

PROMOVENDO O DESEMPENHO EDUCACIONAL VIA MELHORIAS NA GESTÃO ESCOLAR: O CASO DO PROGRAMA JOVEM DE FUTURO¹

Ricardo Paes de Barros²

Mirela de Carvalho³

Samuel Franco⁴

Beatriz Silva Garcia⁵

Ricardo Henriques⁶

Laura Muller Machado⁷

O desempenho educacional brasileiro é inquestionavelmente limitado, mesmo quando comparado ao de países com o mesmo gasto por estudante. Esse fraco desempenho decorre, em grande medida, de um uso inadequado dos recursos disponíveis. Diante do desafio, o Instituto Unibanco desenvolveu o Jovem de Futuro, programa voltado à melhoria da gestão escolar, adotado ao longo da última década por 3 mil escolas públicas de onze redes estaduais. Este estudo estima, por meio de um desenho experimental, o impacto desse programa sobre a proficiência dos estudantes ao final da 3ª série do ensino médio. Os resultados encontrados indicam um impacto estatisticamente significativo de 10% de um desvio-padrão, robusto entre redes de ensino e ao longo do tempo. Impactos dessa magnitude estão em linha com os encontrados em avaliações similares realizadas internacionalmente e equivalem ao ganho em proficiência que um estudante alcança ao longo de um ano letivo do ensino médio. Esta avaliação demonstra que é possível desenhar e implementar programas capazes de transformar a gestão de escolas públicas e contribuir para o enfrentamento das graves deficiências educacionais do país.

Palavras-chave: avaliação de impacto experimental; gestão escolar; desempenho escolar.

PROMOTING EDUCATIONAL PERFORMANCE THROUGH IMPROVEMENTS IN THE SCHOOL MANAGEMENT: THE CASE OF THE JOVEM DE FUTURO PROGRAM

Educational performance in Brazil is unquestionably limited, even when compared to countries with the same expenditure per student. This weak performance occurs, largely, to an inappropriate use of the available resources. In order to address this challenge, Instituto Unibanco developed *Jovem de Futuro*: a program focused on improving school management, adopted throughout the last decade by three thousand schools from eleven state school networks. This study estimates, through an experimental design, the impact of such Program on the proficiency of students at the end of 3rd grade of High School (which corresponds to the final grade of High School in Brazil). The results found show a statistically significant impact of 10% from a standard deviation, which is robust between the different teaching networks and throughout time. Impacts of this magnitude

1. DOI: <http://dx.doi.org/10.38116/ppe51n3art1>

2. Professor do Instituto de Ensino e Pesquisa (Insper). *E-mail*: <ricardopb1@insper.edu.br>.

3. Gerente de gestão do conhecimento do Instituto Unibanco. *E-mail*: <mirela.silva@institutounibanco.org.br>.

4. Diretor técnico da Oppen Social. *E-mail*: <samuelfranco@oppen.social>.

5. Analista sênior de pesquisas socioeducacionais do Instituto Unibanco. *E-mail*: <bsgarcia028@gmail.com>.

6. Superintendente executivo do Instituto Unibanco. *E-mail*: <ricardo.henriques@institutounibanco.org.br>.

7. Professora do Insper. *E-mail*: <lauramm1@insper.edu.br>.

are aligned with the findings of similar evaluations carried out internationally and are equivalent to the gain in proficiency that a student reaches throughout one school year in High School. This evaluation demonstrates that it is possible to design and implement programs that can transform the management of public schools and contribute to facing the severe educational flaws in the country.

Keywords: experimental impact evaluation; school management; school performance.

JEL: C93; I20; I21.

1 INTRODUÇÃO

1.1 Fraco e desigual desempenho educacional

O Brasil tem metas para a educação básica bem estabelecidas. Algumas são requerimentos constitucionais, como é o caso do art. 208 da Constituição Federal de 1988 (CF/1988), que estabelece a obrigatoriedade e a gratuidade da educação básica a todos dos 4 aos 17 anos de idade; outras são definidas por lei, como as vinte metas do Plano Nacional de Educação (PNE),⁸ ou refletem o compromisso e a preocupação da sociedade civil, como as cinco metas do movimento Todos pela Educação (TPE). Em conjunto, visam garantir adequados insumos e resultados em todas as etapas do sistema educacional.

O Brasil definitivamente não deve alcançar suas metas de resultado (tabela 1). Na sequência, buscamos demonstrar esse fato. Dado que o objetivo do programa Jovem de Futuro é promover melhores resultados no ensino médio, centramos a atenção nas metas relacionadas ao acesso, conclusão e aprendizado nessa etapa do ciclo educacional.

No que se refere ao acesso à escola, a meta 3 do PNE estabelece que, em 2016, 100% dos jovens no grupo etário de 15 a 17 anos já deveriam estar frequentando a escola ou terem concluído a educação básica. Apesar disso, em 2017, 9% desse grupo etário ainda não havia concluído a educação básica e permanecia fora da escola (tabela 1). A meta 3 do PNE também estabelece que, até 2024, 85% dos jovens nesse grupo etário deveriam estar frequentando o ensino médio ou já tê-lo concluído. Conforme a tabela 1 revela, em 2017, apenas 70% dos jovens nesse grupo etário estavam frequentando o ensino médio ou já o haviam concluído. Como o crescimento da taxa líquida de matrícula desse grupo etário ao longo da última década foi de 1,4 ponto percentual (p.p.) ao ano (a.a.), mantido o ritmo atual, deve-se alcançar em 2024 uma taxa de apenas 80% – portanto, aquém dos 85% estabelecidos pela meta 3 do PNE.

Em termos de conclusão, uma das metas é que, até 2021, 90% dos jovens de 19 anos tenham concluído o ensino médio (meta 4 do TPE). No entanto, em 2017, apenas 59% dos jovens de 19 anos já haviam concluído o ensino médio

8. Definidas pela Lei nº 13.005, de 25 de junho de 2014.

(tabela 1). Como o progresso nessa taxa tem sido inferior a 2 p.p. por ano, mantida a velocidade atual, não se deve chegar sequer a uma taxa de conclusão do ensino médio de 70% em 2021 – ficando, portanto, muito aquém do estabelecido pela meta 4 do TPE.

Por fim, a meta de aprendizado é que 70% dos estudantes ao final do ensino médio tenham proficiência em língua portuguesa e matemática considerada adequada⁹ (meta 3 do TPE). Nesse aspecto, a situação do país é ainda pior. Em 2017, menos de 30% dos estudantes tinham, ao final do ensino médio, aprendizado adequado em língua portuguesa, e em matemática, menos de 10%. Em língua portuguesa, o progresso tem sido extremamente lento (inferior a 0,6 p.p. por ano), ao passo que, em matemática, essa porcentagem vem declinando. Mantido o ritmo atual, certamente a meta de jovens com aprendizado adequado não será cumprida.

TABELA 1
Situação atual e evolução dos indicadores de acompanhamento das metas educacionais no Brasil

Meta	Indicador	Situação atual – 2017 (%)	Progresso anual médio na última década – 2005-2015 (p.p.)	Situação prevista para o ano de referência da meta (%)
PNE				
Universalizar, até 2016, o atendimento escolar para toda a população de 15 a 17 anos (...).	Porcentagem de jovens de 15 a 17 anos que frequentam escola ou já concluíram a educação básica	91	0,6	91
(...) e elevar, até o final do período de vigência deste PNE [2024], a taxa líquida de matrículas no ensino médio para 85%.	Porcentagem de jovens de 15 a 17 anos que frequentam o ensino médio ou já concluíram a educação básica	70	1,4	80
Todos pela Educação				
Até 2021, (...) 90% ou mais dos jovens brasileiros de 19 anos deverão ter completado o ensino médio.	Porcentagem de jovens de 19 anos que concluíram o ensino médio	59	1,5	65
Até 2021, 70% ou mais dos estudantes terão aprendido o que é adequado para seu ano (300 pontos na escala Saeb em Língua Portuguesa).	Porcentagem de estudantes do 3º ano do ensino médio com pontuação acima do nível considerado adequado no Saeb	29	0,5	31
Até 2021, 70% ou mais dos estudantes terão aprendido o que é adequado para seu ano (350 pontos na escala Saeb em Matemática).	Porcentagem de estudantes do 3º ano do ensino médio com pontuação acima do nível considerado adequado no Saeb	9	-0,1	9

Fontes: Para PNE, Inep (2019a, p. 61-66); para Todos pela Educação, Todos pela Educação (2019, p. 42-70).
Elaboração dos autores.

Obs.: Para o cálculo do progresso anual, no caso dos resultados divulgados pelo Todos pela Educação, utilizou-se o quinquênio 2012-2017.

9. Trezentos pontos na escala Sistema de Avaliação da Educação Básica (Saeb) em língua portuguesa e 350 em matemática.

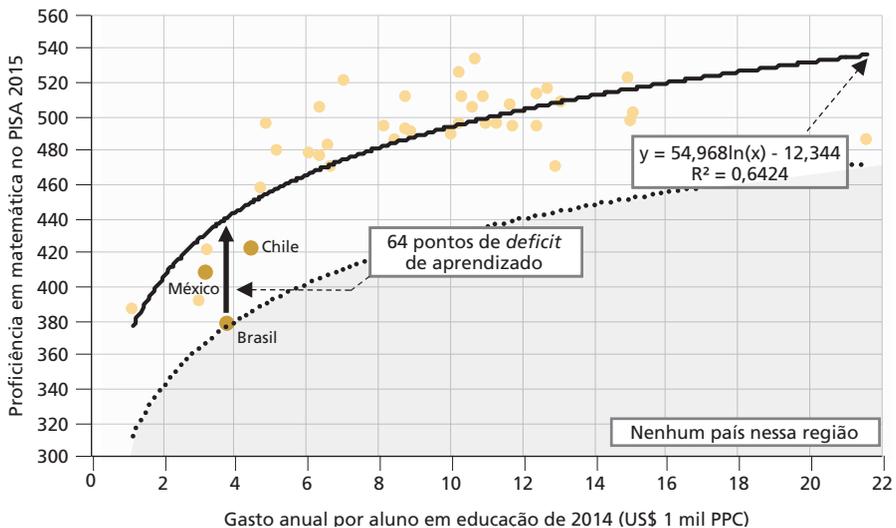
1.2 Esforço fiscal e ineficiência

A incapacidade brasileira de alcançar suas metas certamente não pode ser justificada por uma correspondente falta de atenção da sociedade com a educação. O país dedica 6% do seu produto interno bruto (PIB) à educação (Inep, 2019b). Segundo a Resolução 14 da Declaração de Incheon,¹⁰ os gastos públicos com educação devem ser de 4% a 6% do PIB ou 15% a 20% do total das despesas públicas. O esforço fiscal brasileiro já está, portanto, perfeitamente alinhado com o compromisso internacional.

Dada a magnitude do gasto público com educação, o desempenho educacional brasileiro deveria ser muito superior ao alcançado. Conforme ilustra o gráfico 1, o aprendizado em matemática no Programme for International Student Assessment (PISA) no Brasil encontra-se 64 pontos¹¹ abaixo do que se deveria esperar de países com o mesmo gasto por estudante.¹² Essa distância e esse grau de *aparente ineficiência* no uso dos recursos educacionais são tão elevados que nenhum outro país participante do PISA apresenta uma distância similar ou superior.

GRÁFICO 1

Relação entre desempenho em matemática no PISA 2015 e gasto anual por estudante em 2014



Fontes: OECD (2016, p. 44; 2017, p. 177).

Elaboração dos autores.

Obs.: Para comparação dos gastos entre países, OECD (2017) converteu todos os valores em dólares americanos de 2014 ajustados para a paridade do poder de compra (PPC). Os gastos se referem ao gasto público por estudante nos anos finais do ensino fundamental e do ensino médio.

10. Disponível em: <<https://bit.ly/3vite3i>>. Acesso em: 24 jul. 2019

11. Na escala PISA, 64 pontos equivalem a 64% de um desvio-padrão na distribuição de proficiência entre estudantes.

12. O PISA avalia o desempenho de estudantes de 15 anos, boa parte ainda no ensino fundamental. Para medir a ineficiência no ensino médio, seria necessário contar com uma medida que avaliasse estudantes ao final do ensino médio. No entanto, nenhuma avaliação dessa natureza comparável internacionalmente está disponível.

