de aprendizagem), (5) protocolo direcionado às escolas para o uso eficaz do horário de planejamento coletivo com os professores, (6) comunidades de práticas para diretores, coordenadores pedagógicos, supervisores e jovens, as quais tratarão de assuntos pedagógicos, e (7) programa de convivência na escola.

CONCLUSÃO

Como visto ao longo deste livro, a avaliação do Jovem de Futuro mostrou que ele foi capaz de afetar positivamente o aprendizado e a aprovação de jovens no ensino médio, reduzindo desigualdades. Este resultado é ainda mais significativo se considerarmos sua escala e o fato de ter sido implementado em diversas regiões do país. Os impactos identificados são coerentes com os verificados em estudos internacionais que também investigaram o efeito da gestão escolar no desempenho dos alunos.

Apesar de todo o esforço para utilizar o que há de mais cientificamente rigoroso em termos de avaliação de impacto, o Instituto Unibanco tem consciência de que nenhum método é perfeito ou fornece todas as respostas necessárias ao aperfeiçoamento de um programa. Essa conclusão é ainda mais evidente quando se trata da avaliação de um fenômeno tão complexo quanto o aprendizado de estudantes, em centenas de escolas públicas, em diferentes contextos. Investimentos em diversas avaliações vêm sendo feitos nos últimos cinco anos, com resultados que serão abordados num futuro livro.

No entanto, a avaliação experimental, quando feita de forma rigorosa, cumpre o importante papel de fornecer as melhores evidências possíveis sobre o impacto de um programa na população atendida. Ela também é fundamental para subsidiar uma reflexão constante sobre o funcionamento do próprio Programa, em busca de aperfeiçoamento contínuo.

Para o Instituto Unibanco, uma gestão baseada em evidências significa criar nos gestores a capacidade de aprender mais com a prática em cada ciclo, de modo a avançar constantemente. Esse princípio vale tanto para as escolas como para os sistemas de ensino, assim como para o próprio programa Jovem de Futuro, que está em permanente transformação, para contribuir cada vez mais para a melhoria da qualidade do ensino no país.

Ao divulgar os resultados de avaliação de seu principal programa de forma transparente, o Instituto Unibanco procura ainda cumprir sua missão de prestar contas à sociedade sobre o resultado dos esforços realizados na parceria com os estados, contribuindo para o debate público a respeito de seu mais importante objetivo: a garantia dos direitos de aprendizagem de todos os jovens brasileiros.

BIBLIOGRAFIA

ANGRIST, J. D.; PISCHKE, J. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press, 2008.

_____ et al. Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment. *American Economic Review*, Nashville, v. 92, n. 5, p. 1535-58, 2002.

_____; BETTINGER, E.; KREMER, M. Long-term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia. *American Economic Review*, Nashville, v. 96, n. 3, p. 847-62, 2006.

ÄRLESTIG, H.; DAY, C.; JOHANSSON, O. (Eds.). *A Decade of Research on School Principals: Cases from 24 Countries*. Nova York: Springer, 2016.

BANERJEE, A. et al. Can Information Campaigns Raise Awareness and Local Participation In Primary Education? *Economic and Political Weekly*, Mumbai, v. 42, n. 15, p. 1365-72, 2007.

BARROS, R. P. *Desenho da avaliação de impacto do Projeto Jovem de Futuro*. São Paulo: Instituto Unibanco, 2016.

_____ et al. *O impacto do Projeto Jovem de Futuro sobre a aprendizagem em escolas públicas*. São Paulo: Instituto Unibanco, 2016.

_____ et al. *Assessment of the Impact of the Jovem de Futuro Program on Learning*. Nova York: The World Bank, 2018.

_____ et al. *Avaliação de impacto da iniciativa Jovem de Futuro sobre o aprendizado*. São Paulo: Instituto Unibanco, 2020 (no prelo).

