

# Avaliação Econômica do Programa Excelência em Gestão Educacional (projeto piloto)

Maria Carolina Nogueira Dias<sup>1</sup>  
Clarissa Gondim Teixeira<sup>2</sup>

1 Fundação Itaú Social  
2 Itaú Unibanco

## Resumo

---

O PROGRAMA "EXCELÊNCIA EM GESTÃO EDUCACIONAL" É UMA INICIATIVA DE FORMAÇÃO CONTINUADA DE GESTÃO DO TRABALHO PEDAGÓGICO E DE INSERÇÃO DO PROFISSIONAL "COORDENADOR DE PAIS" NAS ESCOLAS, CUJA FUNÇÃO É FAVORECER VÍNCULOS ENTRE FAMÍLIA E INSTITUIÇÃO ESCOLAR POR MEIO DE VISITAS DOMICILIARES. COMO ESTRATÉGIA DE GESTÃO, OPTOU-SE POR EMPREGAR DIFERENTES ABORDAGENS AVALIATIVAS COMPLEMENTARES: AVALIAÇÃO DE PROCESSO E MÉTODOS MISTOS. ESSE TEXTO TRATA DA PARTE QUANTITATIVA DA AVALIAÇÃO DE IMPACTO, QUE MEDE O EFEITO DO PROGRAMA SOBRE O DESEMPENHO DOS ALUNOS NAS NOTAS DE LÍNGUA PORTUGUESA E MATEMÁTICA, USANDO DADOS DO SARESP. FORAM OBTIDOS IMPACTOS POSITIVOS E SIGNIFICATIVOS EM ALGUMAS SÉRIES, PORTANTO NÃO GENERALIZADOS, PORÉM ROBUSTOS AOS MÉTODOS. A ANÁLISE CONJUNTA DAS AVALIAÇÕES PERMITE CONCLUIR QUAIS OBJETIVOS FORAM FACTÍVEIS DE SEREM CONCRETIZADOS E AS HIPÓTESES INTERPRETATIVAS CONTRIBUEM PARA O AMADURECIMENTO DO REAL ALCANCE DO PROGRAMA, FAVORECENDO AJUSTES OPORTUNOS.

## Abstract

---

*THE "EXCELLENCE IN EDUCATION MANAGEMENT PROGRAM" IS AN INITIATIVE OF CONTINUING TRAINING OF EDUCATIONAL MANAGEMENT IN SCHOOLS AND THE INTRODUCTION OF A "PARENTING COORDINATOR" WHOSE FUNCTION IS TO PROMOTE LINKAGES BETWEEN FAMILY AND SCHOOL VIA HOME VISITS. THERE WERE EMPLOYED DIFFERENT COMPLEMENTARY EVALUATION APPROACHES: MONITORING AND MIXED EVALUATION METHODS. THIS PAPER ADDRESSES THE QUANTITATIVE PART OF THE IMPACT ASSESSMENT THAT MEASURES THE EFFECT OF THE PROGRAM ON STUDENT PERFORMANCE IN PORTUGUESE AND MATHEMATICS USING SARESP DATA. WE OBTAINED POSITIVE AND SIGNIFICANT EFFECTS IN SOME GRADES, THEREFORE NOT A GENERALIZED RESULT, BUT ROBUST METHODS. THE JOINT ANALYSIS OF THE EVALUATIONS SUGGESTS WHICH GOALS WERE FEASIBLE TO BE IMPLEMENTED. THE INTERPRETATIVE HYPOTHESES CONTRIBUTED TO ENHANCE THE SCOPE OF THE PROGRAM FAVORING TIMELY ADJUSTMENTS.*

### **PALAVRAS-CHAVE:**

Avaliação de impacto; Desempenho; Gestão escolar; Relação família-escola.

## Introdução

O programa “Excelência em Gestão Educacional” é uma iniciativa da Fundação Itaú Social em parceria com a Secretaria de Educação do Estado de São Paulo. Entre os anos de 2009 e 2011, foi desenvolvido um projeto-piloto, com coordenação técnica do Instituto Fernand Braudel, em dez escolas com baixo IDESP (Índice de Desenvolvimento da Educação do Estado de São Paulo) em 200 , que oferecem Ensino Fundamental II e Ensino Médio pertencentes à Diretoria Regional de Ensino Leste 3, na cidade de São Paulo.