Existem também indicações de que a ineficiência no uso dos recursos educacionais brasileiros esteja aumentando: enquanto o progresso educacional tem sido lento, o gasto público cresce a taxas acentuadas. Entre 2010 e 2014, no Brasil, os gastos públicos com a educação básica por estudante cresceram em termos reais – segundo a Organização para a Cooperação e Desenvolvimento Econômico (OCDE), mais de 50% (OECD, 2017, p. 179). Nenhum outro país para o qual se tem essa informação apresentou uma taxa de crescimento nos gastos com educação acima da brasileira.¹³ O país definitivamente não está sendo capaz de traduzir uma maior disponibilidade de recursos em maior acesso, progresso ou aprendizado dos estudantes.

Mais do que isso, no Brasil, a maioria da elevada desigualdade em resultados educacionais existente não resulta de diferenças no gasto por estudante, e sim de diferenças na *aparente eficiência* com que esses recursos são utilizados. De fato, 88% dos diferenciais entre redes municipais de ensino, com relação à proficiência em matemática ao final do ensino fundamental, deve-se às diferenças entre redes com o mesmo gasto por estudante e, portanto, em grande medida, às diferenças na *aparente eficiência* com que os recursos disponíveis são utilizados.^{14,15}

Em suma, no Brasil, um elevado e desigual grau de ineficiência na utilização dos recursos públicos alocados à educação parece explicar boa parte dos fracos e desiguais resultados educacionais alcançados.

1.3 Gestão e eficiência

Pode-se entender gestão como um conjunto de atividades voltadas à melhoria na eficiência com que os recursos disponíveis são utilizados. Isto é, atividades cujo objetivo é alcançar melhores resultados com os mesmos recursos ou alcançar resultados preestabelecidos com a utilização de um volume menor de recursos. Nesse caso, gestão e eficiência estariam intimamente relacionadas: o objetivo primordial da gestão seria obter ganhos de eficiência.

13. Embora essas sejam as estimativas internacionalmente comparáveis da OCDE, segundo os Indicadores Financeiros Educacionais, divulgados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), o gasto por estudante na educação básica brasileira cresceu, em termos reais, 26% entre 2010 e 2014 (Inep, 2019b). Crescimento bem inferior ao estimado para o Brasil pela OCDE, mesmo assim, superior ao de todos os demais países, exceto Turquia (OECD, 2017, p. 179).

14. Essa estimativa corresponde ao R^2 (porcentagem da variância explicada) da regressão log-linear, em que as unidades de análise são as redes municipais de ensino, a variável dependente é a proficiência média em matemática dos estudantes ao final do 9º ano do ensino fundamental na Prova Brasil de 2015, divulgados por Inep (2017), e a variável explicativa é o logaritmo natural do gasto municipal por estudante em 2014, obtida pelo Sistema de Informações sobre Orçamentos Públicos em Educação – Siope (disponível em: <<https://bit.ly/3BOvEli>>). Foram utilizadas informações válidas para 3.120 redes municipais de ensino.

15. Denominamos de *aparente eficiência*, ao invés de *eficiência*, uma vez que certamente também existem outros fatores que podem explicar diferenças de resultados entre redes com o mesmo gasto por estudante: diferenças de custo de vida local e grau de pobreza familiar são candidatos típicos.

Instituições, nos mais diversos setores econômicos, alocam um montante significativo de recursos para aprimorar sua gestão. É de se esperar, portanto, que gestão leve a maior eficiência. A despeito de razões teóricas e ampla evidência anedótica, quantificações sistemáticas dessa relação são raras. Uma notável exceção é o estudo de Bloom *et al.* (2013), que comprova, com base em uma avaliação experimental, que o acesso à assessoria em gestão, de fato, tem impacto sobre produtividade e, portanto, sobre eficiência.

Resta verificar se melhorias em gestão são também capazes de promover maior eficiência na provisão pública de educação. Como menciona Lück (2009, p. 24), a gestão escolar visa

promover a organização, a mobilização e a articulação de todas as condições materiais e humanas necessárias para garantir o avanço dos processos socioeducacionais dos estabelecimentos de ensino, orientados para a promoção *efetiva* da aprendizagem dos alunos.

Infelizmente, existem poucos estudos que buscam corroborar empiricamente essa pressuposição. Os estudos realizados por Fryer Junior (2014; 2017) são exceções. Com base em uma avaliação experimental, o autor encontra que estudantes em escolas cujo diretor teve acesso a um programa de formação em gestão de 300 horas têm um aprendizado 10% a 20% de um desvio-padrão acima dos estudantes em escolas cujo diretor não teve acesso a essa formação (Fryer Junior, 2017). Ele também, com base em uma avaliação experimental, encontra evidência de que a adoção por escolas públicas de boas práticas em gestão, derivadas de escolas comunitárias¹⁶ de alto desempenho, têm impacto significativo, de 15% a 20% de um desvio-padrão, sobre o aprendizado dos seus estudantes (Fryer Junior, 2014).

Outro passo nessa direção foi dado por Bloom *et al.* (2015), em estudo que investiga a associação entre as competências de gestão dos diretores escolares e o aprendizado em matemática em escolas brasileiras. Esses autores constataam que o aumento de um desvio-padrão nas habilidades dos gestores está associado a um aumento de 10% de um desvio-padrão no aprendizado dos estudantes. Vale ressaltar que essa relação é bem menos intensa que a obtida pelo mesmo estudo para outros países.

1.4 Escopo

Existe, portanto, evidência de que a ineficiência no uso dos recursos é um dos determinantes centrais do fraco e desigual desempenho educacional brasileiro, e também alguma evidência de que ações dirigidas à promoção de melhor gestão podem promover maior eficiência na provisão pública da educação. Assim, para o Brasil alcançar um melhor e mais equitativo desempenho educacional, seria recomendável

16. Tradução aproximada para *charter schools*.

incorporar à sua política educacional programas voltados à melhoria da gestão das redes de ensino e das escolas. Em consonância com essa perspectiva, diversas iniciativas voltadas à promoção de melhorias na gestão educacional têm sido adotadas no país, destacando-se, entre elas, a Escola de Gestores; o Programa Nacional de Capacitação de Conselheiros Municipais de Educação (Pró-Conselho) e o Programa de Apoio aos Dirigentes Municipais de Educação (Pradime),¹⁷ do MEC; o Formar, da Fundação Lemann;¹⁸ o Gestão em Foco;¹⁹ e o Jovem de Futuro, do Instituto Unibanco, entre outras. Resta verificar se essas ações são de fato eficazes na promoção de maior eficiência e, conseqüentemente, melhor desempenho educacional.

O objetivo geral deste estudo é, com base em uma avaliação de natureza experimental, investigar a efetiva capacidade do programa Jovem de Futuro de promover maior proficiência em matemática e língua portuguesa via melhorias na gestão. Esse objetivo geral abarca três objetivos específicos inter-relacionados, conforme descrito adiante.

- 1) Identificar a existência e avaliar a magnitude do impacto do programa sobre a proficiência dos estudantes.
- 2) Avaliar se a magnitude desse impacto é substantivamente relevante, dada a dimensão do problema educacional brasileiro.
- 3) Avaliar a sensibilidade da magnitude desse impacto às condições locais onde o programa é implantado e em que medida as inovações incorporadas ao seu desenho, em uma década de experiência, têm sido frutíferas.

2 JOVEM DE FUTURO

2.1 Desenho

O Jovem de Futuro é um programa do Instituto Unibanco,²⁰ implantado em parceria com redes estaduais de ensino, voltado à promoção de maior eficiência na utilização dos recursos educacionais por meio da adoção de modelos de gestão

17. O Pró-Conselho estimula a criação de novos conselhos municipais de educação, o fortalecimento daqueles já existentes e a participação da sociedade civil na avaliação, definição e fiscalização das políticas educacionais, entre outras ações (mais informações em: <[https:// bit.ly/39C5w9D](https://bit.ly/39C5w9D)>; acesso em: 23 jul. 2019). O Pradime foi criado com o objetivo de fortalecer e apoiar os dirigentes da educação municipal na gestão dos sistemas de ensino e das políticas educacionais (mais informações em: <<https://bit.ly/3HulcGd>>; acesso em: 23 jul. 2019). O Programa Nacional Escola de Gestores da Educação Básica Pública visa “qualificar os gestores das escolas da educação básica pública a partir do oferecimento de cursos de formação a distância. A formação dos gestores é feita por uma rede de universidades públicas parceiras do MEC” (mais informações em: <<https://bit.ly/3hg1ODg>>; acesso em: 23 jul. 2019).

18. O Programa Formar é uma parceria entre a Fundação Lemann e redes públicas de educação para apoiar os educadores em seu desenvolvimento profissional e no aperfeiçoamento das práticas de gestão e dos processos pedagógicos. Mais informações sobre esse programa em: <<https://bit.ly/3kKZq8Z>>. Acesso em: 23 jul. 2019.

19. O Gestão em Foco é um programa que visa ao fortalecimento da gestão escolar na rede estadual do Paraná. Ele está previsto na ação nº 3 do programa Minha Escola Tem Ação (Meta). O Gestão em Foco se constitui de várias ações, entre elas, a formação continuada voltada à equipe diretiva da escola, prevista no art. 20, inciso II, item d, da Lei nº 18.590/2015. Mais informações em: <<https://bit.ly/34YVkJG>>. Acesso em: 23 jul. 2019.

20. Informações mais detalhadas sobre o programa Jovem de Futuro podem ser obtidas em Barros *et al.* (2017, p. 20-38).

para resultados. O programa, que originalmente se centrava na gestão escolar, vem progressivamente expandindo sua atuação a todos os níveis do sistema de ensino: da gestão escolar à administração central da Secretaria de Educação, passando pelas administrações regionais.

Iniciado em 2008, o programa vem, desde então, sendo continuamente aperfeiçoado. Atualmente, em sua terceira versão, conta com quatro elementos complementares. Primeiro, oferece formação em gestão voltada para resultados, composta por 72 horas presenciais para técnicos das secretarias, dirigentes regionais e supervisores escolares; e 48 horas presenciais, complementadas por 120 horas a distância, para gestores escolares e coordenadores pedagógicos.

Segundo, institui um circuito formal, em que estão presentes as quatro fases clássicas da gestão, conforme descrito a seguir.

- 1) Planejamento participativo voltado para resultados (metas) e fortemente baseado em evidência.
- 2) Execução do planejado.
- 3) Monitoramento, avaliação e análise dos resultados alcançados, incluindo momentos para troca de experiências entre escolas da mesma regional.
- 4) Identificação de ajustes, mudanças de rota necessárias e replanejamento das ações.

Além do circuito de gestão para as escolas, também se adota um correspondente circuito para as regionais e outro exclusivo da Secretaria de Educação.

Terceiro, instrumentaliza os gestores com metas, protocolos e práticas de gestão que facilitam, estimulam e promovem um melhor desempenho nas quatro fases do circuito de gestão. Por fim, conta com supervisores escolares: agentes ligados à administração regional, externos à escola, que têm, por finalidade, monitorar e assessorar as escolas em cada uma das quatro etapas da gestão. A utilização de supervisores escolares é um elemento de vital importância para o funcionamento do Jovem de Futuro.

Embora tenha potencial para ser adaptado para as diferentes etapas da educação, desde sua concepção, o programa foi aplicado apenas em escolas públicas estaduais de ensino médio. Uma vez adotado por uma rede estadual, sua implantação é apoiada pelo Instituto Unibanco por, ao menos, seis anos.

2.2 Difusão

Adotado pela primeira vez em 2008, em 45 escolas nas regiões metropolitanas de Belo Horizonte e Porto Alegre, o Jovem de Futuro, em mais de uma década de contínua expansão, já atingiu mais de 3 mil escolas de onze redes estaduais.²¹

Sua expansão ocorreu em três fases (gerações) que se sobrepõem (tabela 2). A primeira geração, considerada piloto, envolveu cerca de duzentas escolas – distribuídas entre cinco aglomerados urbanos, em quatro redes estaduais –,²² e aconteceu entre 2008 e 2013.²³ As escolas atendidas nessa primeira geração representavam apenas de 2% a 4% do total de escolas com ensino médio regular das redes atendidas.