_____.; FRANCO, S. *Relatório metodológico: processo de aleatorização da avaliação de impacto do Jovem de Futuro*. São Paulo: Instituto Unibanco, 2018. Disponível em: <https://cedoc.observatoriodeeducacao.org.br/item/?cod=123456789_5232>. Acesso em: 8 maio 2020.

BLOOM, N. et al. Does Management Matter? Evidence from India. *The Quarterly Journal of Economics*, Cambridge, v. 128, n. 1, p. 1-51, 2013.

BLOOM, N.; LEMOS, R.; SADUN, R.; VAN REENEN, J. Does Management Matter in Schools?. *The Economic Journal*, Londres, v. 125, n. 584, p. 647-74, 2015.

BLUNDELL, R.; BOZIO, A.; LAROQUE, G. Extensive and Intensive Margins of Labour Supply: Work and Working Hours in the US, the UK and France. *Fiscal Studies*, Londres, v. 34, n. 1, p. 1-29, 2013.

BOTHWELL, L.; GREENE, J.; PODOLSKY, S.; JONES, David. Assessing the Gold Standard: Lessons from the History of RCTs. *New England Journal of Medicine*, Boston, n. 374. p. 2175-81, 2016. DOI: <10.1056/NEJMms1604593>.

BRASIL. INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP). Matrizes e escalas. Disponível em: <<http://provabrazil.inep.gov.br/escalas-de-proficiencia>>. Acesso em: 2 maio 2020.

_____. Sistema de Avaliação da Educação Básica (Saeb), 2017. Disponível em: <<http://portal.inep.gov.br/educacao-basica/saeb/resultados>>. Acesso em: 2 maio 2020.

BRASIL. MINISTÉRIO DA EDUCAÇÃO (MEC). Plano Nacional de Educação, 2014. Disponível em: <<http://portal.mec.gov.br/arquivos/pdf/pne.pdf>>. Acesso em: 2 maio 2020.

COLEMAN, J. S. et al. *Equality of Educational Opportunity*. Washington: US Government Printing Office, 1966.

DEMING, W. Edwards. *Out of the Crisis*. Cambridge: MIT Press, 2018.

DOBBIE, W.; FRYER, R. G. Getting Beneath the Veil of Effective Schools: Evidence from New York City. *The American Economic Journal: Applied Economics*, Santa Clara, v. 5, n. 4, p. 28-60, 2013.

DUFLO, E.; DUPAS, P.; KREMER, M. School Governance, Teacher Incentives, and Pupil-Teacher Ratios: Experimental Evidence from Kenyan Primary Schools. *Journal of Public Economics*, Amsterdã, v. 123, p. 92-110, 2015.

_____. Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *American Economic Review*, Nashville, v. 101, n. 5, p. 1739-74, 2011.

DUNCAN, G. J.; MAGNUSON, K. Investing in Preschool Programs. *The Journal of Economic Perspectives*, Nashville, v. 27, n. 2, p. 109-32, 2013.

FLESSA, J.; BRAMWELL, D.; FERNÁNDEZ, M.; WEINSTEIN, J. School Leadership in Latin America 2000-2016. *Educational Management Administration & Leadership*, Londres, v. 46, n. 2, p. 182-206, 2018.

FULLAN, M. Large-Scale Reform Comes of Age. *Journal of Educational Change*, Dordrecht, n. 10, p. 101-13, 2009.

FRYER, R. G. Injecting Charter School Best Practices in to Traditional Public Schools: Evidence from Field Experiments. *The Quarterly Journal of Economics*, Cambridge, v. 129, n. 3, p. 1355-1407, 2014.

_____. Management and Student Achievement: Evidence from a Randomized Field Experiment. *NBER Working Papers*, Cambridge, n. 23437, 2017. Disponível em: <<https://www.nber.org/papers/w23437>>. Acesso em: 2 maio 2020.

GERTLER, P. J. et al. *Impact Evaluation in Practice*. Nova York: The World Bank, 2016.

