

O IMPACTO DO PROJETO JOVEM DE FUTURO SOBRE A APRENDIZAGEM EM ESCOLAS PÚBLICAS

Ricardo Paes de Barros, *Economista Chefe do Instituto Ayrton Senna e Professor no Insper*
Rosane Mendonça, *Universidade Federal Fluminense*
Samuel Franco, *OPE Sociais e IETS*
Gabriela Gall Rosa, *Insper*

1. INTRODUÇÃO

O Jovem de Futuro é um Projeto do Instituto Unibanco que tem como objetivo melhorar a aprendizagem dos estudantes em escolas públicas estaduais de Ensino Médio por meio de uma gestão voltada para resultados. Iniciado em 2008, já foi implantado em dez estados brasileiros¹ e até 2015 tinha beneficiado quase 2,5 mil escolas estaduais e 1,6 milhão de estudantes de Ensino Médio.

Nesse estudo, apresentamos a avaliação de impacto do Jovem de Futuro, que possui desenho experimental e tem sido conduzida em todos os estados que o adotaram. O Projeto transforma práticas de gestão das escolas e, portanto, o impacto investigado é uma medida agregada da aprendizagem dos alunos que frequentavam as escolas durante o período em que a avaliação foi realizada. O texto não se limita a descrever os resultados encontrados e inclui uma apresen-

1 Os estados que têm ou já tiveram o Projeto são: Ceará, Espírito Santo, Goiás, Mato Grosso do Sul, Minas Gerais, Pará, Piauí, Rio de Janeiro, Rio Grande do Sul e São Paulo.

tação do método utilizado para isolar o impacto e também propõe uma reflexão que nos permite julgar o quão elevado é o efeito apresentado.

2. METODOLOGIA

2.1. Impacto sobre o quê?

Uma vez que a unidade de análise da avaliação de impacto do Projeto Jovem de Futuro é a escola, a pergunta a ser respondida é o quanto “ter participado do Projeto” garantiu às escolas um melhor desempenho acadêmico.

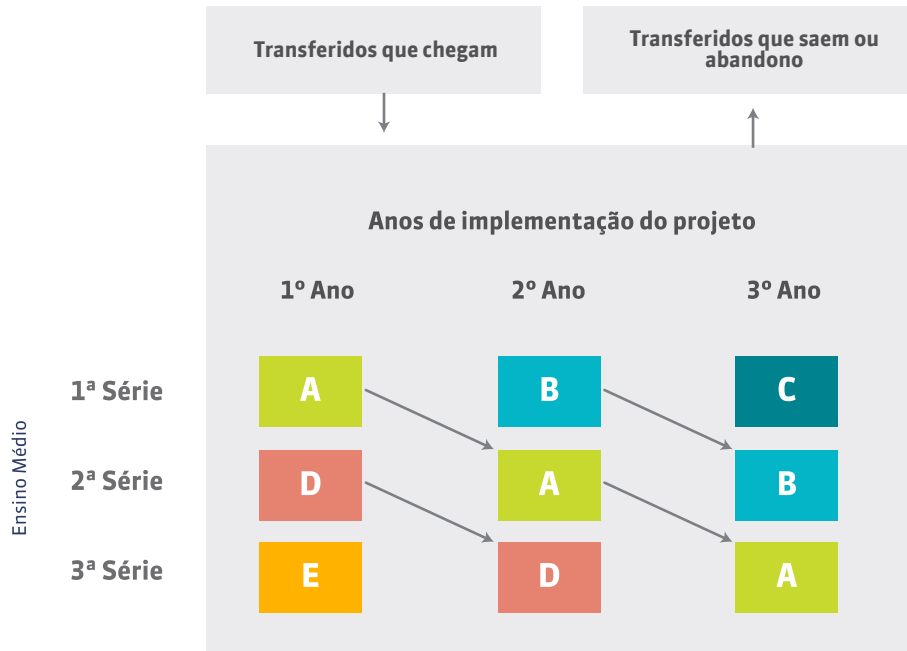
O desempenho acadêmico de uma escola, por sua vez, pode ser traduzido por diversos tipos de resultado, como proficiências em disciplinas específicas, taxas de aprovação, reprovação ou abandono, etc. Selecionado o resultado de interesse, é necessário ainda especificar que vamos expressá-lo como uma **característica** específica da distribuição dos alunos que frequentavam uma determinada **série** dessa escola **ao fim da janela de tempo da avaliação**.

Nossa opção foi avaliar o impacto do Jovem de Futuro sobre dois resultados tradutores do desempenho acadêmico: as proficiências em Língua Portuguesa e Matemática. Assim sendo, a formulação completa do que está sendo estudado é: o impacto sobre as **médias** (características da distribuição) das distribuições dos alunos da escola segundo as **proficiências em Língua Portuguesa e Matemática** (resultados tradutores do desempenho acadêmico dos alunos) **ao final da 3ª série** (série selecionada), **três anos após o início da implantação do Projeto** (janela de tempo da avaliação).

O fato da avaliação ser realizada com os alunos da 3ª série do Ensino Médio após três anos da data de início do Projeto é de particular importância. Afinal, esse é o primeiro momento em que se pode capturar o efeito de “cursar todas as séries dessa etapa de ensino” em uma escola com as práticas de gestão instituídas pelo Jovem de Futuro. Este seria o impacto pleno. Avaliações em qualquer outra série ou em uma janela de tempo de avaliação inferior a três anos captariam o impacto de uma exposição parcial dos estudantes ao Projeto, pois não consideraríamos aqueles estudantes que completaram as três séries do Ensino Médio na escola modificada pelo Projeto.