TABELA 2
Escolas atendidas pelo programa Jovem de Futuro, por ano de entrada (2008-2018)

Redes	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	Nº total de escolas atendidas	Participação na rede atendida (%)
Minas Gerais	20			24								44	2
Rio Grande do Sul	25			21								46	4
São Paulo		41				36						77	2
Rio de Janeiro		15				15						30	3
Ceará					100	124	169	46		211	3	653	102
Goiás					179	120	281		74	10	18	682	108
Mato Grosso do Sul					99	96	76					271	88
Pará					131	254	70	6	0	20	6	487	86
Piauí					73	163	176	6	0	25	42	485	99
Espírito Santo								150	0	66	41	257	89
Rio Grande do Norte										143		143	50
1ª geração	45	56	0	45	0	51	0	0	0	0	0	197	2
2ª geração	0	0	0	0	582	757	772	46	0	0	0	2.157	82

(Continua)

21. Em 2019, o Jovem de Futuro foi novamente expandido para atender a boa parte da rede estadual de Minas Gerais. As informações apresentadas sobre a difusão do programa não incluem essa expansão mais recente. Informações sobre quais escolas estão sendo ou foram atendidas, e desde quando, eram inicialmente acompanhadas pelo Monitoramento Físico-Financeiro (MFF), sistema que foi posteriormente aprimorado e substituído pelo Sistema de Gestão de Projetos (SGP). Todas as informações utilizadas neste estudo sobre o atendimento às escolas das redes participantes foram extraídas desses sistemas.

22. O programa foi adotado nas regiões metropolitanas das capitais das redes estaduais de Minas Gerais, Rio Grande do Sul, Rio de Janeiro e São Paulo. Foi também adotado em São Paulo, no aglomerado urbano do Vale do Paraíba, formado por São José dos Campos e Jacareí.

23. Essas datas se referem ao intervalo entre o momento em que as primeiras escolas começaram a ser atendidas e o momento em que as últimas escolas dessas redes começaram a ser atendidas. A presença do programa nessas redes ocorreu por um período muito mais longo.

(Continuação)

Redes	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	Nº total de escolas atendidas	Participação na rede atendida (%)
3ª geração	0	0	0	0	0	0	0	162	74	475	110	821	28
3ª geração (incluindo as gerações anteriores)	0	0	0	0	0	0	0	334	620	908	110	2.072	71
Todas as gerações	45	56	0	45	582	808	772	208	74	475	110	3.175	27

Legenda:

- 1ª geração
- 2ª geração
- 3ª geração

Fontes: MFF; SGP; e Censos Escolares 2014 a 2018.

Elaboração dos autores.

Obs.: Para o percentual de atendimento, foi considerado o número médio de escolas entre os anos de 2014 e 2018 com oferta de ensino médio na rede estadual. Devido à possibilidade do fechamento de escolas, o percentual da rede atendida pode superar 100%.

Na primeira geração, o programa estava dirigido apenas à gestão escolar; baseava-se mais em mudanças de mentalidade e cultura do que na efetiva oferta de protocolos de gestão; contava com envolvimento direto do Instituto Unibanco nas escolas,²⁴ ao contrário do que caracterizou as gerações seguintes; e garantia a cada escola um apoio financeiro anual para a implementação do seu plano de ação, da ordem de R\$ 100 por estudante.²⁵

A segunda geração foi iniciada em 2012, em cinco redes estaduais (Ceará, Goiás, Mato Grosso do Sul, Pará e Piauí), envolvendo ao todo 2,2 mil escolas, representando mais de 80% do total de escolas dessas redes com ensino médio regular. Vale ressaltar que o programa não é necessariamente adequado para todo tipo de escola. Assim, algumas redes optaram por não incluir algumas de suas escolas, tipicamente aquelas muito pequenas. Essa geração contou com programas de formação, protocolos e práticas de gestão mais estruturados que os utilizados na primeira. O atendimento do Instituto Unibanco às escolas passou a ser indireto, por intermédio da formação e da utilização de supervisores da rede estadual para atendimento às escolas. A despeito de a atuação ocorrer via secretarias estaduais de educação, o foco permaneceu na promoção de melhorias na gestão escolar. O apoio financeiro à implantação do plano de ação das escolas passou a ser responsabilidade do setor público, mais especificamente do Ministério da Educação, por meio do

24. Cada escola recebia assessoria em gestão diretamente do Instituto Unibanco.

25. Inicialmente, o apoio por estudante era dado para estudantes em todas as etapas da educação básica. Posteriormente, passou a ser proporcional apenas à matrícula no ensino médio.

Programa Ensino Médio Inovador (ProEMI).²⁶ Em 2015, o programa Jovem de Futuro deixou de estar acoplado ao ProEMI, marcando o final da segunda geração.

A terceira geração iniciou-se em 2015 em três estados (Pará, Piauí e Espírito Santo), expandiu-se, no ano seguinte, para Goiás, e dois anos depois, para Ceará e Rio Grande do Norte, alcançando, ao todo, cerca de 2,1 mil escolas,²⁷ o que representava mais de 70% do total de escolas com ensino médio regular dessas redes. Nessa terceira geração, o apoio financeiro foi formalmente extinto e o escopo ampliado para abarcar a promoção de melhorias, tanto na gestão escolar como na gestão da administração central e das regionais das redes estaduais. Os protocolos, procedimentos e formações se tornaram ainda mais estruturados e formalizados, além de ter havido a implantação de um circuito de gestão sincronizado e operante em todos os níveis de gestão.

3 IMPLEMENTAÇÃO DO PROGRAMA E POSSIBILIDADES DE AVALIAÇÃO

O modelo de implantação do programa Jovem de Futuro, em cada rede estadual, apresenta duas características que facilitam sobremaneira a identificação e estimação da magnitude do seu impacto médio:²⁸ i) implantação progressiva, com as últimas escolas sendo atendidas no mínimo três anos após o atendimento das primeiras; e ii) seleção, por sorteio, da ordem de entrada de ao menos parte das escolas.

Em cada uma das redes estaduais que adotaram o programa, a entrada das escolas foi progressiva, com as últimas escolas atendidas iniciando o atendimento sempre no mínimo três anos após as primeiras. Assim, ao longo de três anos, chamados de janela de avaliação, sempre coexistem escolas atendidas e não atendidas, permitindo, ao menos ao longo dessa janela, contrastar o desempenho dessas escolas.

Além disso, para uma parcela significativa de escolas programadas para serem atendidas,²⁹ o ano de admissão foi definido por sorteio. Das 3,3 mil escolas programadas para serem atendidas, mais da metade (1,7 mil escolas) teve seu momento de ingresso definido por sorteio (tabela 3).³⁰

26. Mais informações sobre o ProEMI em: <<https://bit.ly/3pf8QMJ>>. Acesso em: 23 jul. 2019.

27. Pará, Piauí, Goiás e Ceará participaram da segunda geração do programa. Nesses estados, a partir de 2015, ocorreu tanto expansão da cobertura como migração de escolas atendidas pelo modelo da segunda para o da terceira geração. O número de escolas atendidas pelo modelo da segunda geração mencionado no texto (2,1 mil) contabiliza as escolas que já praticavam o modelo da segunda geração e migraram para a terceira geração.

28. Como em todas as avaliações experimentais, a identificação de outras características da distribuição do impacto além da média, tais como a mediana ou o desvio-padrão, requer hipóteses adicionais.

29. Nem todas as escolas programadas para participar do programa participaram. No Mato Grosso do Sul, por exemplo, a rede suspendeu o programa antes da participação de todas as escolas programadas. Essa suspensão ocorreu após o final da janela de avaliação e, portanto, em nada prejudica a inclusão dessa rede na avaliação de impacto do programa.

30. Nem todas as escolas que participaram do programa tiveram o momento que sua participação seria iniciada definida por sorteio. Informações sobre quais escolas participaram do sorteio e sobre o resultado do sorteio não constam dos sistemas MFF e SGP, que monitoravam a implementação do programa. O Instituto Unibanco, no entanto, conta com registros detalhados de todos os sorteios realizados. Todas as informações utilizadas neste estudo sobre os sorteios foram extraídas desses registros.

O conjunto de escolas que teve seu momento de ingresso definido por sorteio foi selecionado no instante em que a rede aderiu ao programa, com um subconjunto sempre sendo selecionado para atendimento imediato³¹ e outro para atendimento apenas após três anos desde o início da adoção do programa pela rede (isto é, após o final da janela de avaliação). A tabela 4 apresenta a distribuição desse conjunto de escolas, segundo a rede e o ano definido por sorteio para ingresso no programa.

TABELA 3
Escolas programadas para serem atendidas e que tiveram o momento de ingresso no programa definido por sorteio

Rede	Número de escolas programadas para serem atendidas			
	Programadas para serem atendidas	Não tiveram o momento de ingresso no programa definido por sorteio	Tiveram o momento do ingresso no programa definido por sorteio	Escolas programadas para serem atendidas que tiveram o momento de ingresso definido por sorteio (%)
Minas Gerais	48	0	48	100
Rio Grande do Sul	50	0	50	100
São Paulo	81	1	80	99
Rio de Janeiro	30	0	30	100
Ceará	653	275	378	58
Goiás	682	537	145	21
Mato Grosso do Sul	296	171	125	42
Pará	487	374	113	23
Piauí	485	210	275	57
Espírito Santo	257	36	221	86
Rio Grande do Norte	243	0	243	100
Todas as redes	3.312	1.604	1.708	52

Fontes: Sistemas de monitoramento do atendimento, MFF e SGP, e registros do Instituto Unibanco sobre os procedimentos e resultados dos sorteios.

Elaboração dos autores.

Obs.: Nem todas as escolas programadas para participar do programa o fizeram. No Mato Grosso do Sul, por exemplo, a rede suspendeu o programa antes que todas as escolas que estavam programadas para participar tivessem participado. No Rio Grande do Norte, cem escolas passaram a ser atendidas apenas a partir de 2020.

31. Exceções são o Ceará e o Piauí, que aderiram ao programa em 2012, mas só em 2013 e 2015, respectivamente, promoveram o ingresso das primeiras escolas cujo momento de entrada foi definido por sorteio (tabela 4). No Pará, a adesão ao programa ocorreu em etapas: dois conjuntos de escolas foram progressivamente selecionados, em 2012 e 2015, para ter o momento de ingresso no programa definido por sorteio. Em 2015, no Pará e no Piauí, quando a terceira geração foi adotada, o programa foi "reiniciado", no sentido de que uma nova definição do momento de atendimento foi realizada, incluindo não apenas escolas que ainda não haviam sido beneficiadas, mas também escolas que já haviam sido beneficiadas. Dessa forma, em alguns casos, definidos por sorteio, escolas que estavam sendo atendidas pela segunda geração do programa tiveram o atendimento interrompido. No Pará, os conjuntos de escolas envolvidas nos dois sorteios não foram disjuntos. Algumas escolas (precisamente catorze) participaram dos dois sorteios e, por isso, participam tanto da avaliação da segunda como da terceira geração nesse estado. Por esse motivo, o número de escolas no Pará que tiveram o momento do ingresso no programa definido por sorteio na tabela 3 (em que as que participaram de dois sorteios são contabilizadas uma única vez) é menor que o correspondente número de escolas por ano previsto para atendimento definido na tabela 4 (em que cada nova entrada no programa é contabilizada).

Neste estudo, obtém-se um estimador não enviesado da magnitude do impacto do programa, aproveitando-se essas duas características notáveis do seu modelo de implantação: *seleção por sorteio* do momento de ingresso, para mais da metade das escolas atendidas; e *implantação progressiva*, com mais de metade das escolas participantes do sorteio recebendo atendimento imediato e mais de um quarto recebendo atendimento apenas ao final da janela de avaliação. Estatisticamente, as escolas escolhidas por sorteio para participar apenas após o final da janela de avaliação apontam, ao longo desta janela, para o que teria acontecido com as escolhidas para atendimento imediato, também por sorteio, caso essas não fossem atendidas. Dessa forma, uma medida não enviesada da magnitude do impacto do programa, ao longo da janela de avaliação, é dada pela diferença entre o desempenho das escolas escolhidas para serem atendidas de imediato e o desempenho das escolhidas para atendimento posterior à janela.

TABELA 4
Escolas participantes de sorteio segundo a rede e o ano previsto para o atendimento pelo programa Jovem de Futuro

Redes	Ano de entrada (2008-2018)																Total	Distribuição de escolas segundo tipo de atendimento	
	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	Imediato	Após a janela de avaliação				
1ª geração																			
Minas Gerais	20			28													48	42	58
Rio Grande do Sul	25			25													50	50	50
São Paulo		40					40										80	50	50
Rio de Janeiro		15					15										30	50	50
2ª geração																			
Ceará						125	170	46	26								367	34	7
Goiás					120					25							145	83	17
Mato Grosso do Sul					100				25								125	80	20
Pará – 1ª etapa					25				25								50	50	50
3ª geração																			
Espírito Santo								151		34	36						221	68	16
Pará – 2ª etapa								45				42					87	52	48
Piauí								139		16	120						275	51	44
Rio Grande do Norte											143			100			243	59	41

(Continua)

(Continuação)

Redes	Ano de entrada (2008-2018)														Distribuição de escolas segundo tipo de atendimento		
	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	Total	Imediato	Após a janela de avaliação	
1ª geração	45	55	0	53	0	55	0	0	0	0	0	0	0	208	48	52	
2ª geração	0	0	0	0	245	125	170	96	51	0	0	0	0	687	54	15	
3ª geração	0	0	0	0	0	0	0	335	0	193	198	0	100	826	58	36	
Todos as redes	45	55	0	53	245	180	170	431	51	193	198	0	100	1.721	55	29	

Legenda:

	Entrada imediata
	Ao longo da janela de avaliação
	Após a janela de avaliação

Fonte: Registros do Instituto Unibanco sobre os procedimentos e resultados dos sorteios.