HENRIQUES, R. M. S.; ROCHA, H. C. Ensino médio no Espírito Santo: juventudes e garantia do direito à aprendizagem. In: HARTUNG, P. (org.). *Espírito Santo: como o governo capixaba enfrentou a crise, reconquistou o equilíbrio fiscal e inovou em políticas sociais*. Vitória: Governo do Espírito Santo, 2018.

HERMAN, R. et al. *Turning Around Chronically Low-Performing Schools: A Practice Guide* (NCEE 2008-4020). Washington: National Center for Education Evaluation and Regional Assistance; Institute of Education Sciences, 2008.

HILL, C. J. et al. Empirical Benchmarks for Interpreting Effect Sizes in Research. *Child Development Perspectives*, Chicago, v. 2, n. 3, p. 172-77, 2008.

INSTITUTO UNIBANCO. *Projeto Jovem de Futuro*, 2010.

_____. *Ensino Médio Inovador e Jovem de Futuro: por uma escola mais participativa, eficiente, criativa e transformadora*, 2013.

_____. *O Plano em Ação: circuito de gestão contínuo, integrado e com foco nos resultados de aprendizagem*, 2013.

_____. Sumário executivo sobre a memória e história do Projeto Jovem de Futuro, 2015.

_____. *Investimento social em larga escala*, 2020a (No prelo).

_____. *Teoria da mudança da gestão em educação*, 2020b (No prelo).

LEITHWOOD, K. et al. *How Leadership Influences Student Learning*. Minneapolis; Toronto: Center for Applied Research and Educational Improvement; Ontario Institute for Studies in Education, 2004.

TODOS PELA EDUCAÇÃO. Cenários da educação. Disponível em: <<https://www.todospelaeducacao.org.br/pag/cenarios-da-educacao>>. Acesso em: 2 maio 2020.

_____. *De olho nas metas: primeiro relatório de acompanhamento das 5 metas do movimento todos pela educação*. Organizado pela Comissão Técnica do Compromisso Todos pela Educação. São Paulo, 2008.

ROBINSON, V. *Student-Centered Leadership*. San Francisco: John Wiley & Sons, 2011.

ROBINSON, V. M. J.; LLOYD, C. A.; ROWE, K. J. The Impact of Leadership on Student Outcomes: An Analysis of the Differential Effects of Leadership Types. *Educational Administration Quarterly*, Columbus, v. 44, n. 5, p. 635-74, 2008.

WEINSTEIN, J.; HERNÁNDEZ, M. Birth Pains: Emerging School Leadership Policies in Eight School Systems of Latin America. *International Journal of Leadership in Education: Theory and Practice*, Londres, v. 19, n. 3, p. 241-63, 2016.

Em 2019, o prêmio Nobel de economia foi concedido aos pesquisadores Esther Duflo, Abhijit Banerjee e Michael Kremer por seus estudos de combate à pobreza. O trio utilizou o método mais rigoroso para mensurar o impacto de políticas públicas: a avaliação experimental. Nela, dois grupos são divididos aleatoriamente, de modo que a única diferença observável entre eles seja a participação ou não em determinada ação.

Desde 2008, é também assim que o Instituto Unibanco avalia seu principal programa, o Jovem de Futuro. Métodos menos precisos, como a simples comparação de resultados antes e depois, poderiam até gerar números mais vistosos para a opinião pública, mas seriam ilusórios, pois não indicariam o real efeito do programa nas escolas públicas de ensino médio por ele beneficiadas. A utilização do que há de mais rigoroso no campo da avaliação só foi possível graças à relação de confiança construída com estados parceiros.

Este livro conta a história dessas avaliações e de seus resultados. Eles mostram que, mesmo atuando em larga escala e em estados com realidades bastante distintas, o Jovem de Futuro teve impactos positivos no aprendizado e na aprovação de jovens do ensino médio, com redução de desigualdades.



9 786586 069532