O Diagrama 1 nos ajuda a ilustrar essas ideias. Note que após três anos de implantação do Projeto, estarão cursando a 3ª série três tipos de estudantes. Primeiro, aqueles que no início do Projeto estavam na 1ª série e foram sempre progredindo (caixinha A – três anos em uma escola com gestão modificada, sendo que em cada ano de implantação foi cursada uma série diferente). Quando esse grupo chega ao final da 3ª série, estamos medindo o impacto pleno. Segundo, os estudantes que no início do Projeto estavam na 2ª série e que alternaram reprovação e aprovação (caixinha B – três anos em uma escola com gestão modificada, porém cursaram apenas duas séries distintas no período). Terceiro, os que no início do Projeto estavam na 3ª série e a repetiram duas vezes consecutivas (caixinha C – três anos em uma escola com gestão modificada, porém cursaram a mesma série em todo o período). Os estudantes representados pelas caixinhas D e E sequer chegaram à 3ª série ao final dos três anos de intervenção.

DIAGRAMA 1:
Fluxo na escola durante os 3 anos de implementação do Projeto

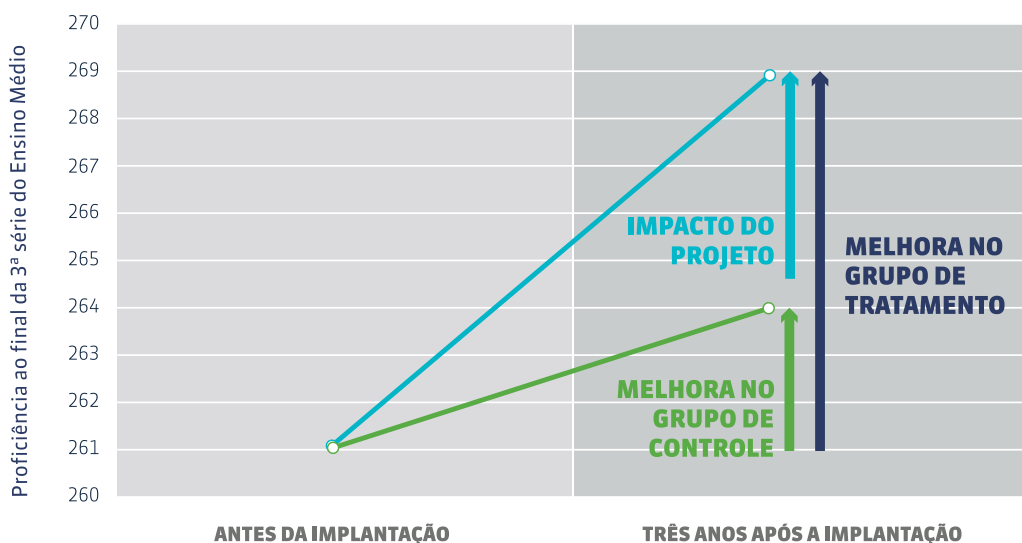


2.2. Construindo o grupo de controle

Toda avaliação de impacto com desenho experimental baseia-se no contraste entre a melhoria no desempenho de um grupo de escolas beneficiadas pelo Projeto (grupo de tratamento), antes e depois da intervenção (janela de tempo da avaliação), e a correspondente melhoria em um grupo de escolas não beneficiadas (grupo de controle) – veja Gráfico 1. Assim, é fundamental em qualquer avaliação de impacto a especificação de como os grupos de tratamento e controle foram obtidos. Esse é o objetivo desta seção.

Em termos conceituais, o impacto de um Projeto sobre um conjunto de escolas é, por definição, o contraste entre o que seria o desempenho dessas escolas na presença do Projeto e o que teria acontecido caso não tivessem sido beneficiadas. Como, ao final, cada escola vivenciará apenas uma das duas situações, a outra será sempre uma conjectura, um contrafactual, impossível de ser observado.

GRÁFICO 1:
Representação gráfica do conceito de impacto com desenho experimental



Tipicamente o que se deseja estimar é o impacto de uma intervenção sobre um grupo de escolas beneficiadas e, portanto, o que **não** se observa é o que teria acontecido com elas caso não fossem beneficiadas. Exatamente com o objetivo

de preencher essa lacuna informacional é que necessitamos de um grupo de controle, o qual evidentemente deve ser formado por escolas não beneficiadas. Mas essa condição de não estar entre as beneficiadas não é suficiente. Esse grupo precisa também ser selecionado de forma muito especial, uma vez que a evolução do desempenho de um verdadeiro grupo de controle precisa ser idêntica à evolução que o grupo de escolas beneficiadas teria caso não tivessem sido atendidas.

Embora ainda exista debate sobre como grupos de controle podem ser obtidos, uma questão já se encontra bem difundida há pelo menos um século: quando a seleção dos beneficiários é feita com base em um sorteio, os não beneficiados formam um grupo de controle ideal.

No caso da avaliação de impacto do Projeto Jovem de Futuro, os grupos de tratamento e controle foram obtidos a partir de um processo que pode ser descrito em três etapas:

1ª ETAPA: em cada um dos estados participantes do Projeto, um grupo ou a totalidade das escolas foi convidada a participar.