A iniciativa, inspirada na reforma de ensino executada na cidade de Nova York, teve dois objetivos: (i) fortalecer as práticas da coordenação escolar e do corpo docente na oferta de uma aprendizagem de qualidade, criando mecanismos de gestão e supervisão do trabalho pedagógico; (ii) intensificar a participação dos pais nos esforços de melhoria do aprendizado, desenvolvendo estratégias que auxiliassem professores e gestores na aproximação família-escola, a partir da criação da função de coordenador de pais.

Esses objetivos desdobraram-se em duas estratégias. A primeira delas, denominada Tutoria, caracterizou-se pela oferta de formação continuada customizada, de caráter prático e modelar, realizada no cotidiano da escola, especialmente na sala de aula, por um par avançado (professor ou professora) também atuante na rede pública de ensino, chamado aqui de tutor. O foco da tutoria foi trabalhar com professores coordenadores e professores de Língua Portuguesa e Matemática, com base em sua adesão voluntária.

A segunda estratégia implicou a inserção de um novo profissional na escola, o Coordenador de Pais. Lançando mão de visitas domiciliares para alunos com risco de evasão escolar, organizando atividades para pais e alunos, mobilizando voluntários, ordenando o espaço escolar e atendendo a pais em espaços formais e informais, a intenção era favorecer aproximações e vínculos entre família e escola. O trabalho foi realizado por um profissional por escola, dedicado exclusivamente às atividades descritas e com grande conhecimento da comunidade.

Como estratégia de gestão, optou-se por empregar diferentes abordagens avaliativas complementares. Em uma Avaliação de Processo, procurou-se monitorar as atividades cotidianas desenvolvidas pela equipe, com vistas a alimentar o processo de gestão do programa e a produção de relatórios de acompanhamento. Com a Avaliação de Resultados com Métodos Mistos, optou-se de metodologias quantitativas e qualitativas. Essas análises foram feitas ao final do terceiro ano no projeto piloto e serão repetidas durante e após a expansão do programa para outras Diretorias Regionais de Ensino.

Os dados qualitativos foram obtidos a partir da aplicação de questionários, realização de grupos focais e de entrevistas em profundidade com diretores, família, professores, tutores e alunos. Os indicadores de “Planejamento”, “Gestão” e “Avaliação”, que avaliam os resultados do desempenho do professor, mostram que houve avanços nesses três itens. Na dimensão desempenho do aluno, verificamos

desempenhos crescentes, positivos e significativos nos indicadores de “Relação entre alunos” e “Relação com conhecimento”. Por sua vez, na área de gestão pedagógica, observamos que há uma melhor compreensão dos profissionais sobre sua ação, porém não se verifica mudança em sua prática cotidiana.

Procurou-se identificar o impacto do programa no desempenho dos alunos nas notas de língua portuguesa e matemática para as séries possíveis de serem seguidas nas bases de dados do Saresp (Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo), usando as metodologias quantitativas de Avaliação de Impacto: pareamento e diferenças em diferenças. Trataremos com mais detalhes exclusivamente sobre a avaliação de impacto nas próximas sessões.

## Metodologia

Esta avaliação tem foco sobre os três anos do programa piloto (2009 a 2011) e analisa os impactos sobre a melhoria de desempenho em Língua Portuguesa e Matemática, medido pelas notas dessas disciplinas no exame do Saresp. O exame é censitário e aplicado a alunos da rede estadual paulista quem frequentam Ensino Fundamental (5º, 7º e 9º anos) e Ensino Médio (3º ano). Por ser censitário e ter exames comparáveis no tempo, o uso do Saresp permitiu acompanhar o desempenho dos mesmos alunos ao longo dos anos.

A avaliação foi feita acompanhando cinco grupos de alunos no tempo que estão representados no quadro que segue. Este mostra todas as séries que passaram pelo programa em seus 3 anos de duração, nas colunas 2009, 2010 e 2011. As linhas mostram a trajetória do aluno ao longo do tempo. Em todos os casos, os anos de 2007 e 2008 correspondem aos anos de linha de base (antes do início do piloto) e os anos de 2010 e 2011 aos períodos pós-tratamento (após o início). Sendo assim, estamos avaliando dois (linhas claras) e três anos (linhas escuras) de exposição ao programa.

No Saresp, apenas é possível verificar o primeiro e o último ano das linhas claras e escuras. Os demais não são verificáveis, dado que o Saresp existe apenas para o 5º, 7º e 9º anos do ensino fundamental e 3º ano do ensino médio. Portanto, foram avaliados os alunos:

- que cursavam 5º ano do Ensino Fundamental em 2008 e o 7º ano em 2010;
- que cursavam 5º ano do Ensino Fundamental em 2007 e o 9º ano em 2011;
- que cursavam 7º ano do Ensino Fundamental em 2008 e o 9º ano em 2010;
- que cursavam 9º ano do Ensino Fundamental em 2008 e o 3º ano do Ensino Médio em 2011 ;
- que cursavam 9º ano do Ensino Fundamental em 2007 e o 3º ano do Ensino Médio em 2010.