Elaboração dos autores.

Obs.: 1. A 2ª etapa no Pará inclui 14 escolas que também fizeram parte da 1ª etapa no Pará. Nesta tabela, essas escolas precisam ser contabilizadas duas vezes. Por esse motivo, o total para o Pará ($50 + 87 = 137$ escolas) nessa tabela é maior que o total apresentado na tabela 3: 113 escolas cujo momento de ingresso no programa foi definido por sorteio.

2. No caso do Ceará, que adotou um sorteio sequencial, nove escolas que deveriam participar do segundo sorteio foram excluídas. Essas escolas participaram do primeiro sorteio, no qual tiveram seu atendimento postergado. Portanto, deveriam participar do segundo sorteio, mas foram excluídas. Assim, nem foram atendidas nem tiveram uma data para atendimento definida por sorteio. Por esse motivo, foram excluídas da contabilidade apresentada nesta tabela. Adicionalmente, duas escolas que deveriam participar do terceiro sorteio também deixaram de participar e por isso também foram excluídas desta tabela.

Vale ressaltar que apenas as escolas envolvidas no sorteio são utilizadas na avaliação. A experiência e as informações sobre as 1,6 mil escolas que estavam programadas para serem atendidas, mas cujo momento de admissão não foi definido por sorteio, não são utilizadas.³²

4 SORTEIO E AGRUPAMENTOS

4.1 Sorteio

Em cada rede, a definição do ano de admissão ao programa das escolas envolvidas na avaliação poderia ter sido feita com base em um único sorteio. Como resultado desse sorteio, seriam definidas, desde o início, quais escolas participariam de imediato no 1º ano do programa, e quais começariam a participar a partir do 2º, 3º e 4º anos, respectivamente. Esse procedimento é denominado *sorteio único*.

32. Em 2013, ocorreu uma expansão do número de escolas atendidas tanto no Pará como no Piauí, em que o momento do atendimento foi definido por sorteio. Esse cronograma de atendimento foi respeitado por dois anos. Em 2015, no entanto, motivados pela oportunidade de adotar a terceira geração, optou-se por "reiniciar" o programa nas redes desses estados. Esse reinício envolveu uma completa redefinição do cronograma de atendimento das escolas. As que estavam sendo atendidas e as que ainda não tinham tido essa oportunidade passaram a ter iguais chances de atendimento imediato. Como a seleção por sorteio realizada em 2013 foi respeitada por dois anos, ela poderia ser utilizada para avaliar o impacto da segunda geração do programa ao longo de seus dois primeiros anos. A despeito dessa possibilidade, a experiência dessas escolas nesse período não é aproveitada neste estudo. Em concordância com essa opção, na tabela 3, essas escolas são tratadas como não tendo a data de atendimento definida por sorteio e, por conseguinte, são excluídas da tabela 4.

Uma alternativa seria realizar uma sequência de sorteios cujo primeiro, anterior ao início do programa na rede, contemplando todas as escolas na avaliação, definisse apenas quais escolas seriam atendidas de imediato. Ao final do 1º ano, haveria um segundo sorteio, limitado àquelas que ainda não haviam sido selecionadas para atendimento, definindo quais seriam atendidas no início do 2º ano. Finalmente, ao término do 2º ano, um terceiro sorteio seria realizado, com as escolas ainda não atendidas, definindo as que deveriam ser atendidas no 3º e 4º anos, respectivamente. Esse procedimento é denominado *sorteio sequencial*.

Na maioria das redes, o ano de admissão das escolas foi definido com base em um *sorteio único*. No Ceará, Piauí e no Espírito Santo, no entanto, o modelo de sorteio adotado foi o *sequencial*. Para efeito da estimação do impacto do programa, esses dois tipos de sorteios são plenamente equivalentes.

4.2 Agrupamentos

O modelo de implantação do programa prevê que as escolas que irão participar do sorteio sejam primeiro organizadas em agrupamentos, do modo mais homogêneo possível, a partir de características observadas, para então terem a ordem de atendimento definida por sorteio dentro de cada agrupamento. Esse procedimento de “sorteio por agrupamento” pode ser tanto do tipo *único* como do tipo *sequencial*. Em qualquer um dos casos, a cada momento, sorteios independentes por agrupamento são realizados. É de fundamental importância ressaltar que os agrupamentos são sempre formados antes da realização dos sorteios, e, quando um processo sequencial de sorteio é adotado, os agrupamentos iniciais são mantidos intactos ao longo de todo o processo.

Embora o sorteio por agrupamentos torne o processo mais complexo, ele traz vantagens tanto para a implementação do programa como para sua avaliação, que justificam sua adoção, a saber: i) maior igualdade de oportunidade; e ii) maior eficiência estatística na estimação da magnitude do impacto do programa.

Do ponto de vista da igualdade de oportunidades, sorteios dentro de agrupamentos são preferíveis, pois fica *garantido* que escolas representando cada um dos agrupamentos estarão presentes, tanto no grupo das que irão participar de imediato como entre as que deverão aguardar atendimento. O processo promove maior igualdade de oportunidades, na medida em que funciona como um sistema de cotas.

4.3 Ganho de precisão devido aos agrupamentos

Escolas podem sempre ser agrupadas procurando-se minimizar a desigualdade entre aquelas no mesmo agrupamento.³³ Dessa forma, escolas no mesmo agrupamento serão mais homogêneas e, graças a essa maior homogeneidade, o uso do agrupamento anterior ao sorteio permite aumentar a precisão das estimativas de impacto.

Com vistas a ilustrar esse argumento, considere-se a situação em que n agrupamentos são formados, cada um por m_0 (escolas a serem atendidas de imediato) e m_1 , que precisam aguardar atendimento. Nesse caso, a média entre agrupamentos das diferenças intra-agrupamentos, entre a média do resultado de interesse entre as escolas com atendimento imediato no agrupamento e a correspondente média entre as escolas aguardando atendimento no mesmo agrupamento é uma estimativa do impacto do programa. A variância desse estimador, λ^2 , é dada por:

$$\lambda^2 = \frac{1}{n} \frac{m_0 + m_1}{m_0 m_1} \tau^2, \quad (1)$$

em que, para simplificar a argumentação, foi assumido que o impacto de programa é o mesmo para todas as escolas e que a variância do resultado de interesse na ausência do programa entre escolas no mesmo agrupamento, τ^2 , é a mesma em todos os agrupamentos.

Na ausência de agrupamentos, o impacto seria estimado pela diferença entre a média do resultado de interesse entre as $n \cdot m_0$ escolas sorteadas para atendimento imediato e a correspondente média entre as $n \cdot m_1$ escolas sorteadas para atendimento futuro, cuja variância é dada por:

$$\delta^2 = \frac{1}{n} \frac{m_0 + m_1}{m_0 m_1} \sigma^2, \quad (2)$$

em que σ^2 denota a variância do resultado de interesse na ausência do programa para o conjunto de todas as escolas.

Dado que a variância total pode ser sempre expressa como a soma da variância intra-agrupamentos com a variância entre os agrupamentos, isto é,

$$\sigma^2 = \tau^2 + Var[\mu_g], \quad (3)$$

33. Dado que a desigualdade total entre as escolas é fixa, esse procedimento equivale a agrupar escolas, com vistas à maximização da desigualdade entre agrupamentos.

em que μ_g denota a média do resultado de interesse na ausência do programa no agrupamento g , segue que $\sigma^2 > \tau^2$, e daí que $\delta^2 > \lambda^2$, exceto se $\mu_g = \mu$ para todo $g = 1, \dots, n$, caso em que não existiria vantagem para a avaliação no agrupamento de escolas antes do sorteio.

Assim, o ganho relativo em precisão devido ao pareamento é dado pela porcentagem da variância do resultado de interesse na ausência do programa, que se deve às diferenças entre agrupamentos:

$$\phi = \frac{Var[\mu_g]}{\sigma^2}, \tag{4}$$

uma vez que segue dessa notação que $\tau^2 = (1 - \phi)\sigma^2$ e, portanto,

$$\frac{\delta^2 - \lambda^2}{\delta^2} = \phi. \tag{5}$$

Não é imediato obter estimativas desse ganho, uma vez que são requeridas estimativas da variância do resultado de interesse na ausência do programa. Uma aproximação, no entanto, pode ser obtida, removendo-se do resultado observado nas escolas atendidas uma estimativa da magnitude do impacto do programa na rede a que a escola pertence.³⁴

TABELA 5
Ganho de precisão devido à utilização de agrupamentos

Geração/rede	Progresso na proficiência ao longo de uma janela de avaliação de três anos (%)		Tamanho médio dos agrupamentos
	Língua portuguesa	Matemática	
1ª geração	46	49	2,7
Minas Gerais	16	24	12,0
São Paulo	52	55	2,0
Rio de Janeiro	70	68	2,0
2ª geração	27	27	6,9
Ceará	18	24	14,5
Goiás	19	14	5,8
Mato Grosso do Sul	16	33	5,0
Pará – 1ª etapa	56	39	2,0
3ª geração	59	53	2,4
Espírito Santo	68	65	3,2
Pará – 2ª etapa	70	64	2,1

(Continua)

34. Outra alternativa seria estimar a variância intra-agrupamento do resultado de interesse na ausência do programa, utilizando-se apenas as escolas selecionadas para receber atendimento após a janela de avaliação. No caso deste estudo, no entanto, essa estratégia não é viável, uma vez que, na vasta maioria dos agrupamentos, apenas uma escola foi selecionada para atendimento após a janela de avaliação.

(Continuação)

Geração/rede	Progresso na proficiência ao longo de uma janela de avaliação de três anos (%)		Tamanho médio dos agrupamentos
	Língua portuguesa	Matemática	
Piauí	39	30	2,0
Todas as gerações	44	43	3,3

Fontes: Arquivos administrativos do Instituto Unibanco; estimativas obtidas a partir dos registros do Instituto Unibanco sobre os procedimentos e resultados dos sorteios e dos sistemas de avaliação estaduais da educação básica.

Minas Gerais – Programa de Avaliação da Educação Básica do Sistema Mineiro de Avaliação e Equidade da Educação Pública – Proeb/Simave (disponível em: <<https://bit.ly/3MgTm3E>>); São Paulo – Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo – Saresp (disponível em: <<https://bit.ly/3NVjLF9>>); Rio de Janeiro – Sistema de Avaliação da Educação do Estado do Rio de Janeiro – Saerj (disponível em: <<https://bit.ly/3xclaAO>>); Ceará – Sistema Permanente de Avaliação da Educação Básica do Ceará – Spaece (disponível em: <<https://bit.ly/3x2xEuF>>); Goiás – Sistema de Avaliação Educacional do Estado de Goiás – Saego (disponível em: <<https://bit.ly/3NPyt0y>>); Mato Grosso do Sul – Sistema de Avaliação da Educação da Rede Pública de Mato Grosso do Sul – Saems (disponível em: <<https://bit.ly/3xaXo9j>>); Pará – Sistema Paraense de Avaliação Educacional – SisPAE (disponível em: <<https://bit.ly/3NVAM2m>>); Espírito Santo – Programa de Avaliação da Educação Básica do Espírito Santo – Paebes (disponível em: <<https://bit.ly/3Mccuj7>>); Piauí – Sistema de Avaliação Educacional do Piauí – Saepi (disponível em: <<https://bit.ly/3NfOnS3>>).