2ª ETAPA: as escolas que declararam interesse em participar foram agrupadas por semelhança. Esses grupos deveriam ser homogêneos com relação à evolução do desempenho que as escolas do grupo teriam na ausência do Projeto. O ideal seria que os grupos fossem formados apenas por duas escolas (pares de escola, portanto). A maior parte dos grupos acabou atendendo a esse critério, embora em alguns casos, o número de escolas do grupo chegue a 12 (veja Tabela 1).

3ª ETAPA: Organizou-se um sorteio público que, para cada grupo, selecionou aleatoriamente quais escolas seriam atendidas de imediato e quais teriam atendimento postergado. Em cada grupo, pelo menos uma escola teve o atendimento postergado para o fim da janela de tempo da avaliação.

Esse processo vem sendo realizado desde o início do Projeto em 2008. O Diagrama 2 apresenta o momento em que as escolas envolvidas na avaliação foram inicialmente sorteadas em cada um dos estados que são ou já foram parceiros do Instituto Unibanco. Ao longo desses sete anos de avaliação, muitas escolas foram sorteadas a partir dos grupos formados e, com isso, foram criadas opor-

tunidades para a avaliação de impacto do Projeto. Temos hoje 409 grupos de escolas e destes, 141 foram utilizados na avaliação de impacto do Projeto.

TABELA 1:
Número e tamanho dos grupos de escolas utilizados nos sorteios para seleção de tratamento e controles

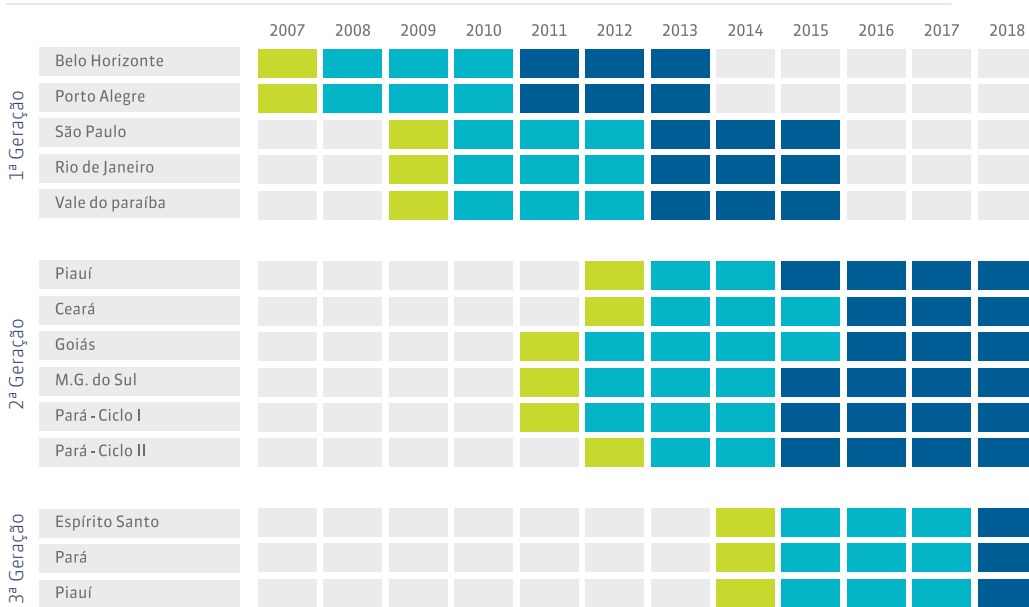
GERAÇÃO	ÁREA	NÚMERO DE GRUPOS		TAMANHO DOS GRUPOS (NÚMERO DE ESCOLAS)	
		TOTAL	UTILIZADOS NA AVALIAÇÃO	GRUPO DE TRATAMENTO	GRUPO DE CONTROLE
1a	Minas Gerais (RM* de Belo Horizonte)	4	4	5	7
	Rio Grande do Sul (RM* de Porto Alegre)	25	0	1	1
	São Paulo (RM* de São Paulo)	20	17	1	1
	São Paulo (Vale do Paraíba)	20	20	1	1
	Rio de Janeiro (RM* do Rio de Janeiro)	15	12	1	1
2a	Pará	44	15	1	1
	Piauí	25	0	7	1
	Ceará	26	25	5	1
	Goiás	25	25	5	1
	Mato Grosso do Sul	25	23	4	1
3a	Pará	42	0	1	1
	Piauí	134	0	1	1
	Espírito Santo	70	0	2	1
TOTAL		475	141	-	-

*Região Metropolitana

Fonte: Elaboração própria a partir de registros administrativos do Instituto Unibanco

DIAGRAMA 2:

Janela de tempo de avaliação



Legenda:

- Amarelo Linha de base
- Ciano Janela de avaliação
- Azul Pós Janela de Avaliação

2.3. Impacto medido de que forma?

Para medir o desempenho acadêmico das escolas, entendido nessa avaliação de impacto como proficiências em Língua Portuguesa e Matemática, optamos por utilizar somente medidas derivadas de fontes secundárias, independentes do Projeto Jovem de Futuro. Em quase todos os estados onde o Projeto foi implantado estavam em funcionamento “sistemas locais de avaliação da aprendizagem dos alunos” que abrangem **todas** as escolas da rede². O fato da proficiência ser aferida para todas as escolas e não apenas para aquelas envolvidas na avaliação de impacto do Projeto Jovem de Futuro é muito importante para

2 O único estado que não contava com avaliação externa para a 3ª série do Ensino Médio durante o período de avaliação de impacto do Jovem de Futuro foi o Rio Grande do Sul e, por essa razão, não foi possível incluir suas escolas no presente estudo.

garantir a validade interna da avaliação, isto é, para assegurar que o contraste entre as escolas de tratamento e controle, em termos de proficiência, seja um estimador não enviesado do impacto do Projeto.