■ **TABELA 1: ESQUEMA DOS ANOS AVALIADOS**

Linha de Base		Anos de Programa		
2007	2008	2009	2010	2011
				6º Ano
			6º Ano	7º Ano
	5º Ano	6º Ano	7º Ano	8º Ano
5º Ano	6º Ano	7º Ano	8º Ano	9º Ano
6º Ano	7º Ano	8º Ano	9º Ano	1º EM
7º Ano	8º Ano	9º Ano	1º EM	2º EM
8º Ano	9º Ano	1º EM	2º EM	3º EM
9º Ano	1º EM	2º EM	3º EM	
1º EM	2º EM	3º EM		

Fonte: Elaboração própria.

Devido a essa estratégia de acompanhar os alunos no tempo, o número de escolas considerado na avaliação passou a depender da oferta de ensino durante esse período. Por exemplo, para o acompanhamento dos alunos entre o 5º e o 7º anos, tivemos que considerar apenas as escolas que oferecem os dois ciclos do Ensino Fundamental, isto é, seis das dez do grupo de tratamento.

Além disso, o número de alunos encontrados na linha de base e no ano durante o programa também é bastante reduzido, uma vez que os alunos não podem ter repetido o ano, evadido ou mudado de escola nesse período. Os números de instituições e de alunos para cada amostra são mostrados a seguir, assim como o percentual de alunos que se mantiveram na amostra em relação à linha de base. Nota-se que, nas escolas piloto, o atrito de alunos, ou seja, alunos que repetiram o ano, evadiram ou mudaram de escola nesse período, é mais elevado que a média da RMSP.

Ressalta-se que, apesar do número reduzido de alunos, os resultados encontrados estão alinhados aos resultados obtidos quando con-

sideramos a totalidade dos matriculados em cada ano. Porém o resultado para o total de alunos é mais sujeito a viés devido à ausência de grupo controle pelas características fixas não observadas, conforme explicado a seguir.

## 2.1 DEFINIÇÃO DO GRUPO CONTROLE

Usamos as bases do SARESP, Censo Escolar e IDESP para caracterizar as escolas e os alunos das escolas piloto e demais unidades escolares elegidas para o grupo de controle. Para essa escolha, partimos do grupo de alunos de escolas estaduais da região metropolitana de São Paulo (RMSP), dado que as do interior do estado têm características e contextos bastante distintos. Verificamos que as escolas da RMSP cujo IDESP situa-se entre os 5% piores do estado tinham em média um desempenho próximo às que participaram do piloto, pois fazer parte desse grupo foi critério de inclusão no projeto. Trata-se de um conjunto de escolas bastante parecido com as unidades que participaram do piloto em diversas características relevantes para a análise do perfil socioeconômico do aluno e do perfil da instituição escolar, conforme será mostrado.

■ **TABELA 2: NÚMERO DE ESCOLAS E ALUNOS AVALIADOS**

	Total de escolas no programa	Total de alunos das escolas piloto avaliados
5º - 7º (2008-2010)	6	375
5º - 9º (2007-2011)	6	296
7º - 9º (2008-2010)	9	523
9º - 3ºEM (2008-2011)	10	295
9º - 3ºEM (2007-2010)	10	279

Fonte: Elaboração própria, SARESP.

Uma segunda estratégia adotada foi usar como grupo de controle as demais escolas da Leste 3 com melhor desempenho no IDESP. São escolas melhores na média do que as que participaram, mas que fazem parte do mesmo contexto e, por isso, têm a vantagem de serem influenciadas do mesmo modo durante esse período. De fato, a análise dos gráficos das notas médias desses grupos mostra que as trajetórias são paralelas, apesar de as notas médias das demais escolas da Leste 3 estarem em um patamar superior em relação às escolas piloto e as de pior IDESP das outras diretorias de ensino.

Ainda no que tange ao desempenho, a próxima tabela traz a nota média no exame do Saresp para as escolas tratadas, para as da RMSP de pior IDESP, para as demais escolas da Leste 3 e todas as da RMSP na linha de base e após o início do piloto nos anos avaliados. A tabela mostra que, em geral, os alunos dos colégios que participaram do programa e os estudantes das escolas da RMSP de pior IDESP partem de um desempenho médio similar entre eles e inferior aos alunos das outras instituições escolares da Leste 3 e da RMSP na linha de base. Isso justifica a adoção das escolas da RMSP de pior IDESP como grupo de controle.