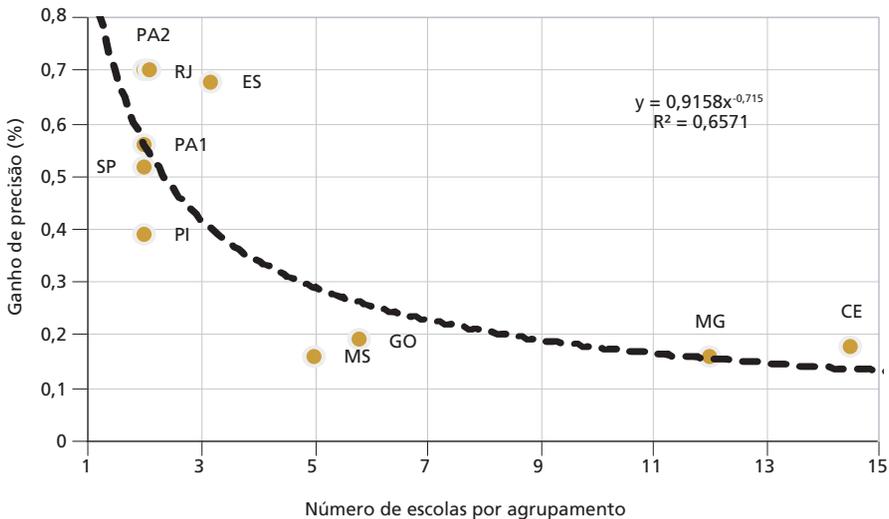
Elaboração dos autores.

Obs.: O ganho de precisão para cada geração foi obtido como a média simples dos correspondentes valores das redes atendidas. O valor para o conjunto das gerações é a média simples dos valores de cada geração

Utilizando-se esse procedimento, estimativas foram obtidas do ganho de eficiência decorrente da utilização do agrupamento anterior ao sorteio (tabela 5). Conforme essa tabela revela, em média, a utilização de agrupamentos é capaz de reduzir a variância da estimativa de impacto do programa por rede em mais de 40%, com essa redução variando substancialmente de acordo com o tamanho dos agrupamentos (gráfico 2) e o procedimento utilizado para construí-los.

GRÁFICO 2

Ganho de precisão devido ao uso de agrupamentos: língua portuguesa



Elaboração dos autores.

Obs.: PA1 = Pará – 1ª etapa; PA2 = Pará – 2ª etapa.

4.4 Procedimentos utilizados para a formação dos agrupamentos

Como resultado do protocolo de implementação do programa Jovem de Futuro, as 1,7 mil escolas que tiveram a data de ingresso no programa definida por sorteio foram previamente organizadas em 533 agrupamentos (tabela 6).

Tanto o tamanho dos agrupamentos como o número de escolas designadas para atendimento imediato, para atendimento ao longo da janela de avaliação e para atendimento apenas após o final da janela de avaliação variam substancialmente entre redes (tabela 6).³⁵ O tamanho dos agrupamentos varia de duas escolas (Rio Grande do Sul, São Paulo, Rio de Janeiro e Pará – 1ª etapa) a mais de dez, nos casos de Minas Gerais e Ceará. No Ceará, o menor agrupamento tinha dez escolas, e o maior, vinte. Em todas as redes, exceto Minas Gerais, uma única escola em cada agrupamento foi selecionada para receber atendimento apenas após o final da janela de avaliação.

Em praticamente todas as redes, os agrupamentos foram obtidos em dois estágios.³⁶ Primeiro, as escolas foram estratificadas em função de sua localização, para, em seguida, dentro de cada estrato espacial, serem agrupadas buscando-se a maior similaridade possível com relação a um conjunto de características. As características selecionadas e o método utilizado para maximizar a similaridade entre escolas em um mesmo agrupamento variou substancialmente entre redes.

Vale ressaltar que as diferenças entre redes, tanto na metodologia utilizada para a construção dos agrupamentos quanto no número de escolas e de agrupamentos e distribuição das escolas no agrupamento, segundo o momento do atendimento, resultaram de especificidades locais, limitações orçamentárias e outras necessidades do processo de implementação do programa. Essas diferenças não refletem necessidades do desenho da avaliação ou representam qualquer impedimento a ele. Vale ressaltar que essa variabilidade na forma de implementação da avaliação ilustra a grande flexibilidade do desenho experimental.

35. Existem também diferenças na configuração de agrupamentos de uma mesma rede, mas essas diferenças são pequenas.

36. O Pará – 1ª etapa é uma exceção: os agrupamentos não tiveram estratificação espacial.

TABELA 6
Quantitativo de escolas participantes do sorteio e agrupamentos formados por rede e geração

Geração	Rede	Número de escolas no sorteio	Número de agrupamentos por tipo						Número de escolas, por agrupamento, segundo o momento do atendimento em relação à janela de avaliação			
			Total	Pares	Trios	Quadras	Quintas	Sextetos	Outros	Início	Meio	Final
1 ^a	Minas Gerais	48	4	-	-	-	-	-	4	5	-	7
	Rio Grande do Sul	50	25	-	-	-	-	-	-	1	-	1
	São Paulo	80	40	-	-	-	-	-	-	1	-	1
	Rio de Janeiro	30	15	-	-	-	-	-	-	1	-	1
	Total	378	26	-	-	-	-	-	26	3 a 6	5 a 13	1
2 ^a	Goiás	145	25	-	-	-	5	20	-	4 a 5	-	1
	Mato Grosso do Sul	125	25	-	-	-	25	-	-	4	-	1
	Pará – 1 ^a etapa	50	25	-	-	-	-	-	-	1	-	1
3 ^a	Espírito Santo	221	70	-	59	11	-	-	-	3 a 2	0 ou 1	0 ou 1
	Pará – 2 ^a etapa	87	42	39	3	-	-	-	-	1 a 2	-	1
	Piauí	275	136	133	3	-	-	-	-	1 a 2	0 ou 1	0 ou 1
	Rio Grande do Norte	243	100	57	43	-	-	-	-	1 a 2	-	1
	Todas as localidades	1.732	533	334	108	11	30	20	30	-	-	-

Fonte: Estimativas obtidas a partir dos registros do Instituto Unibanco sobre os procedimentos e resultados dos sorteios. Elaboração dos autores.

4.5 Agrupamentos utilizados na avaliação

Cada um dos 533 agrupamentos gerados durante a implementação do programa representa uma oportunidade para, ao longo da janela de avaliação, observar-se o desempenho de escolas atendidas diante do desempenho de escolas ainda não atendidas e, portanto, uma oportunidade para se avaliar o impacto do programa.

Nem todos esses agrupamentos, no entanto, puderam ser utilizados com esse propósito, uma vez que sua utilização pressupõe: i) que o resultado de interesse tenha sido medido para as escolas no agrupamento antes e ao final da janela de avaliação; ii) que ao menos uma escola no agrupamento tenha efetivamente sido atendida continuamente desde o início da janela de avaliação; e iii) que ao menos uma escola no agrupamento tenha aguardado atendimento ao longo de toda a janela de avaliação.

A falta de informação sobre o resultado de interesse, no início e ao final da janela de avaliação, levou a que 125 agrupamentos (24%) fossem excluídos da avaliação.³⁷ Esse é o caso dos 100 agrupamentos no Rio Grande do Norte, onde a janela de avaliação vai do início de 2017 ao final de 2019 e, no momento em que este texto foi escrito, medidas para proficiência referentes ao final de 2019 ainda não estavam disponíveis. Também os 25 agrupamentos no Rio Grande do Sul tiveram que ser abandonados, por falta de informação confiável sobre proficiência: no período em que o programa foi implantado, a rede não contava com um sistema anual de avaliação externa da proficiência dos estudantes ao final do ensino médio. A retirada dessas duas redes, embora limite o escopo, não constitui uma ameaça à validade interna da avaliação. Mesmo sem elas, continua sendo possível obter um estimador não enviesado do impacto do programa para as nove redes remanescentes.

Pelo próprio desenho da implementação do programa, em todos os agrupamentos existe ao menos uma escola escolhida para atendimento imediato que foi efetivamente atendida por ao menos três anos consecutivos. O mesmo, no entanto, não se verifica com relação à existência de ao menos uma escola no agrupamento que não tenha sido atendida ao longo de toda a janela de três anos de avaliação. Em 60 agrupamentos, todas as escolas foram atendidas antes do final da janela de avaliação. Esse entrave ocorre na terceira geração nos estados do Piauí e do Espírito Santo. No Piauí, optou-se, em 26 dos 136 agrupamentos, por iniciar o atendimento das últimas escolas ainda não atendidas um ano antes do final da janela de avaliação. No Espírito Santo, a mesma decisão foi tomada, envolvendo 34 dos 70 agrupamentos. Como esses 60 agrupamentos foram escolhidos por sorteio, sua eventual exclusão não representa qualquer ameaça à validade interna da avaliação.

37. Neste estudo, todas as informações do resultado de interesse derivam das avaliações estaduais da educação básica.

Vale ressaltar, no entanto, que esses agrupamentos oferecem informação válida sobre o impacto dos dois primeiros anos do programa. Assim, ao invés de se descartarem completamente esses agrupamentos, foram obtidas estimativas da magnitude do impacto de três anos, combinando-se um estimador da magnitude do impacto dos dois primeiros anos que os inclui com uma estimativa da magnitude do impacto do terceiro ano do programa que os exclui.

TABELA 7
Quantitativo de escolas e agrupamentos descartados, utilizados parcialmente ou plenamente

Disponibilidades e perdas	Rede	Agrupamentos		Escolas	
		Número	%	Número	%
Disponível	-	533	100	1.732	100
Perdas devido à falta de informação	Rio Grande do Sul	25	5	50	3
	Rio Grande do Norte	100	19	243	14
Disponível com informação sobre resultado de interesse no início e no final da janela de avaliação	-	408	77	1.439	83
Todas as escolas foram atendidas antes do final da janela de avaliação	Piauí	26	5	26	2
	Espirito Santo	34	6	34	2
Informação disponível e com alguma escola não atendida até o final da janela de avaliação	-	348	65	1.379	80

Fontes: Estimativas obtidas a partir dos sistemas de monitoramento do atendimento, MFF e SGP, e registros do Instituto Unibanco sobre os procedimentos e resultados dos sorteios.
Elaboração dos autores.

Assim, sem qualquer ameaça à validade interna da avaliação, 185 agrupamentos foram ou descartados (125) ou utilizados apenas parcialmente (60). Trata-se, portanto, de uma perda completa de 24% e parcial de 11% dos agrupamentos inicialmente disponíveis (tabela 7). Em relação às escolas, as perdas são relativamente menores, 17% e 4%, respectivamente.

Essas perdas, infelizmente, não foram as únicas. Em todas as redes, ao longo dos três anos da janela de avaliação, escolas foram fechadas, não realizaram a avaliação externa ou abandonaram o programa por motivos não previstos no plano original de implementação. Por esses motivos, 26 agrupamentos e, por conseguinte, 59 escolas tiveram que ser retirados da avaliação (tabela 8).³⁸ Caso fosse possível confirmar que os eventos (fechamento de escola, não participação na avaliação etc.) que levaram à retirada desses 26 agrupamentos da análise teriam ocorrido mesmo na ausência do programa, sua exclusão não teria consequências sobre a validade interna da avaliação. Não há, no entanto, evidência de que a ocorrência desses eventos tenha sido independente da implantação do programa. A exclusão desses agrupamentos e

38. Adicionalmente, catorze escolas (1% do total) em agrupamentos incluídos foram retiradas pelos mesmos motivos (tabela 8).

escolas pode estar, em alguma medida, enviesando os resultados obtidos na avaliação. Como apenas 6% dos agrupamentos e 5% das escolas precisaram ser excluídos, o viés que podem causar deve ser correspondentemente limitado.

TABELA 8
Disponibilidade e perdas de agrupamentos e escolas

Rede	Agrupamentos			Escolas			Total de perdas (%)
	Potencialmente disponíveis	Perdas	Perdas (%)	Potencialmente disponíveis	Perdas vinculadas à exclusão de agrupamentos	Perdas em agrupamentos não excluídos	
1ª geração	59	6	10	158	12	1	8
Minas Gerais	4	0	0	48	0	1	2
São Paulo	40	3	8	80	6	0	8
Rio de Janeiro	15	3	20	30	6	0	20
2ª geração	101	12	12	689	30	12	6
Ceará	26	0	0	369	0	9	2
Goiás	25	0	0	145	0	2	1
Mato Grosso do Sul	25	2	8	125	10	1	9
Pará – 1ª etapa	25	10	40	50	20	0	40
3ª geração	248	8	3	583	17	1	3
Piauí	136	5	4	275	10	0	4
Pará – 2ª etapa	42	3	7	87	7	0	8
Espírito Santo	70	0	0	221	0	1	0
Total	408	26	6	1.430	59	14	5

Fontes: Estimativas obtidas a partir dos sistemas de monitoramento do atendimento, MFF e SGP, e registros do Instituto Unibanco sobre os procedimentos e resultados dos sorteios.
Elaboração dos autores.