Quando são utilizadas medidas de proficiência especialmente desenvolvidas para a avaliação de impacto do Projeto, mesmo sem a intenção, estaremos introduzindo algum tipo de viés na estimativa. Seja porque as escolas de controle e tratamento podem ter incentivos diferenciados para ir bem ou mal nos testes, seja porque os dois grupos podem ter os mesmos incentivos para se comportar de maneira atípica ao saberem do experimento, por exemplo, colocando excesso de empenho ou excesso de desmotivação na participação nos testes. Se os vieses fossem diferentes para tratamento e controle, estaríamos anulando a vantagem do sorteio, pois o que se observaria no grupo de controle não corresponderia ao que ocorreria com as escolas beneficiadas na ausência do Projeto. Consequentemente, o impacto estimado captaria tanto o efetivo benefício do Projeto como também diferenças devido à imperfeita comparabilidade das medidas de proficiência entre tratamento e controle.

Não queremos com isso dizer que os “sistemas locais de avaliação da aprendizagem dos alunos” meçam de forma não enviesada a real proficiência dos estudantes. Existem diversas razões não relacionadas com o Projeto Jovem de Futuro para que essas medidas sobre-estimem (em particular, quando os alunos são treinados para fazer a prova) e para que subestimem (quando existe desinteresse dos alunos pelo resultado de um exame que não fará parte de seu histórico escolar) a real proficiência dos estudantes. No entanto, como essa mensuração é desvinculada do Projeto, não existe razão para que esses vieses sejam diferentes nas escolas de tratamento e controle.

3. ESTIMANDO O IMPACTO

3.1. O Projeto Jovem de Futuro tem impacto?

Estatisticamente falando, nosso problema consiste em verificar se existe suficiente evidência para rejeitar a hipótese de que o Projeto não tem qualquer impacto sobre a proficiência dos alunos em escolas avaliadas.

Caso o Projeto não gere impacto, então a melhoria no desempenho das escolas beneficiadas, considerando o período antes e depois da intervenção, não deve ser estatisticamente diferente da correspondente melhoria observada nas escolas de controle. Ou seja, se o Projeto não tem impacto, a chance das escolas beneficiadas melhorarem mais do que as escolas de controle deve ser 50%. Isso pode ser entendido da seguinte forma: se “melhorar mais a proficiência” é algo independente do Jovem de Futuro, então é natural que, de forma aleatória, algumas escolas beneficiadas melhorem mais que as de controle e outras de controle melhorem mais do que as beneficiadas. Se esse é um efeito aleatório, uma escola beneficiada pelo Jovem de Futuro tem 50% de chance de ir melhor do que uma escola de controle. Por isso, independentemente do número de escolas em cada grupo e da divisão interna entre beneficiadas e controle, caso o Projeto não tenha impacto, os dois grupos têm as mesmas chances de melhoria.

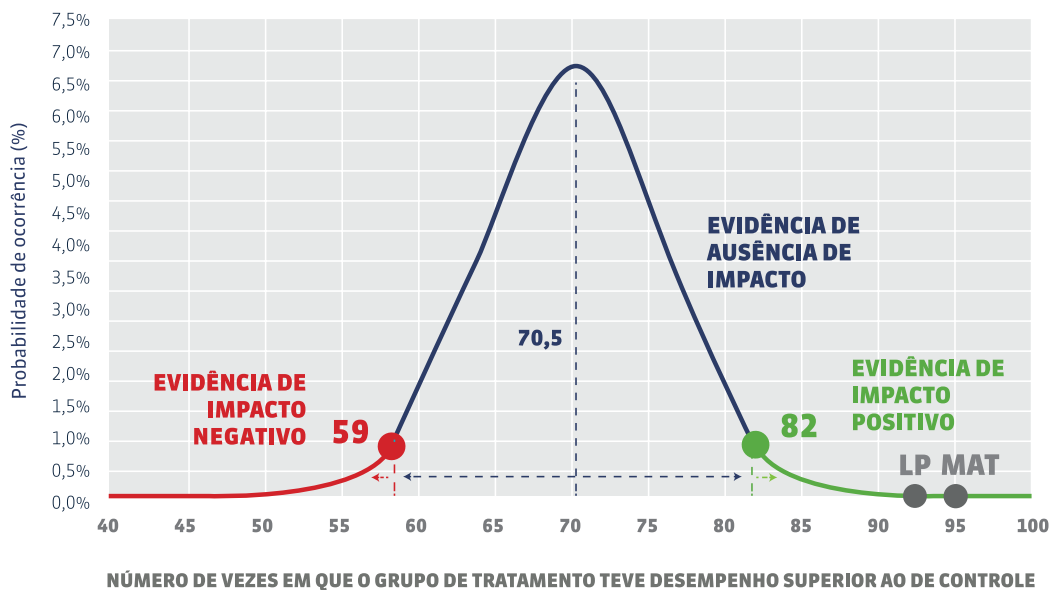
Dos 475 grupos de escolas formados nos estados e participantes do sorteio, foram considerados para a avaliação de impacto apenas 141. Isso porque foi necessário excluir grupos em quatro estados, seja porque não dispunham de um sistema externo de avaliação da aprendizagem dos alunos na 3ª série do Ensino Médio, seja porque o Jovem de Futuro ainda não completou os três anos de implementação.