No entanto, os estudantes do 9º ano do EF (2011) e do 3º ano do EM (2010) (apenas Língua Portuguesa) têm desempenho na linha de base mais próximo às demais escolas da Leste 3, corroborando a importância de trabalhar também com esse grupo como controle. Ressalta-se que essas diferenças iniciais no desempenho serão corrigidas ao parear alunos usando a nota inicial como critério.

Sobre o crescimento no desempenho, captado pela coluna "diferença", no 3º ano do EM (2010), 9º ano do EF (2010) e 7º ano do EF (2010) há variações discretas nas trajetórias de desempenho entre os grupos de tratamento e controle. Conta-se apenas uma diferença mais acentuada na trajetória de crescimento dos alunos que cursaram o 3º ano do EM (2011), tanto em Língua Portuguesa quanto em Matemática, no qual a média das escolas tratadas apresenta uma diferença de cerca de 7 pontos acima da média das escolas controle.

Não podemos atribuir as diferenças mencionadas exclusivamente ao efeito do projeto Excelência em Gestão Educacional, pois as trajetórias do desempenho ao longo do tempo podem ser influenciadas pelo programa,

mas também por outros fatores relacionados com a composição dos grupos que fazem que os grupos não sejam comparáveis. Por essa razão, só podemos atribuir causalidade aos efeitos do programa após o pareamento dos

alunos de tratamento e controle segundo o conjunto de características citadas anteriormente. Depois desse processo, verificamos que apenas o resultado em Língua Portuguesa se mantém, como será mostrado a seguir.

■ **TABELA 3: DESEMPENHO MÉDIO DOS ALUNOS DA AMOSTRA NO SARESP POR TIPO DE ESCOLA E GERAÇÃO**

3º ano do EM (2011)						
	MATEMÁTICA			LÍNGUA PORTUGUESA		
	2008	2011	Diferença	2008	2011	Diferença
Tratamento	235,19	261,97	25,78	221,45	254,88	33,43
RMSP de pior IDESP	238,40	258,60	19,20	226,05	252,03	25,98
Leste 3	240,91	260,78	19,87	228,68	356,13	27,44
RMSP	247,39	265,52	17,24	235,98	263,19	26,33
3º ano do EM (2010)						
	MATEMÁTICA			LÍNGUA PORTUGUESA		
	2007	2010	Diferença	2007	2010	Diferença
Tratamento	218,51	257,32	38,81	239,39	254,20	14,81
RMSP de pior IDESP	220,02	255,05	35,03	230,59	250,34	19,75
Leste 3	227,44	258,78	31,34	244,93	258,51	13,59
RMSP	231,47	264,43	32,96	249,80	263,06	13,26
9º ano do EF (2011)						
	MATEMÁTICA			LÍNGUA PORTUGUESA		
	2007	2010	Diferença	2007	2010	Diferença
Tratamento	181,55	235,99	54,44	185,87	224,91	39,04
RMSP de pior IDESP	170,97	234,82	63,85	177,92	219,39	41,46
Leste 3	180,21	236,62	56,41	183,59	222,96	39,37
RMSP	180,24	240,76	59,51	186,23	226,54	39,57
9º ano do EF (2010)						
	MATEMÁTICA			LÍNGUA PORTUGUESA		
	2008	2010	Diferença	2008	2010	Diferença
Tratamento	199,43	232,25	32,82	195,84	221,31	25,47
RMSP de pior IDESP	199,52	233,74	34,21	196,26	217,98	21,72
Leste 3	203,63	236,06	32,43	200,70	224,18	23,48
RMSP	207,87	240,34	32,47	206,17	227,72	21,55
7º ano do EF (2010)						
	MATEMÁTICA			LÍNGUA PORTUGUESA		
	2008	2010	Diferença	2008	2010	Diferença
Tratamento	176,71	203,60	26,88	168,28	197,79	29,50
RMSP de pior IDESP	176,07	199,92	23,85	165,04	191,54	26,50
Leste 3	178,87	199,94	21,07	167,05	192,89	25,84
RMSP	176,81	201,01	24,20	166,36	192,10	25,74

Fonte: Elaboração própria, SARESP.