5 METODOLOGIA

Para cada um dos 382 agrupamentos efetivamente disponíveis para a avaliação, um estimador não enviesado do impacto pode ser obtido pela diferença entre o desempenho médio dos estudantes nas escolas atendidas e não atendidas. Esse estimador, embora não enviesado, tem baixa precisão, devido ao limitado número de escolas em cada agrupamento. Estimadores mais precisos por rede de ensino podem ser obtidos como a média dos estimadores para os diversos agrupamentos que fazem parte da rede. O impacto do programa para o conjunto das redes pode, então, ser obtido com a média ponderada dos estimadores obtidos por rede. Esta seção descreve, sequencialmente, a metodologia utilizada para obter estimadores do impacto por agrupamento, por rede e global (conjunto das redes).

5.1 Estimando o impacto por agrupamento

Para efeito dessa avaliação, a origem do tempo é ajustada por agrupamento, de tal forma que $t = 0$ denota o ano letivo imediatamente anterior ao início do atendimento e $t = j$ denota o j -ésimo ano letivo desde o início do programa em escolas no agrupamento. Considera-se, como janela de avaliação, J os primeiros três anos da adoção do programa em um agrupamento. Assim, $J = \{1,2,3\}$. O objetivo desta subseção é obter um estimador não enviesado do impacto do Jovem de Futuro para cada agrupamento em cada momento, ao longo da janela de avaliação, isto é, para cada $t \in J$. Como toda a análise é realizada por agrupamento, a notação deveria estar correspondentemente indexada por agrupamento. No entanto, para simplificar a exposição, esse indexador é omitido.

Considera-se, como parte do grupo de beneficiários, B , apenas as escolas que foram atendidas continuamente ao longo de toda a janela de avaliação, isto é, desde o início da adoção do programa, $t = 1$. Esse conjunto, portanto, não varia ao longo da janela de avaliação e, como vimos,³⁹ não é vazio, $B \neq \emptyset$. Por sua vez, o conjunto das escolas atendidas apenas após o ano t , \bar{B}_t ,⁴⁰ pode variar ao longo da janela de avaliação e se tornar, inclusive, vazio, antes do final da janela de avaliação.⁴¹ Para que um agrupamento seja útil à avaliação, ao menos algumas escolas no agrupamento precisam não ser atendidas desde o início, isto é, precisa-se, ao menos, ter que $\bar{B}_1 \neq \emptyset$. Note-se que $\bar{B}_t \subseteq \bar{B}_s$ sempre que $t \geq s$. Na medida em que parte das escolas inicia seu atendimento ao longo da janela de avaliação, o conjunto das escolas ainda não atendidas efetivamente declina.

Para cada escola i e ano t , d_{it} indica se essa escola foi sorteada para iniciar seu atendimento em algum ano até t (inclusive),⁴² $d_{it} = 1$, ou se foi sorteada para atendimento apenas após o ano letivo t (exclusive), $d_{it} = 0$. Dessa forma, $i \in \bar{B}_t$, se e somente se $d_{it} = 0$, e $i \in B$, se e somente se $d_{i1} = 1$. O conjunto de escolas úteis para a avaliação do impacto do t -ésimo ano do programa é dado por $\bar{B}_t \cup B$.⁴³ Na medida em que algumas escolas podem iniciar seu atendimento ao longo da janela de avaliação, $\bar{B}_t \cup B$ pode estar estritamente contido no conjunto de todas as escolas no agrupamento.⁴⁴

39. Como relatado na seção anterior, essa hipótese é válida para todos os 382 agrupamentos a serem utilizados na avaliação.

40. Esse conjunto inclui todas as escolas atendidas a partir do início do ano letivo $t + 1$.

41. De fato, para 26 agrupamentos na rede do Piauí e 34 na rede do Espírito Santo, $\bar{B}_3 = \emptyset$.

42. Nessa avaliação, toda escola que ingressa no programa permanece ao menos até o final da janela de avaliação. Logo, se $d_{it} = 1$, então temos necessariamente que $d_{is} = 1$ para todo $s \geq t$.

43. Implicitamente, está sendo considerado que o impacto do t -ésimo ano do programa num agrupamento é o impacto desse ano letivo sobre escolas que participam continuamente desde o primeiro ano do programa nesse agrupamento. É, portanto, o impacto do t -ésimo ano sobre escolas que participam do programa em t anos. Escolas que participaram do t -ésimo ano do programa, mas iniciaram sua participação somente após o primeiro ano, não são consideradas na avaliação do impacto do t -ésimo ano do programa.

44. Vale ressaltar, no entanto, que $\bar{B}_1 \cup B$ forma o conjunto de todas as escolas no agrupamento.

Adicionalmente, para cada escola i no agrupamento, denota-se por y_{it}^b o progresso no resultado de interesse ao longo do ano t para essa escola i caso tenha sido beneficiada (b) pelo programa continuamente desde o início da janela de avaliação, e por y_{it}^a , o progresso no mesmo resultado para essa mesma escola i ao longo do mesmo ano t , caso só venha a ter acesso ao programa em algum momento após (a) o final do ano t .⁴⁵

Com base nessa notação, o progresso observado (o) no resultado de interesse para a escola i ao longo do ano t , y_{it}^o , entre as escolas úteis à avaliação de impacto no ano t , isto é, $i \in \bar{B}_t \cup B$, é dado por

$$y_{it}^o = (1 - d_{it})y_{it}^a + d_{it} \cdot y_{it}^b, \quad (6)$$

desde que $i \in \bar{B}_t \cup B$.

Supondo que a magnitude do impacto ao longo do ano t , β_t , seja homogênea, isto é, que o programa influencie o progresso no resultado de interesse de todas as escolas na mesma magnitude, segue que $y_{it}^b = y_{it}^a + \beta_t$.

Nesse caso, a diferença entre o progresso médio ao longo do ano t alcançado pelas m_0 escolas sorteadas para atendimento imediato, $m_0 = \#B$, e o progresso médio ao longo do ano t das m_t escolas sorteadas para atendimento apenas após t anos, $m_t = \#\bar{B}_t$, que é denominado de $\hat{\beta}_t$, é um estimador não enviesado da magnitude do impacto ao longo do ano t , β_t . De fato, como

$$\hat{\beta}_t = \frac{1}{m_0} \sum_{i \in B} y_{it}^o - \frac{1}{m_t} \sum_{i \in \bar{B}_t} y_{it}^o, \quad (7)$$

$y_{it}^o = y_{it}^a$ se $i \in B_t$, e $y_{it}^o = y_{it}^b$ se $i \in B$, segue que

$$\hat{\beta}_t = \frac{1}{m_0} \sum_{i \in B} y_{it}^b - \frac{1}{m_t} \sum_{i \in \bar{B}_t} y_{it}^a, \quad (8)$$

como $y_{it}^b = y_{it}^a + \beta_t$, segue que

$$\hat{\beta}_t = \beta_t + \frac{1}{m_0} \sum_{i \in B} y_{it}^a - \frac{1}{m_t} \sum_{i \in \bar{B}_t} y_{it}^a, \quad (9)$$

e, portanto,

$$E[\hat{\beta}_t] = \beta_t + E[y_{it}^a | i \in B] - E[y_{it}^a | i \in \bar{B}_t]. \quad (10)$$

45. O progresso para uma escola ao longo do ano t , caso ela tenha tido acesso ao programa no meio da janela de avaliação (isto é, em algum ano s anterior a t , mas posterior ao início da janela de avaliação, $t > s > 1$) não está sendo aqui especificado. Considerar esses casos não é necessário, dado que as escolas que aderem ao programa no ano letivo $1 < s < t$ só são úteis à avaliação de impacto até o ano $s - 1$.

Como, em cada agrupamento, o momento do atendimento das escolas foi escolhido ao acaso, segue que

$$E[y_{it}^a | i \in B] = E[y_{it}^a | i \in \bar{B}_t] = E[y_{it}^a], \quad (11)$$

e daí, que $E[\hat{\beta}_t] = \beta_t$, como se queria mostrar. Adicionalmente,

$$\text{Var}[\hat{\beta}_t] = \text{Var} \left[\frac{1}{m_0} \sum_{i \in B} y_{it}^a - \frac{1}{m_t} \sum_{i \in \bar{B}_t} y_{it}^a \right] = \left(\frac{m_0 + m_t}{m_0 m_t} \right) \text{Var}[y_{it}^a]. \quad (12)$$

Na sequência, apresenta-se a metodologia utilizada para obter-se a média por rede desses estimadores por agrupamento. Para isso, torna-se necessário utilizar explicitamente um indexador de agrupamento, g . Assim, o estimador do impacto do programa, ao longo do ano t , referente no agrupamento g , passa a ser denotado por $\hat{\beta}_{gt}$.

5.2 Estimando o impacto por rede de ensino

Denota-se por E_{rt} o conjunto de agrupamentos da rede r relevantes para a avaliação do impacto do programa ao longo do ano t , $t \in J$. Esse conjunto é formado por todos os agrupamentos na rede r , para os quais, ao final do ano t , ainda existiam escolas não atendidas, isto é, agrupamentos para os quais $\bar{B}_{gt} \neq \emptyset$, portanto, $m_{gt} > 0$.

Assumindo-se que a magnitude do impacto, ao longo do ano t , é a mesma em todos os agrupamentos da mesma rede r , segue que a média, $\hat{\beta}_{rt}$, dos estimadores obtidos para cada agrupamento, $\hat{\beta}_{gt}$, também é um estimador não enviesado dessa magnitude:

$$\hat{\beta}_{rt} = \frac{1}{n_{rt}} \sum_{g \in E_{rt}} \hat{\beta}_{gt}, \quad (13)$$

em que $n_{rt} = \#E_{rt}$ denota o número de agrupamentos da rede r relevantes para a avaliação do impacto ao longo do ano t .

Uma vez obtidos estimadores não enviesados para o impacto do programa em cada rede r ao longo de cada ano t na janela de avaliação, um estimador por rede r do impacto ao longo de toda a janela de avaliação, denominado de $\hat{\beta}_r$, pode ser obtido pela soma:

$$\hat{\beta}_r = \hat{\beta}_{r1} + \hat{\beta}_{r2} + \hat{\beta}_{r3}. \quad (14)$$

Na medida em que o tamanho e a natureza dos agrupamentos em uma mesma rede são muito similares, segue que a covariância entre $\hat{\beta}_{rs}$ e $\hat{\beta}_{rt}$, $\text{Cov}[\hat{\beta}_{rs}, \hat{\beta}_{rt}] = \sigma_{rst}$ pode ser obtida para todo s e t , tal que $s \leq t$, via⁴⁶

46. Note que para obter essa expressão utilizamos o fato que $B_t \subseteq B_s$ sempre que $s \leq t$.

$$\hat{\sigma}_{rst} = \frac{1}{n_{rt} - 1} \sum_{g \in E_{rt}} (\hat{\beta}_{gs} - \hat{\beta}_{rs}) (\hat{\beta}_{gt} - \hat{\beta}_{rt}). \quad (15)$$

Dessa forma, um estimador de $Var [\hat{\beta}_r] = \tau_r^2$ pode ser obtido via

$$\hat{\tau}_r^2 = \sum_{t=1}^3 \hat{\sigma}_{rtt} + 2 \sum_{t=1}^2 \sum_{s=t+1}^3 \hat{\sigma}_{rst}. \quad (16)$$

5.3 Estimando o impacto para o conjunto das redes

Por fim, sob a hipótese de que o impacto é homogêneo (isto é, de igual magnitude em todas as redes), um estimador não enviesado da magnitude do impacto do Jovem de Futuro para o conjunto das redes de ensino, $\hat{\beta}$, pode ser obtido via qualquer uma das combinações lineares

$$\hat{\beta} = \frac{\sum_r \alpha_r \hat{\beta}_r}{\sum_r \alpha_r}. \quad (17)$$

Dado que os estimadores por rede, $\hat{\beta}_r$, são estocasticamente independentes, segue que, sob a hipótese de que o impacto é homogêneo, a combinação linear que leva ao estimador de menor variância é aquela em que

$$\alpha_r = \frac{1}{\hat{\tau}_r^2}. \quad (18)$$

Neste caso,

$$Var [\hat{\beta}] = \frac{1}{\sum_r \frac{1}{\hat{\tau}_r^2}}. \quad (19)$$

6 A NATUREZA E MAGNITUDE DO IMPACTO DO PROGRAMA SOBRE O APRENDIZADO

6.1 A natureza do impacto do programa sobre o aprendizado

Neste estudo, estima-se a magnitude do impacto de três anos de exposição ao Jovem de Futuro sobre a proficiência em matemática e língua portuguesa ao final da 3ª série do ensino médio.⁴⁷ É importante ressaltar que a *escola* é a unidade de intervenção do programa e, também, a unidade utilizada para medir seu impacto. Assim, o que se busca avaliar é o impacto do programa sobre o desempenho da *escola*, medido pela proficiência média de seu corpo discente ao final da 3ª série do ensino médio.