Consideremos, por exemplo, a proficiência média em Língua Portuguesa ao final da 3ª série. Em cada um dos 141 grupos de escolas considerados na avaliação de impacto, medimos se a melhoria nas beneficiadas (antes e depois do Projeto) foi maior do que nas de controle. Se o Projeto não tem qualquer impacto, é esperado que em metade dos grupos, escolas beneficiadas tenham se saído melhor do que as de controle. Portanto, o esperado é encontrar 70 grupos em que há impacto (beneficiadas se saindo melhor).

Mas é evidente que, da mesma forma que alguém que lança 141 moedas não deveria esperar obter exatamente 70 caras ou coroas, não deveríamos esperar que, na ausência de impacto do Projeto, obtivéssemos exatamente 70 casos em que as escolas beneficiadas melhoraram mais do que as de controle. Na verdade, a probabilidade de se obter exatamente 70 caras e 71 coroas em 141 lançamentos é de apenas 6,7%. Assim sendo, utilizando um intervalo de confiança para tornar o argumento mais preciso, se o Projeto não tem impacto,

existiria 95% de chance de se observar entre 59 e 82 grupos em que as escolas beneficiadas se saem melhor do que as de controle. Para rejeitar a hipótese nula (hipótese de que o Projeto não tem impacto), o número de grupos em que se observa que beneficiadas se saem melhor do que controles precisa ser superior a 82 (veja Gráfico 2).

GRÁFICO 2:
Distribuição dos grupos da avaliação de impacto em que tratamento melhorou acima do controle segundo a probabilidade de ocorrência



Observamos que, três anos depois do início do Jovem de Futuro, em 92 dos 141 grupos de escolas, a melhoria na proficiência em Língua Portuguesa dos alunos na 3ª série das escolas beneficiadas foi superior à das escolas de controle. Existe, portanto, sólida evidência para rejeitarmos a hipótese de que o Projeto não tem qualquer impacto sobre o desempenho acadêmico das escolas.

É mínima a chance de olharmos para esses 92 grupos em que as beneficiadas se saíram melhor e concluirmos equivocadamente que o Projeto tem impacto. A chance desse erro seria uma em 10.000, portanto, extremamente improvável.

No caso da Matemática, vimos um resultado ainda melhor. Em 95 dos 141 grupos, as escolas beneficiadas se saíram melhor do que as de controle. A chance de estarmos cometendo um equívoco ao concluirmos que o Jovem de Futuro tem impacto na proficiência de matemática é uma em 100.000.

Para completar a consistência dessa argumentação, vamos olhar para a situação das escolas antes da implantação do Projeto. Conhecemos a proficiência de todas elas no ano imediatamente anterior ao da data de início do Jovem de Futuro. Como o Projeto não pode ter impacto sobre o passado e, na medida em que o sorteio foi bem conduzido, devemos esperar que em 70 grupos, as escolas beneficiadas tenham um ponto de partida melhor do que as de controle. Como visto anteriormente, estatisticamente falando, deveríamos considerar um intervalo de confiança em torno dos 70 casos, que vai de 59 a 82. Os resultados encontrados confirmam integralmente a expectativa de que não há diferenças iniciais relevantes entre tratamento e controle. Com relação à proficiência média em Língua Portuguesa, há 65 dos 141 grupos em que as beneficiadas tinham situação de partida melhor. No caso da proficiência média em Matemática, encontramos que em 68 dos 141 grupos, o ponto de partida das escolas beneficiadas era melhor.

Podemos concluir com elevadíssimo grau de certeza que o Projeto Jovem de Futuro possui impacto na proficiência média de Língua Portuguesa e Matemática dos alunos de 3ª série do Ensino Médio.

3.2. Qual a magnitude do impacto do Projeto Jovem de Futuro?

Saber que o Projeto gerou impacto é bom, mas talvez não o suficiente, pois a expectativa dos gestores é que esse impacto seja de magnitude relevante. Para saber a magnitude do impacto do Jovem de Futuro, o procedimento utilizado é muito similar ao que vimos na seção anterior. Até então, rejeitamos a hipótese de que o Projeto não tem impacto, considerando o número de grupos em que se observou que as escolas beneficiadas melhoraram mais do que as de controle. Nesta seção, qualificamos a análise, considerando a informação sobre quantos pontos a mais na Escala Saeb³, as escolas beneficiadas melhoram

3 A Escala Saeb é a escala brasileira na qual expressamos a proficiência em Língua Portuguesa e Matemática, construída pelo Ministério da Educação.

comparadas com as de controle. A precisão dessas estimativas pode ser obtida a partir da variabilidade dos diferenciais na melhoria em proficiência entre grupos, conforme indicado nas Tabelas 2A e 2B.