■ **TABELA 4: DESCRIÇÃO DAS VARIÁVEIS DE ESCOLAS DETERMINANTES DE DESEMPENHO ESCOLAR POR GRUPOS DE ESCOLAS – 9º ANO DO EF EM 2008**

Variáveis de escola determinantes do desempenho	10 Escolas Piloto	Leste 3		RMSP pior IDESP		RMSP	
IDESP 4ª série	2.22	2.48		2.03		2.58	
IDESP 8ª série	1.64	1.93	**	1.52		2.19	**
IDESP 3º ano EM	0.65	0.94	**	0.68		1.20	***
Proporção de servidores que exercem o cargo há menos de 1 ano:							
- professores de Matemática	0.48	0.25	**	0.32	*	0.28	**
- professores de Língua Portuguesa	0.32	0.26		0.38		0.28	
- diretores	0.40	0.44		0.51		0.40	
Número de alunos que há na escola	1,485	1,575		1,348		1,418	
Número de ciclos que a escola oferta	2.80	2.39	**	2.37	**	2.24	***
Na escola há laboratório de ciências	0.50	0.21	*	0.34		0.42	
Proporção de crianças pobres na escola	0.38	0.34	*	0.36		0.30	**
Proporção de crianças com irmãos de 0 a 5 anos	0.36	0.34		0.36		0.31	**

Nota: \*\*\* 1% de significância; \*\* 5% de significância; \* 10% de significância.

As características determinantes do desempenho escolar e da escolha das instituições do projeto são descritas abaixo para as escolas piloto, as demais escolas da Leste 3 e para as de baixo IDESP da RMSP, assim como as da RMSP de modo geral. Essas características incluem variáveis da escola e do aluno.

Entre as variáveis de escola, destaca-se o IDESP. Nas escolas piloto, esse índice é próximo ao das demais escolas da RMSP com baixo desempenho, em um patamar inferior aos outros grupos. Em termos de rotatividade de diretores e professores, as escolas piloto não são estatisticamente diferentes dos demais grupos de escolas, com exceção dos professores de Matemática, que possuem maior rotatividade entre

as escolas piloto. Tanto essas como as unidades da RMSP de pior IDESP estão em um nível mais elevado de rotatividade do corpo docente e gestor. Parece haver uma correlação entre essas variáveis, de forma que, ou o desempenho dos alunos é afetado por essa rotatividade, ou essa característica reflete as difíceis condições de trabalho nesses espaços de ensino.

O porte das escolas piloto, descrito em termos de número de alunos matriculados, não difere das escolas dos outros grupos. Contudo, como são apenas 10 escolas piloto, predomina entre elas a oferta dos três ciclos de ensino: Fundamental I e II e Ensino Médio, diferente da média das demais escolas, em que é mais comum a oferta de dois níveis.



A variável sobre presença de laboratório de ciências mostra que essa ainda não é uma realidade para a maior parte da rede de ensino estadual, composta por mais de 30% de famílias pobres e com um número maior de filhos (ter irmãos com uma idade muito abaixo é um indicador de família mais numerosa, frequentemente correlacionada com pior condição socioeconômica). Mais uma vez, as escolas piloto se assemelham nessas características às escolas da RMSP com pior IDESP.

Sobre as variáveis de aluno, o sexo das crianças se aproxima de 60% de meninas em todos os grupos. A idade média dos alunos é

relevante por ser um indicativo da repetência e também não se diferencia de maneira significativa entre os grupos. Entre os alunos da RMSP com pior IDESP, há maior percentual de crianças que frequentaram a pré-escola – quase 90% –, enquanto nos demais grupos de escolas esse percentual é inferior a 85%.

O nível de escolaridade da mãe é um bom indicativo socioeconômico do aluno. Essa variável é uniforme entre as escolas da Leste 3, porém, em comparação às mães dos alunos das outras escolas da RMSP, as mães dos alunos da Leste 3 têm escolaridade menor: com 17% sem qualquer instrução (9% na RMSP), 62%

■ **TABELA 5: DESCRIÇÃO DAS VARIÁVEIS DE ALUNOS DETERMINANTES DE DESEMPENHO ESCOLAR POR GRUPOS DE ESCOLAS – 9º ANO DO EF EM 2008**

Variáveis dos alunos determinantes do desempenho	Escolas Piloto	Leste 3	RMSP pior IDESP	RMSP	
Percentual de alunos do sexo feminino	0.57	0.58	0.58	0.57	
Idade média dos alunos	14.91	14.92	14.87	14.85	*
Percentual de alunos negros	0.38	0.35	0.41	0.43	
Percentual de alunos que fez pré-escola	0.82	0.80	0.89	0.85	***
Percentual de mães com Ensino Fundamental até 4a série	0.33	0.33	0.23	0.26	**
Percentual de mães com Ensino Fundamental entre 5a e 8a série	0.28	0.27	0.28	0.27	
Percentual de mães com até o Ensino Médio	0.17	0.21	0.28	0.29	***
Percentual de mães com até o Ensino Superior	0.03	0.03	0.06	0.06	**
Percentual de mães sem instrução	0.16	0.08	0.11	0.08	***

Nota: \*\*\* 1% de significância; \*\* 5% de significância; \* 10% de significância.

com até o Fundamental completo (54%), 18% com até o Médio completo (30%) e apenas 3% com Ensino Superior (7%).