47. Vale ressaltar que o impacto do programa sobre o *nível* da proficiência ao final da 3ª série é idêntico a seu impacto sobre o *progresso* na proficiência. Afinal o programa não pode ter impacto sobre o valor da proficiência anterior a seu início.

Embora fosse extremamente importante medir o impacto do programa sobre a proficiência de um dado grupo de estudantes – por exemplo, aqueles que iniciaram a 1ª série no início da janela de avaliação –, não é esse impacto que a avaliação é capaz de identificar. Devido à evasão, à repetência e às transferências entre escolas, nem todos os estudantes concluintes da 3ª série em um dado ano estavam na 1ª série da mesma escola três anos antes. Na medida em que esses fluxos escolares são influenciados pelo programa, o impacto estimado para a escola capta tanto o impacto do programa sobre o *aprendizado* quanto seu impacto sobre a *composição* do corpo discente.

Vale também ressaltar que, mesmo que o programa não tivesse qualquer impacto sobre evasão, repetência e transferências, o impacto estimado neste estudo não seria necessariamente igual ao impacto que o programa tem sobre o aprendizado dos estudantes que estavam na 1ª série no início da janela de avaliação. Para que esses dois impactos fossem os mesmos, seria necessário que todos os fluxos escolares ocorressem ao acaso, garantindo que a composição do corpo discente no início da 1ª série fosse estatisticamente equivalente à alcançada três anos mais tarde, ao final da 3ª série.

6.2 A magnitude do impacto do programa sobre o aprendizado

A tabela 9 apresenta estimativas da magnitude do impacto de três anos de exposição ao Jovem de Futuro sobre o desempenho escolar, medido pela proficiência média do corpo discente ao final da 3ª série do ensino médio, por rede, para cada uma das três gerações que marcam a evolução do programa ao longo da década (2008-2018) e, também, para o conjunto das três gerações.⁴⁸ Para isso, utiliza-se a experiência de 1,4 mil escolas em nove redes estaduais. Para efeito da implementação e avaliação do programa, essas 1,4 mil escolas foram organizadas em 382 agrupamentos. Destes agrupamentos, 322 permitem estimar o impacto de três anos de exposição, enquanto os 60 restantes complementam permitindo aprimorar as estimativas do impacto dos dois primeiros anos de exposição ao programa.

Beneficiaram-se da primeira geração cerca de 200 escolas em quatro redes estaduais. Nessa avaliação, utilizamos a experiência de 145 escolas em três redes de ensino organizadas em 53 agrupamentos (tabela 9). As estimativas obtidas apontam para um impacto de 6 e 5 pontos na escala Saeb em língua portuguesa e matemática, respectivamente. Ambas as estimativas são estatisticamente significativas ao nível de 1% (tabela 9).

48. A proficiência média em língua portuguesa e matemática dos estudantes ao final da 3ª série do ensino médio, para cada escola na avaliação, foi extraída dos sistemas estaduais de avaliação da educação básica de cada rede na avaliação: Proeb, Saesp, Saerj, Spaace, Saego, Saems, SisPAE, Saepi e Paebes, conforme detalhado nas fontes das tabelas 5 e 9.

Na segunda geração, o programa beneficiou 2,2 mil escolas em cinco estados (tabela 2). Da avaliação, participaram cerca de 650 escolas de quatro redes estaduais, organizadas em 89 agrupamentos (tabela 9). A magnitude dos impactos estimados sobre a proficiência foi de 6 pontos na escala Saeb, tanto em língua portuguesa quanto em matemática – portanto, um pouco acima do obtido na primeira geração. Todas as estimativas obtidas para o impacto são estatisticamente significativas ao nível de 1% (tabela 9).

TABELA 9
Impactos sobre o desempenho em língua portuguesa e matemática por geração e localidade

Geração/localidade	Número de agrupamentos	Língua portuguesa			Matemática		
		Impacto médio (pontos na escala Saeb)	Erro-padrão (pontos na escala Saeb)	p-valor (%)	Impacto médio (pontos na escala Saeb)	Erro-padrão (pontos na escala Saeb)	p-valor (%)
Todas as gerações	382	4,4	0,9	0	4,8	0,8	0
1ª geração	53	5,5	2,1	0	5,3	1,7	0
Minas Gerais	4	9,7	4,7	2	11,8	6,0	2
São Paulo	37	4,4	2,5	4	4,4	1,9	1
Rio de Janeiro	12	5,5	6,8	21	6,9	5,0	8
2ª geração	89	5,6	1,6	0	5,8	1,4	0
Ceará	26	7,9	2,4	0	7,1	2,3	0
Goiás	25	5,2	4,0	10	5,5	3,9	8
Mato Grosso do Sul	23	3,1	3,4	18	4,0	2,1	3
Pará – 1ª etapa	15	2,0	4,6	33	10,1	4,7	2
3ª geração	240	3,1	1,4	1	3,7	1,2	0
Espírito Santo	70	2,5	2,4	15	4,7	2,9	5
Pará – 2ª etapa	39	4,8	2,8	4	3,5	2,1	5
Piauí	131	2,6	2,1	11	3,4	1,8	3

Fontes: Estimativas obtidas a partir dos sistemas de monitoramento do atendimento do programa, MFF e SGP, e dos registros do Instituto Unibanco sobre os procedimentos e resultados dos sorteios e dos dados dos sistemas de avaliação estaduais da educação básica. Minas Gerais – Proeb/Simave (disponível em: <<https://bit.ly/3MgTm3E>>); São Paulo – Saresp (disponível em: <<https://bit.ly/3NVJLF9>>); Rio de Janeiro – Saerj (disponível em: <<https://bit.ly/3xclaAO>>); Ceará – Spaec (disponível em: <<https://bit.ly/3x2xEuF>>); Goiás – Saego (disponível em: <<https://bit.ly/3NPYt0y>>); Mato Grosso do Sul – Saems (disponível em: <<https://bit.ly/3xaXo9j>>); Pará – SisPAE (disponível em: <<https://bit.ly/3NVAM2m>>); Espírito Santo – Paebs (disponível em: <<https://bit.ly/3Mccuj7>>); Piauí – Saepi (disponível em: <<https://bit.ly/3NfOnS3>>).

Por fim, na terceira geração foram beneficiadas cerca de 2,1 mil escolas em seis estados. Participaram da avaliação mais de 550 escolas de três redes estaduais, organizadas em 240 agrupamentos. As estimativas obtidas para essa geração apontam para um impacto de magnitude um pouco menor: 3 e 4 pontos na

escala Saeb em língua portuguesa e matemática, respectivamente. A precisão dessas estimativas é maior que nas gerações anteriores, devido ao aumento no número de agrupamentos utilizados. No entanto, como a magnitude estimada é menor, o *p*-valor é ligeiramente maior, mas, mesmo assim, da ordem de (ou inferior a) 1%.

Agregando as experiências das três gerações, obtêm-se impactos globais de 4 e 5 pontos na escala Saeb sobre a proficiência em língua portuguesa e em matemática, respectivamente, ambos estatisticamente significantes ao nível de 1% (tabela 9). Esses impactos sobre a proficiência em língua portuguesa e em matemática correspondem, respectivamente, a 9% e 12% de um desvio-padrão, uma vez que o desvio-padrão do aprendizado entre estudantes em língua portuguesa e matemática, ao final do ensino médio, é de 48 e 40 pontos, respectivamente.⁴⁹

6.3 Diferenças na magnitude do impacto do programa entre gerações

Com vistas a interpretar a evolução da magnitude do impacto de um programa ao longo do tempo, é conveniente reconhecer a interveniência de três fatores. Em primeiro lugar, a experiência com o programa deve levar a aperfeiçoamentos e, portanto, a aumentos na magnitude do impacto.

Em segundo lugar, na medida em que aumenta a escala da operação, a qualidade da implementação deve sofrer em alguma medida, levando a uma queda na magnitude do impacto. No caso do Jovem de Futuro, ocorreu um grande aumento na escala de operação entre a primeira e a segunda geração (o número de escolas atendidas passou de 200 para 2,2 mil), mas também um significativo aperfeiçoamento do desenho. O ligeiro aumento na magnitude do impacto entre essas gerações aponta para a validade da hipótese de que os ganhos devidos ao aperfeiçoamento devem ter suplantado eventuais perdas decorrentes do aumento na escala.

Em terceiro lugar, deve-se sempre reconhecer que o impacto medido de um programa raramente é absoluto. Invariavelmente, o que se mede é o impacto do programa em relação a uma alternativa. À medida que o tempo passa, é esperado que programas similares sejam desenvolvidos. Assim, espera-se que o impacto medido de um programa inovador vá declinando ao longo do tempo, na medida em que suas características inovadoras vão sendo adotadas pelas alternativas (Duncan e Magnuson, 2013). No caso do Jovem de Futuro, embora o progresso das alternativas possa ser um candidato para explicar o pequeno declínio do seu impacto entre a segunda e a terceira geração, o substancial aprimoramento no desenho do programa, ocorrido entre essas duas gerações, torna a validade dessa conjectura pouco provável.

49. Esse desvio-padrão refere-se à distribuição da proficiência dos estudantes ao final da 3ª série do ensino médio da rede pública no Brasil segundo o Saeb 2015.

No caso do Jovem de Futuro, a explicação mais provável para o pequeno declínio na magnitude do impacto entre a segunda e a terceira geração é o crescimento no impacto do programa sobre o fluxo escolar (Henriques, Carvalho e Barros, 2020). Conforme esses autores demonstram, o impacto do Jovem de Futuro de terceira geração sobre a taxa de aprovação é muito superior ao encontrado na primeira ou segunda geração. Na medida em que os que não seriam aprovados na ausência do programa têm uma proficiência em média inferior à dos que seriam aprovados independentemente do programa, um aumento na aprovação deve levar a uma redução na proficiência dos aprovados e a uma medida de impacto de menor magnitude.

A comparação, entre gerações, das estimativas da magnitude do impacto do Jovem de Futuro claramente aponta para uma tendência em “U invertido”: a magnitude do impacto cresce da primeira para a segunda geração e decresce da segunda para a terceira geração. Essas comparações, no entanto, devem ser tratadas com extrema cautela, uma vez que as diferenças encontradas não são estatisticamente significativas. De fato, como nas três gerações o erro-padrão das estimativas da magnitude do impacto fica sempre em torno de 1,5 ponto na escala Saeb, e as diferenças observadas nunca superam 2,5 pontos, não é possível rejeitar a hipótese de que o impacto é invariante, mesmo assumindo-se erros de até 30%. De fato, o *p*-valor do teste Q de Cochran para homogeneidade do impacto entre as três gerações é de 43% para língua portuguesa e de 50% para matemática (tabela 10).⁵⁰

TABELA 10
Teste para heterogeneidade do impacto entre gerações

Matéria	Estatística	Geração				<i>p</i> -valor (%)
		Primeira	Segunda	Terceira	Todas	
Língua portuguesa	Impacto (pontos na escala Saeb)	5,5	5,6	3,1	4,4	–
	Erro-padrão (pontos na escala Saeb)	2,1	1,6	1,4	0,9	–
	Q-Cochran	–	–	–	–	43
Matemática	Impacto (pontos na escala Saeb)	5,3	5,8	3,7	4,8	–
	Erro-padrão (pontos na escala Saeb)	1,7	1,4	1,2	0,8	–
	Q-Cochran	–	–	–	–	50

Fontes: Estimativas obtidas a partir dos sistemas de monitoramento do atendimento do programa, MFF e SGP, e dos registros do Instituto Unibanco sobre os procedimentos e resultados dos sorteios e dos dados dos sistemas de avaliação estaduais da educação básica.

50. Também as disparidades entre redes não apontam para qualquer heterogeneidade do impacto do programa. De fato, o *p*-valor do teste Q de Cochran para homogeneidade do impacto entre as dez redes (note-se que são nove redes, mas, quanto ao Pará, temos duas estimativas) é de 81% para língua portuguesa e de 84% para matemática.