TABELA 2A:
Impacto do Jovem de Futuro: Proficiência em Língua Portuguesa

GERAÇÃO	ÁREA	IMPACTO SOBRE O APRENDIZADO NA 3ª SÉRIE, 3 ANOS APÓS O INÍCIO DO PROJETO				
		Média (pontos na Escala Saeb)	Número de agrupamentos	Erro-padrão (pontos na Escala Saeb)	Limite inferior (95% de confiança)	Estatística para detectar heterogeneidade
	Média simples dos experimentos	4	141	1,4	2,1	-
	Média ponderada das áreas	5	-	1,7	1,7	
1a	Minas Gerais (RM* de Belo Horizonte)	10	4	4,7	2,0	1,1
	São Paulo (RM* de São Paulo)	5	17	4,2	-1,5	0,2
	São Paulo (Vale do Paraíba)	3	20	2,9	-1,3	-0,4
	Rio de Janeiro (RM* do Rio de Janeiro)	6	12	6,8	-5,6	0,1
2a	Pará	2	15	4,6	-5,6	-0,6
	Ceará	5	25	2,7	0,5	0,1
	Goiás	5	25	4,0	-1,4	0,1
	Mato Grosso do Sul	3	23	3,4	-2,5	-0,4

*Região Metropolitana

Fonte: Elaboração própria a partir de informações das avaliações externas estaduais

TABELA 2B:**Impacto do Jovem de Futuro: Proficiência em Matemática**

GERAÇÃO	ÁREA	IMPACTO SOBRE O APRENDIZADO NA 3ª SÉRIE, 3 ANOS APÓS O INÍCIO DO PROJETO				
		Média (pontos na Escala SAEB)	Número de agrupamentos	Erro-padrão (pontos na Escala SAEB)	Limite inferior (95% de confiança)	Estatísticas para detectar heterogeneidade
	Média simples dos experimentos	6	141	1,2	3,8	-
	Média ponderada das áreas	5	-	1,0	3,0	7,9
1a	Minas Gerais (RM* de Belo Horizonte)	12	4	6,0	1,9	1,2
	São Paulo (RM* de São Paulo)	1	17	2,1	-3,0	-1,9
	São Paulo (Vale do Paraíba)	7	20	2,9	2,8	1,0
	Rio de Janeiro (RM* do Rio de Janeiro)	7	12	5,0	-1,4	0,4
2a	Pará	10	15	4,7	2,4	1,2
	Ceará	5	25	2,1	1,8	0,3
	Goiás	6	25	3,8	-0,5	0,3
	Mato Grosso do Sul	4	23	2,1	0,6	-0,3

*Região Metropolitana

Fonte: Elaboração própria a partir de informações das avaliações externas estaduais

Uma primeira estimativa simples da magnitude do impacto do Projeto sobre a proficiência média dos alunos ao final da 3ª série do Ensino Médio é de 4 pontos na Escala Saeb em Língua Portuguesa e de 6 pontos na Escala Saeb em Matemática.

No entanto, como o número de escolas em cada grupo é distinto e a variabilidade do desempenho entre escolas pode mudar entre os estados, a forma estatís-

ticamente mais eficiente de agregar as estimativas de impacto dos 141 grupos não é necessariamente igual à média aritmética simples adotada acima. Em princípio, cada grupo oferece um estimador da magnitude do impacto do Projeto com precisão diferenciada, dependendo do número de escolas presentes no grupo, do balanceamento no número de tratamentos e de controles e também da variabilidade de desempenho entre escolas no estado. Idealmente, a precisão do estimador de impacto será maximizada se as informações dos grupos forem ponderadas por sua precisão, entendida como o inverso da variância.

Quando esse segundo procedimento é utilizado, o impacto sobre a proficiência em Língua Portuguesa aumenta, passando de 4 para 5 pontos na Escala Saeb. No caso da matemática ocorre o contrário. A estimativa de impacto cai, passando de 6 para 5 pontos na Escala Saeb.

Em resumo, nossas estimativas pontuais apontam para um impacto do Projeto de magnitude entre 4 e 6 pontos na Escala Saeb, tendendo a ser um pouco maior em Matemática do que em Língua Portuguesa.

3.3. Heterogeneidade do impacto

Nas Tabelas 2A e 2B apresentamos tanto estimativas agregadas do impacto do Projeto Jovem de Futuro para o conjunto dos estados atendidos, como também estimativas individuais para cada estado. Existe considerável variabilidade entre os impactos aferidos em cada estado. Em Língua Portuguesa, por exemplo, o impacto nos estados vai de 2 a 10 pontos na Escala Saeb. Em Matemática variam em um intervalo ainda maior: entre 1 e 12 pontos na Escala Saeb.

Existem duas possíveis explicações para essa variabilidade: (a) o Projeto realmente tem impacto diferenciado a depender do estado em que é implantado e, nesse caso, dizemos que o impacto do Projeto é heterogêneo; ou (b) trata-se apenas de flutuação estatística resultante do erro de medição. Como conhecemos a precisão das estimativas de impacto para cada estado, é perfeitamente possível avaliar a legitimidade ou a importância dessas duas possibilidades.