Conclui-se que, em termos de variáveis de escola, as escolas piloto se assemelham mais às da RMSP com pior IDESP. No entanto, no que tange às variáveis de aluno, essas se aproximam mais às demais escolas da Leste 3, justificando assim o uso de dois grupos de controle para compreender as eventuais diferenças nos resultados.

## 2.2 DETERMINANTES DO DESEMPENHO ESCOLAR

Mas quanto cada variável descrita contribui de fato para o desempenho escolar dos alunos? Muitas delas são bastante significativas ao explicá-lo. Todavia, em conjunto, explicam apenas 4% de toda a variância das notas de Matemática e 5% no caso de Língua Portuguesa.

É interessante olhar os resultados das relações com a performance nas duas disciplinas em conjunto. Por exemplo, um aumento do número de professores de Matemática com apenas um ano de trabalho em uma escola reduz o desempenho dos alunos em 4 pontos. Essa variável não é relevante no caso dos professores de Língua Portuguesa. As

meninas têm um desenvolvimento pior que os meninos em média 3 pontos em Matemática, porém estão 10 pontos acima em Língua Portuguesa. Os alunos com idade maior que a esperada para o ano em que se encontram têm em média um desempenho pior nas duas disciplinas. Os alunos brancos performam melhor que os pardos e negros em média 6 pontos, assim como os alunos que fizeram pré-escola.

A escolaridade das mães influencia positivamente a atuação dos filhos na escola. Quanto maior sua instrução, melhor o desempenho do filho, chegando a 16 pontos a diferença de notas em Língua Portuguesa e 11 em Matemática entre filhos de mães sem escolaridade e aqueles cujas mães têm Ensino Superior.

Além disso, a proporção de alunos provenientes de famílias pobres, com renda familiar menor que R\$850, ou seja, um salário mínimo, tem um desempenho bastante inferior aos demais – 14 pontos em Matemática e 22 em Língua Portuguesa. Por fim, conforme esperado, a prática escolar do aluno está fortemente correlacionada com o desempenho médio da escola traduzido no IDESP referente àquela série – 1 ponto a mais no IDESP se traduz em 7 pontos a mais na nota de Matemática e 10 acima em Língua Portuguesa, em média.

## 2.3 AVALIAÇÃO DE IMPACTO

A metodologia adotada se baseou em duas estratégias: (i) a comparação dos alunos das escolas-piloto, chamados tratados, com um grupo de alunos de características muito semelhantes, chamados controle, e (ii) a observação desses estudantes antes e depois do início do projeto. A possibilidade de acompanhar seu desempenho nesses dois momentos diferentes garante que o resultado esteja livre de uma importante fonte de viés: um grupo de alunos depois do início do piloto diferente dos alunos antes de seu início em características que não foram afetadas pelo programa.

Para averiguar o impacto, deveríamos comparar o desempenho dos alunos nas escolas em duas situações: com e sem a presença do projeto. No entanto, é impossível observá-las simultaneamente. Se a exposição à política for aleatória, pode-se comparar dois grupos: um que recebeu o incentivo (grupo de tratamento) e outro que não (grupo controle). Depois, bastaria calcular a diferença do desempenho nas duas situações para avaliar o impacto do programa. Entretanto, em nosso estudo, a escolha não é aleatória, uma vez que apenas as piores escolas da Leste 3 receberam as tutorias e o coordenador de pais.

Para lidar com esse problema, a metodologia adotada foi o *matching*, ou pareamento. A abordagem consiste em identificar os indivíduos do grupo de controle mais parecidos com os indivíduos do grupo tratado, considerando características observáveis. Desse modo, a aleatoriedade da escolha estaria garantida, condicionada naquelas características.

O pareamento foi feito no nível do aluno<sup>3</sup> usando as variáveis determinantes do desempenho escolar descritas acima. As variáveis escolhidas para o pareamento<sup>4</sup> incluem a nota do aluno antes do programa, características dos alunos (gênero, raça, idade, escolaridade dos pais e se fez pré-escola), dos diretores (tempo de experiência), de professores (tempo de experiência na escola), de escolas (número total de matrículas, oferta de ensino, se tem laboratório de ciências e IDESP 2007) e também características médias do conjunto de alunos das escolas (proporção de famílias com renda menor que R\$ 850, proporção de famílias com filhos com idade entre zero e cinco anos)<sup>5</sup>.