6.4 Relevância da magnitude do impacto estimado

Em que medida um impacto de 5 pontos na escala Saeb, ou 10% de um desvio-padrão, pode ser considerado *sucesso* para um programa de promoção da gestão escolar como o Jovem de Futuro? Nesta subseção, essa questão é encaminhada com base em quatro abordagens complementares.

Em primeiro lugar, vale relembrar o gráfico 1, apresentado no início deste estudo. Esse gráfico revela que a proficiência dos estudantes brasileiros poderia ser 64% de um desvio-padrão melhor se fosse possível alcançar o que outros países com gasto similar por estudante já alcançam. Se considerarmos essa marca como o potencial de melhorias na gestão educacional, o impacto do Jovem de Futuro (10% de um desvio-padrão) representaria 16% desse potencial. Assim, por um lado, os resultados dessa avaliação revelam a capacidade do Jovem de Futuro de promover maior proficiência via melhorias em gestão; por outro lado, esses mesmos resultados também revelam que o programa definitivamente não exaure as possibilidades que uma melhor gestão pode trazer para a proficiência dos estudantes.

Em segundo lugar, uma forma alternativa de contextualizar a eficácia do Jovem de Futuro consiste em contrastar a magnitude do seu impacto com o obtido em avaliações de iniciativas similares, como as realizadas em Fryer Junior (2014; 2017). O resultado dessa comparação revela que o impacto do programa (10% de um desvio-padrão) se encontra dentro do espectro de estimativas obtidas por Fryer Junior (2017) (10% a 20% de um desvio-padrão), mas um pouco abaixo das obtidas por Fryer Junior (2014) (15% a 20% de um desvio-padrão).

Uma terceira abordagem para balizar a eficácia do programa consiste em comparar a magnitude do seu impacto com o que tipicamente um estudante aprende ao longo das três séries do ensino médio. Em média, em uma escola pública brasileira, o ganho “aparente” de proficiência ao longo do ensino médio é de 15 pontos na escala Saeb.⁵¹ Assim, em uma escola típica que adote o Jovem de Futuro, esse ganho de proficiência passaria de 15 para 20 pontos: um significativo adicional de 33%.⁵² Nessa métrica, o impacto do Jovem de Futuro é equivalente ao aprendizado médio ao longo de um ano letivo adicional de ensino médio.

51. Esse ganho *aparente* de proficiência ao longo do ensino médio é obtido subtraindo-se da proficiência ao final do ensino médio em 2017 a proficiência ao final do ensino fundamental em 2014, uma tentativa de comparar coortes similares. Dada a ausência da Prova Brasil em 2014, a proficiência para esse ano foi aproximada utilizando-se a média para os anos de 2013 e 2015. A estimativa é a média simples do estimado para língua portuguesa (13 pontos na escala Saeb) com o estimado para matemática (17 pontos na escala Saeb).

52. Como o desvio-padrão em proficiência, ao final do ensino médio, fica entre 48 e 40 pontos (em língua portuguesa e matemática, respectivamente), 44 pontos em média; 15 pontos na escala Saeb representam 34% de um desvio-padrão. Embora o ganho de proficiência, ao longo do ensino médio brasileiro, seja definitivamente limitado, a estimativa para os Estados Unidos não fica muito acima, sendo de 42% de um desvio-padrão, segundo Hill *et al.* (2008).

Por fim, vale observar que, segundo Inep (2019b), três anos em uma escola pública de ensino médio custava, em 2009, R\$ 7 mil⁵³ por estudante. Segundo Vasconcellos (2015, p. 16), nesse ano, o correspondente custo de três anos de Jovem de Futuro era de R\$ 788,87⁵⁴ por estudante. Segue que o Jovem de Futuro, a um custo equivalente a 11% do que se gasta com o ensino médio, é capaz de promover um ganho de aprendizado da ordem de 33% do que é alcançado ao longo do ensino médio (15 pontos). Esse resultado demonstra que o custo de um ponto adicional em proficiência promovido pelo Jovem de Futuro (R\$ 190 por ponto na escala Saeb)⁵⁵ é inferior à metade do correspondente custo das escolas públicas de ensino médio (R\$ 467 por ponto adicional na escala Saeb).⁵⁶

Em suma, embora o Jovem de Futuro tenha um impacto ligeiramente inferior ao alcançado internacionalmente por programas similares, dado o limitado desempenho das escolas públicas brasileiras, esse impacto representa uma inestimável contribuição à educação brasileira. No entanto, muito mais ainda se pode esperar de futuras versões do programa, uma vez que o impacto estimado representa apenas 16% do aparente impacto potencial que melhorias em gestão podem trazer para a proficiência dos estudantes.

7 CONCLUSÕES E RECOMENDAÇÕES

O Jovem de Futuro foi inicialmente desenhado para promover melhorias na gestão escolar por meio da incorporação de uma mentalidade voltada para resultados. Progressivamente, o programa incorporou a oferta de protocolos e práticas estruturadas, além de introduzir um circuito de gestão voltado para resultados, com especial atenção às correções de rumo e à promoção da troca de experiências entre escolas. Em sua terceira geração, o programa aumentou seu escopo, dirigindo a atenção também para as melhorias na gestão da rede estadual e suas regionais.

Embora o Jovem de Futuro tenha por objetivo influenciar múltiplos aspectos da gestão, seu foco central é o planejamento estratégico, sua execução, avaliação e correção de rumos. Assim, não são diretamente tratados pelo Jovem de Futuro aspectos ligados à governança e à descentralização, nem aspectos ligados ao aperfeiçoamento das rotinas de gestão.

53. Essa estimativa se refere ao custo de três anos de ensino médio em 2009, obtida na publicação *Investimento público direto cumulativo por estudante* de 2015 (Inep, 2019b). Usamos a estimativa para 2009 em valores nominais para maior comparabilidade com a estimativa de custo do Jovem de Futuro que se refere a esse ano.

54. Esse custo está expresso em reais (2008) e se refere à primeira geração. Portanto, como mencionado na seção 2.2, inclui uma transferência às escolas de R\$ 300 por estudante matriculado (R\$ 100 por série). Como essa transferência deixou de estar presente na terceira geração, essa estimativa de custo provavelmente muito superestima o custo atual (terceira geração) do Jovem de Futuro por estudante.

55. Isto é, R\$ 190 = R\$ 875/4,6 pontos na escala Saeb, impacto médio do Jovem de Futuro sobre a proficiência em língua portuguesa e matemática.

56. Isto é, R\$ 467 = R\$ 7 mil/15 pontos na escala Saeb, aprendizado *aparente* ao longo do ensino médio.

Ações desenhadas para a melhoria da gestão deveriam, teoricamente, ser capazes de promover maior eficiência. Embora extremamente limitada, a evidência disponível sobre a eficácia desse tipo de ação (Bloom *et al.*, 2015; Fryer Junior, 2014; 2017) é encorajadora, por apontar que ações adequadamente desenhadas para promover uma melhor gestão tendem a ser eficazes em reduzir significativamente a ineficiência no uso dos recursos educacionais. Assim, uma iniciativa como o Jovem de Futuro é de grande relevância em um país onde os resultados educacionais alcançados são ruins e desiguais, a despeito de um gasto público com educação elevado e crescente.

Os resultados da avaliação do impacto do Jovem de Futuro são encorajadores. Esses resultados apontam para impactos, sobre a proficiência ao final da 3ª série do ensino médio em língua portuguesa e matemática, da ordem de 4 a 5 pontos na escala Saeb, respectivamente, o que equivale a 9% e 12% de um desvio-padrão.

No que concerne a essa magnitude, podem ser feitas as considerações a seguir.

- 1) Está dentro do esperado, dada a evidência internacional com iniciativas similares (Fryer Junior, 2014; 2017).
- 2) Corresponde a um aumento significativo, em comparação ao que um estudante típico aprende ao longo do ensino médio na rede pública brasileira.
- 3) Leva a uma relação custo-efetividade mais favorável que a dos demais gastos com o ensino médio público.
- 4) Representa uma ajuda de imenso valor a um sistema educacional no qual o progresso lento vem se tornando a regra.

No entanto, diante do elevado grau total da *aparente ineficiência* (0,64 desvio-padrão) do sistema educacional brasileiro, o impacto do Jovem de Futuro é uma contribuição apenas parcial. Se adotado em todas as escolas públicas brasileiras, seria capaz de reduzir em 16% as *aparentes ineficiências* existentes. Essa contribuição abre caminho para outras iniciativas em gestão, complementares ou mais abrangentes, que possam trazer contribuições ainda maiores para a melhoria do desempenho educacional brasileiro.

Em suma, este estudo, por um lado, revela que a ineficiência no uso dos recursos educacionais, que tanto prejudica o desempenho educacional brasileiro, pode ser combatida com eficácia por iniciativas voltadas à melhoria da gestão, como o Jovem de Futuro. Por outro lado, revela que o Jovem de Futuro deve ser pensado apenas como um primeiro passo. Muito mais ainda pode e precisa ser feito para a melhoria no uso eficaz do gasto público com educação no Brasil via melhorias na gestão.

REFERÊNCIAS

BARROS, R. P. *et al.* O impacto do projeto jovem de futuro sobre a aprendizagem em escolas públicas. *In*: INSTITUTO UNIBANCO (Org.). **Caminhos para a qualidade da educação pública: impactos e evidências**. São Paulo: Instituto Unibanco, 2017. p. 38-57. Disponível em: <<https://bit.ly/3viESLQ>>. Acesso em: 14 jun. 2019.

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

_____. Does management matter in schools? **The Economic Journal**, v. 125, n. 584, p. 647-674, May 2015.

DUNCAN, G. J.; MAGNUSON, K. Investing in preschool programs. **Journal of Economic Perspectives**, v. 27, n. 2, p. 109-132, 2013.

FRYER JUNIOR, R. G. Injecting charter school best practices into traditional public schools: evidence from field experiments. **The Quarterly Journal of Economics**, v. 129, n. 3, p. 1355-1407, Apr. 2014.

_____. **Management and student achievement: evidence from a randomized field experiment**. Cambridge, Massachusetts: NBER, May 2017. (NBER Working Paper, n. 23437).

HENRIQUES, R.; CARVALHO, M.; BARROS, R. P. (Org.). **Avaliação de impacto em educação: a experiência exitosa do programa Jovem de Futuro em parceria com o poder público**. São Paulo: Instituto Unibanco, 2020.

HILL, C. J. *et al.* Empirical benchmarks for interpreting effect sizes in research. **Child Development Perspectives**, v. 2, n. 3, p. 172-177, Dec. 2008.

INEP – INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Microdados da Aneb e da Anresc 2015**. Brasília: Inep, 2017.

_____. **Relatório do 2º ciclo de monitoramento das metas do Plano Nacional de Educação: 2018**. 2. ed. Brasília: Inep, 2019a. Disponível em: <<https://bit.ly/3OWs2Kv>>. Acesso: 16 jul. 2019.

_____. **Indicadores financeiros educacionais**. Brasília: Inep, 2019b. Disponível em: <<https://bit.ly/3LM2ei6>>. Acesso: 17 jul. 2019.

LÜCK, H. (Org.). **Dimensões da gestão escolar e suas competências**. Curitiba: Editora Positivo, 2009.

OECD – ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **PISA 2015 results: excellence and equity in education**. Paris: OECD Publishing, Dec. 2016. v. 1. Disponível em: <<https://bit.ly/3M1vKAT>>. Acesso em: 3 out. 2018.

_____. **Education at a glance 2017: OECD indicators**. Paris: OECD Publishing, Sept. 2017. Disponível em: <<https://bit.ly/351NhIU>>. Acesso em: 3 out. 2018.

TODOS PELA EDUCAÇÃO. **Anuário brasileiro da educação básica 2019**. São Paulo: Moderna, 2019. Disponível em: <<https://bit.ly/3vfYn7J>>. Acesso em: 24 jul. 2019.

VASCONCELLOS, L. **Monitoramento e avaliação: estudos selecionados para o Brasil**. [s.l.]: Itáu Unibanco, Set. 2015. Disponível em: <<https://bit.ly/3NPtooU>>. Acesso em: 9 maio 2022.

BIBLIOGRAFIA COMPLEMENTAR

BRASIL. Senado Federal. **Constituição da República Federativa do Brasil**. Brasília: Senado Federal, 1988. Disponível em: <<https://bit.ly/3t30NDU>>. Acesso em: 16 jul. 2019.

Originais submetidos em: nov. 2018.

Última versão recebida em: dez. 2020.

Aprovada em: dez. 2020.