Uma estatística comumente utilizada para avaliar se as diferenças de impacto entre os estados é real (o que tornaria o impacto heterogêneo) consiste em

expressar a diferença entre a magnitude do impacto em um estado e a média para o conjunto deles em múltiplos do erro-padrão do estimador de impacto para o estado. Por exemplo, como vimos no caso de Língua Portuguesa, o impacto estimado para o conjunto dos estados foi de 5 pontos na Escala Saeb. Como revela a Tabela 2A, em Minas Gerais a magnitude estimada para o impacto do Projeto sobre a proficiência em Língua Portuguesa foi de 10 pontos na Escala Saeb, com um erro-padrão associado de 4,7. Assim, a diferença em relação ao impacto agregado expressa em múltiplos desse erro-padrão seria de 1,1 ($\frac{9,7-4,6}{4,7} = 1,1$). O valor dessa razão em módulo deve ser próximo à unidade ou menor para afirmarmos que tratam-se puramente de erros de medida. Já valores bem acima da unidade apontam para alguma dose de heterogeneidade do impacto. A evidência disponível é perfeitamente compatível com a argumentação de que a variabilidade observada é fruto exclusivamente de erros aleatórios, naturalmente presentes na mensuração do impacto em cada estado. Se existe alguma evidência de heterogeneidade, ela se restringe a São Paulo – Região Metropolitana. Nesse caso, a estatística detectora da heterogeneidade é -1,9, que tomada em módulo, nos dá um valor bem superior à unidade.

4. INTERPRETANDO O IMPACTO ESTIMADO

4.1. Relevância da magnitude do impacto estimado

Já sabemos que é possível afirmar, de forma generalizada, que o impacto do Projeto Jovem de Futuro é de 5 pontos na Escala Saeb, tanto sobre a proficiência em Língua Portuguesa quanto em Matemática. Resta, agora, verificar qual a relevância dessa magnitude de impacto.

O impacto de uma intervenção é comumente expresso como uma porcentagem do desvio-padrão do resultado sobre o qual se está medindo o impacto. Com vistas a facilitar a comparabilidade, tomemos como parâmetro o desvio-padrão em proficiência tanto em Matemática como em Língua Portuguesa no Brasil, que é da ordem de 33 pontos na Escala Saeb. Assim, um impacto de 5 pontos na Escala Saeb corresponde a 15% de um desvio-padrão.

Considerando os achados de uma importante meta-análise internacional (Hill, Bloom, Black, & Lipsey (2007)⁴), que reúne avaliações de intervenções educacionais voltadas à melhoria do aprendizado de jovens na Educação Média, o impacto médio encontrado em 76 avaliações foi de 24% de um desvio padrão. Se restringirmos o universo aos experimentos, temos 43 avaliações que indicam um impacto médio de 27% de um desvio-padrão. Um pouco mais de um terço das avaliações experimentais incluídas nesta meta-análise apresentaram impactos inferiores ao do Jovem de Futuro.

Ainda recorrendo às meta-análises realizadas no Brasil e no mundo sobre avaliação de impacto⁵, sabemos que 5 pontos na Escala Saeb (ou 15% de um desvio padrão) é um efeito maior do que o observado em intervenções que fazem os estudantes passarem a ter aulas com um professor mais experiente. O impacto de ter aula com professores mais experientes é 10% de um desvio-padrão. Também é maior do que o efeito produzido pela redução do tamanho da sala de aula de 22 para 15 alunos, que também está na ordem dos 10% de um desvio-padrão. Por outro lado, considerando especificamente a qualidade do professor, a literatura indica que o impacto de se estudar com um dos melhores professores da escola (aqueles que mais conseguem bons resultados de aprendizagem com seus estudantes) ao invés de estudar com um dos mais fracos é da ordem de 30% de um desvio-padrão, portanto o dobro do impacto do Jovem de Futuro.

Uma terceira alternativa para se avaliar a relevância de 5 pontos a mais na Escala Saeb é comparar esse efeito com o quanto um aluno aprende ao longo de todo o Ensino Médio. A diferença entre a proficiência média em Matemática e Língua Portuguesa dos alunos ao final do Ensino Médio e a correspondente proficiência ao final da Educação Fundamental foi de 18 pontos na Escala

4 Hill, Carolyn J.; Bloom, Howard S.; Black, Alison Rebeck; Lipsey, Mark W. (2007). Empirical Benchmarks for Interpreting Effect Sizes in Research. MDRC Working Papers on Research Methodology.

5 Consultar o site Caminhos para melhorar o aprendizado (www.paramelhoraroaprendizado.org.br)

Saeb em 2013. Se considerarmos essa diferença como uma boa medida do que tipicamente um aluno aprende ao longo do Ensino Médio e se, a cada série, os alunos adquirem mais ou menos a mesma quantidade nova de conhecimento, o aprendizado por série seria de 6 pontos na Escala Saeb. Assim, o impacto do Jovem de Futuro seria mais ou menos 80% do que se aprende em uma série do Ensino Médio.

Por fim, cabe analisar o quanto os estados brasileiros foram capazes de avançar ao longo da última década em proficiência. A evidência disponível revela que, entre 2003 e 2013, apenas quatro unidades da federação conseguiram melhorar a proficiência no Ensino Médio público em 5 pontos na Escala Saeb ou mais. Assim, o impacto do Projeto é maior do que o progresso alcançado por 85% das Unidades da Federação ao longo de uma década.

4.2. Impacto sobre a escola versus impacto sobre o aluno

Caso não houvesse reprovação, abandono ou mesmo transferência de estudantes entre escolas, **todos** os concluintes do Ensino Médio teriam cursado as três séries na mesma escola. Esse é o caso que poderia nos dar o maior impacto pleno possível e nos permitiria afirmar que o impacto do Projeto sobre o desempenho da escola é igual ao impacto sobre a proficiência dos estudantes que ingressaram na 1ª série do Ensino Médio no primeiro ano de implantação e cursaram todas as três séries na escola beneficiada.