Mesmo controlando pelas variáveis observadas, ainda podem existir aspectos não observados que afetam o resultado de interesse. Se esses forem fixos ao longo do tempo, podemos eliminá-los pelo método de Diferen-

■ **TABELA 6: IMPACTO MÉDIO SOBRE NOTAS  
(EM PONTOS NA ESCALA SAESP)**

	Comparação com RMSP pior Idesp	
	Língua Portuguesa	Matemática
7º ano do EF (2010)	3.1*	3.3**
9º ano do EF (2010)	-	-
9º ano do EF (2011)	-	-5.9**
3º ano do EM (2010)	-	-
3º ano do EM (2011)	5.7**	-

Nota: \*\*\* 1% de significância; \*\* 5% de significância; \* 10% de significância; - não significante estatisticamente.

ças em Diferenças (DD), ou seja, o quanto a diferença entre tratados e controles antes da intervenção mudou depois dela. Para tanto, precisamos de informação sobre o indicador de resultado antes e depois do programa que foi usado em primeira diferença como variável de resultado, equivalente à variação da nota, e também em nível quando a nota antes do programa entra no modelo como variável explicativa.

Com isso, eliminamos as diferenças observáveis e não observáveis entre os grupos trata-

dos e controles que não estão relacionadas com o programa, possibilitando estimar um impacto não visado e afirmar a relação de causa-efeito.

## Resultados

A tabela 6 traz os resultados do impacto causal do programa após o pareamento. Mostramos apenas os resultados que são estatisticamente diferentes de zero com uma confiança superior a 90%, ou seja, aqueles chamados significantes.

3 Foi feita uma tentativa de parrear previamente as escolas, no entanto, não foi bem sucedida uma vez que as características observáveis não diferenciam bem as escolas.

4 Para o grupo de controle composto pelas escolas da Leste 3, usamos apenas as características dos alunos e o IDESP, pois trata-se de um número menor de alunos que não comporta um pareamento usando muitas variáveis.

5 Este último conjunto de variáveis procura captar o grau de vulnerabilidade social dos alunos que frequentam estas escolas.

As estimativas mostram que o programa teve impactos positivos e estatisticamente significativos sobre o desempenho em Língua Portuguesa e Matemática para o grupo de alunos que estava no 5º ano em 2008 e, portanto, cursou o 6º e o 7º durante os dois anos de piloto avaliados. Na escala Saesp, o impacto sobre a nota de Língua Portuguesa foi de 3,1 pontos, e de 3,3 em Matemática.

Também foi verificado um impacto positivo para o 3º ano do Ensino Médio em 2011, após três anos de piloto, de 5,7 pontos em Língua Portuguesa. Esses resultados estão de acordo com os encontrados usando as demais escolas da Leste 3 como comparação. É interessante notar que não foi verificado impacto para o 3º ano do Ensino Médio em 2010, quando os alunos foram expostos a dois anos de programa. Conclui-se que há indicativos de que o último ano do projeto tenha sido mais eficaz para os alunos desse período escolar.

Por fim, foi encontrado um impacto negativo de 5,9 pontos em Matemática para o 9º ano em 2011, que não foi corroborado

pelos resultados usando a Leste 3 como comparação, o que pode sugerir uma especificidade da regional Leste 3 que dificulta o aprendizado. Os resultados encontrados para o 9º ano em 2010 não foram estatisticamente significativos para o nível usual de confiança (90%).

Considerando que, em média, a nota de Língua Portuguesa entre o 5º e o 7º anos dos alunos das escolas tratadas cresceu 29,5 pontos entre 2008 e 2010, o programa é responsável por 12% desse crescimento. Em Matemática, o ganho promovido foi de 11%, já que as notas das escolas tratadas variaram 26,5 pontos. Em relação às de Língua Portuguesa dos alunos do 3º ano, elas cresceram em média 33,4 pontos; logo, o impacto representa 17% do crescimento das notas entre 2008 e 2011.

O efeito mais contundente em Língua Portuguesa pode ser explicado pelo fato de que os professores dessa disciplina aceitaram melhor as tutorias e empregaram os ensinamentos de forma mais efetiva, segundo os relatos dos gestores e o relatório da avaliação qualitativa realizada pela empresa MOVE em 2012.

Entende-se que alunos mais novos convertem os benefícios do programa em desempenho com mais facilidade, uma vez que obtiveram resultados positivos desde o segundo ano do piloto, diferentemente dos estudantes do Ensino Médio. Não foi possível verificar esse resultado para o terceiro ano do piloto devido à indisponibilidade de dados para a respectiva linha de base<sup>6</sup>.

Com relação ao resultado negativo para os alunos que cursavam o 9º ano em 2010, levantamos uma hipótese adicional segundo a qual a ação dos coordenadores de pais, na tentativa de reter alunos que faltam a muitas aulas, ou abandonam a escola, reduz as notas na média, pois os alunos que mais faltam são aqueles que têm pior desempenho. As tendências das taxas de aprovação, reprovação e abandono corroboram essa hipótese. No entanto, não foi possível ter acesso a informações de abandono por aluno, o que impossibilitou uma análise mais aprofundada que possa confirmar a suposição.

## Discussão

A análise após dois anos do programa mostra impacto positivo e estatisticamente significativo no desempenho de Língua Portu-

guesa e Matemática para alunos do 7º ano do Ensino Fundamental, mas nenhum efeito relevante para alunos do 9º ano do Ensino Fundamental ou do 3º do Ensino Médio. A análise após três anos do programa mantém a ausência de resultado para o 9º do Ensino Fundamental, porém mostra impacto relevante em Língua Portuguesa para o 3º ano do Ensino Médio, sugerindo que o programa se tornou mais efetivo para essa série no último ano do piloto. Os resultados são robustos aos métodos e aos diferentes grupos de controles usados.

Portanto, juntando as evidências sobre o impacto do piloto do programa Excelência em Gestão Educacional, concluímos que os objetivos foram parcialmente atingidos. Identificou-se um amadurecimento das ações durante o piloto, o que pode explicar a presença de resultados para o Ensino Médio apenas no último ano do programa. Além disso, há relatos de que as tutorias fizeram mais efeito entre os professores de Língua Portuguesa. Os resultados apresentados são fruto de duas intervenções concomitantes, tutorias e coordenadores de pais, que podem ter, em curto prazo, efeitos contraditórios nas notas, ocasionando um resultado líquido difícil de interpretar.

6 A base de microdados do Saesp 2009 está incompleta.

Vale ressaltar que o próprio processo de avaliação traz elementos importantes que colaboram com a gestão da iniciativa, trazendo amadurecimento e trabalho conjunto entre os atores parceiros. No entanto, o grande desafio é a interpretação conjunta dos resultados. Os dados precisam ser analisados em sua complementaridade e em diálogo com o processo de gestão estabelecido ao longo da intervenção, o que trará ao gestor maior segurança na tomada de decisões.

Os resultados permitem concluir quais objetivos foram factíveis de serem concretizados e as hipóteses interpretativas con-

tribuem para o amadurecimento do real alcance do programa, favorecendo ajustes oportunos. Essas evidências foram importantes para justificar a viabilidade de expansão do projeto, atualmente em curso, demonstrando pontos fortes e ressalvas.

Na medida em que estratégias de tutoria e de coordenadores de país despertam interesse de outros sistemas educacionais, a avaliação do programa "Excelência em Gestão Educacional" oferece aprendizados relevantes para os próximos processos de planejamento, introdução e avaliação das novas parcerias com o setor público nessa área.

## Referências Bibliográficas

---

ABADIE, A., DRUKKER, D.; IMBENS, G. Implementing matching estimators for average treatment effects in Stata. *The Stata Journal*. College Station, v.4, n.3, p.290-311, 2001.

BOYD, D. Grossman, P. Lankford, H. Loeb, S. Wyckoff, J. (2008) Teacher Preparation and Student Achievement. NBER Working Paper No. 14314, September 2008.

ROSENBAUM, P.R., RUBIN, D.B. The central role of the propensity score in observational studies for causal effects. *Biometrika*, Great Britain, v.70, n.1, p.41-55, 1983.

MENEZES-FILHO, N. Os Determinantes do Desempenho Escolar do Brasil. Instituto Futuro Brasil, Ibmec-SP e FEA-USP. Mimeo.

SOARES, T. TEIXEIRA, L. (2006) Efeito do Perfil do Diretor na Gestão Escolar sobre a Proficiência do Aluno. Fundação Carlos Chagas, Estudos em Avaliação Educacional, v. 17, n. 34, maio/ago, 2006.

TAVARES, P. (2012) Os impactos de práticas de gestão escolar sobre o desempenho educacional: evidências para escolas públicas paulistas. São Paulo School of Economics, working paper 7/2012, julho/2012.