No entanto, essa não é a realidade e há muitos entraves no fluxo. Muitos dos estudantes que participaram da avaliação de impacto do Jovem de Futuro ao final da 3ª série do Ensino Médio foram, na realidade, repetentes na 2ª série ou ainda são repetentes (por duas vezes) da 3ª série. Ao menos, nesse caso, todos esses estudantes permaneceram por três anos na escola beneficiada, o que nos permitiria afirmar que o impacto sobre a escola é igual ao impacto sobre os alunos que passaram três anos em uma escola beneficiada, independentemente da série em que se encontravam no primeiro ano de implantação da intervenção.

Contudo, sabemos que há alunos que vieram transferidos de outras escolas e que, portanto, não ficaram expostos por três anos ao projeto Jovem de Futuro e isso faz com que o impacto sobre a escola, que viveu três anos de Projeto, não é o mesmo que o impacto sobre os alunos que cursavam a 3ª série do Ensino Médio no momento final da avaliação.

Ocorre também que o Projeto pode mudar o próprio fluxo escolar. Suponhamos que ele aumente a taxa de aprovação e reduza a de abandono. Ao fazer isso, os estudantes que acabam frequentando a 3ª série do Ensino Médio não seriam os mesmos que a frequentariam na ausência do Projeto. Possivelmente seriam estudantes com proficiência média menor do aqueles que chegariam à 3ª série se a escola não estivesse no Projeto. Afinal, na ausência do Projeto, as taxas de reprovação e abandono seriam mais altas, permitindo chegar à 3ª série apenas os “sobreviventes”. Mais uma vez, portanto, o impacto sobre o desempenho da escola ao final da 3ª série não é igual ao impacto sobre o desempenho dos alunos que frequentavam essa escola ao final da 3ª série, pois o Projeto provocaria mudança na composição do grupo que chega à 3ª série. É possível ainda que bons alunos de outras escolas sejam atraídos para a escola beneficiada e essa experimente uma melhora de resultado sem ter feito, na realidade, qualquer progresso pedagógico com os alunos que lá estavam anteriormente.

Em suma, ao avaliarmos o impacto sobre o desempenho da escola, o que estamos capturando é o quanto uma escola beneficiada evolui a mais do que evoluiria caso o Projeto não existisse. E isso inclui tanto as mudanças na composição de perfil dos estudantes que frequentaram a escola beneficiada em qualquer um dos três anos de Projeto, como também o quanto realmente a aprendizagem teria sido influenciada pelo Projeto.

Para calcular o impacto do Jovem de Futuro sobre o real desempenho dos estudantes e não apenas sobre o desempenho da escola, seria necessário neutralizar o efeito da mudança de composição no perfil dos estudantes. Para isso, bastaria restringir a análise a um universo fixo de alunos, por exemplo, àqueles matriculados no momento em que o Projeto foi inicialmente implantado. Tal

estratégia foi seguida em versões anteriores da avaliação de impacto do Jovem de Futuro, revelando efeitos de magnitude similar ao impacto sobre a escola. Além disso, no passado também foram investigados os impactos sobre o fluxo escolar, mas não foram encontrados efeitos significativos, apontando para baixa modificação no perfil dos estudantes que chegam a 3ª série comparado ao que se observaria caso o Projeto não existisse. Há alguma evidência de que o impacto sobre a escola pode ser similar ao impacto sobre os estudantes.

5. PRINCIPAIS CONCLUSÕES E RECOMENDAÇÕES

Os resultados apontam inequivocamente para um impacto positivo do Projeto sobre a aprendizagem em Língua Portuguesa e Matemática nas escolas públicas. A magnitude do impacto em ambas as disciplinas é de 5 pontos na Escala Saeb ou 15% de um desvio-padrão. Embora o tamanho do impacto varie dependendo do estado em que o Projeto foi implantado, esse efeito é reflexo de flutuações estatísticas provenientes de erros de medida.

Cinco pontos a mais na aprendizagem em escolas beneficiadas pelo Jovem de Futuro não é pouco, principalmente se consideramos o quanto um aluno brasileiro tem aprendido durante toda a etapa do Ensino Médio. Cinco pontos na aprendizagem é 80% do que se aprende tipicamente em uma série do Ensino Médio e, portanto, o efeito do programa é parecido com o que aconteceria caso o Ensino Médio passasse a ter uma série a mais. Ou ainda, cinco pontos em três anos é mais do que 23 dos 27 estados brasileiros melhoraram em proficiência durante toda a última década.

Segundo a literatura internacional, impactos medidos por meio de experimentos podem chegar a 27% de um desvio padrão. É fato que as intervenções com maior capacidade de afetar a aprendizagem estão ligadas ao processo pedagógico e incidem sobre os professores. Mais de um terço das avaliações internacionais mostraram efeitos inferiores aos do Jovem de Futuro. Há espaço para ajustes no desenho e na implementação do Projeto visando a elevação do nível atual de impacto.

Caso pretenda-se tomar como meta alcançar níveis próximos aos 25% de um desvio-padrão, seria necessário aumentar o nível atual de impacto em cerca de 10 pontos percentuais, o que soa extremamente desafiador.

Perante tal desafio e dado o limite da potência das transformações em gestão para alterar a aprendizagem na escola, muito provavelmente será necessário incorporar no desenho atual componentes que atuem diretamente no processo de ensino-aprendizagem. Percebemos preocupação do Instituto Unibanco em discutir e caminhar nessa direção.