

EVOLUÇÃO DA PROFICIÊNCIA EM LEITURA

AVALIAÇÃO DE IMPACTO DO PROGRAMA APRENDER A ESTUDAR TEXTOS (TESTING A SCALABLE METHODOLOGY FOR IMPROVED READING COMPREHENSION DURING PANDEMIC RECOVERY) NA REDE MUNICIPAL DE VÁRZEA PAULISTA/SP

RELATÓRIO TÉCNICO



EVOLUÇÃO DA PROFICIÊNCIA EM LEITURA

AVALIAÇÃO DE IMPACTO DO PROGRAMA APRENDER A ESTUDAR
TEXTOS (TESTING A SCALABLE METHODOLOGY FOR IMPROVED
READING COMPREHENSION DURING PANDEMIC RECOVERY) NA
REDE MUNICIPAL DE VÁRZEA PAULISTA/SP

RELATÓRIO TÉCNICO

Realização:

Laboratório de Estudos e Pesquisas em Educação e Economia Social

Coordenação técnica:

Tiago Ferraz

Pesquisadores:

Déborah Cristina Rodrigues Vitor
Malu Villela Juliani

Financiamento:

Tinker Foundation

Agradecimentos:

Unidade Gestora Municipal de Educação de Várzea Paulista e Secretaria Municipal de
Educação de Francisco Morato

Ribeirão Preto | 2023

Sumário

1. Introdução	5
2. Definição e funcionamento dos instrumentos	7
3. Metodologia	9
4. Dados e características gerais da amostra	10
5. Resultados	12
5.1. Desempenho no ALA	12
5.2. Desempenho no PIRLS	20
5.3. Uma escala alternativa para a medida do PIRLS	26
5.4. Comentários adicionais	31
6. Considerações finais	33
REFERÊNCIAS	34
APÊNDICE	35
1. Atrito na amostra	35

1. Introdução

É durante a infância que são formadas as principais habilidades psicossociais dos indivíduos cujo desenvolvimento pode ser descrito em termos de competências¹ que permitem medir o grau em que um indivíduo é capaz de negociar de maneira adaptativa os desafios presentes (PIZATO et al., 2014). Dessa forma, podemos inferir que obstáculos ao desenvolvimento infantil têm o potencial de comprometer a capacidade do indivíduo em atingir seus objetivos e metas ao longo de toda a vida. A pandemia de COVID-19 tornou ainda mais severos os riscos enfrentados pelos estudantes brasileiros. Num país de amplas desigualdades como o Brasil, seria esperado que uma ruptura no processo de ensino, causada pelo fechamento das escolas neste período, tivesse promovido impactos graves e bastante heterogêneos entre os jovens.

Neste contexto torna-se ainda mais relevante fazer uso de ferramentas de mensuração do aprendizado e de se pensar em políticas públicas voltadas à recomposição dos processos de aprendizagem. Em particular, no que diz respeito à alfabetização, é de suma importância traçar diagnósticos mais precisos e desenhar programas capazes de reverter o enorme impacto sofrido pelas crianças². Potenciais atrasos no processo de alfabetização podem criar dificuldades que se manifestarão por muitos anos, em todas as etapas da educação e em várias outras dimensões das vidas destes jovens.

O presente relatório traz algumas evidências sobre os resultados apresentados pelo programa Aprender a Estudar Textos (AET), uma criação do Laboratório de Educação (Labeledu) em colaboração com Ana Teberosky, professora da Universidade de Barcelona, e Paola Uccelli, professora da Faculdade de Educação da Universidade de Harvard. O programa é voltado a estudantes do 4º e 5º ano do ensino fundamental,

¹ Competências são conjuntos de comportamentos aprendidos a partir de situações adversas presenciadas pelo indivíduo (BERRY; O'CONNOR, 2010).

² Um estudo recente mostra que o percentual de crianças de 7 e 8 anos que não sabem ler e escrever praticamente dobrou entre 2020 e 2022 (BOF et al., 2022).

uma etapa onde as crianças começam a ter seus primeiros contatos com textos informativos, que possuem uma estrutura diferente daquela com a qual este público pode estar habituado. O AET se propõe a fornecer um conjunto de atividades modelares, a serem realizadas durante as aulas de história³, para que os professores, em conjunto com os estudantes, possam superar os desafios da linguagem acadêmica presente nos textos didáticos, influenciando na qualidade das interações que professores e alunos desenvolvem em relação a estes textos com a finalidade de aprender. Dessa forma, o AET busca colaborar para que os estudantes se apropriem da linguagem como ferramenta de acesso ao conhecimento que lhes é disponível.

O município de Várzea Paulista (SP) foi selecionado para receber a intervenção, de modo que os professores das 49 turmas (denominado grupo de tratamento) deste município receberam as atividades do AET. Para estimar o efeito médio do programa, seria preciso encontrar um grupo de comparação adequado (denominado grupo de controle) que seja capaz de emular os resultados que seriam obtidos pelo grupo de tratamento na hipótese em que estes não tivessem passado pelo programa. Para compor o grupo de controle foram selecionadas 49 turmas da rede municipal de Francisco Morato, um município que apresenta proximidade geográfica e possui diversas outras características similares ao grupo de tratamento (cultura local, contexto socioeconômico, etc.).

A avaliação de impacto do programa conta com uma pesquisa de campo realizada com estudantes e professores do 4º ano do ensino fundamental das redes municipais de dois municípios do estado de São Paulo (Várzea Paulista e Francisco Morato). A coleta de dados foi realizada em dois momentos, sendo a etapa inicial (linha de base) realizada no período 4-19/abril de 2022 e etapa final (acompanhamento) entre os dias 16/novembro e 1/dezembro de 2022. Ao todo, participaram da pesquisa 2.743

³ Especificamente nesta implementação do programa optou-se por realizar as atividades durante as aulas de história, mas o conceito do programa pode ser adaptado para outras disciplinas também.

estudantes distribuídos em 98 turmas em 23 escolas, em ambos os municípios⁴. Foram aplicados três conjuntos de instrumentos para mensurar o nível de proficiência em leitura dos estudantes: (i) Avaliação da Linguagem Acadêmica (ALA); (ii) Estudo Internacional de Progresso em Leitura (PIRLS, na sigla em inglês); (iii) um questionário contextual. Estes instrumentos serão apresentados em mais detalhes na próxima seção.

Os principais resultados da análise de impacto mostram que o programa afetou de forma positiva e significativa o desempenho médio dos estudantes em ambas as dimensões mensuradas, com destaque para o nível de proficiência medida pelo PIRLS. Também foi possível verificar que o programa apresentou um efeito maior entre os estudantes que obtiveram as menores pontuações na coleta inicial, o que sugere que o programa se mostrou uma importante ferramenta para atacar desigualdades de aprendizado entre os estudantes.

2. Definição e funcionamento dos instrumentos

É a partir do 4º ano do Ensino Fundamental que os estudantes passam a ter contato com textos informativos, de diferentes áreas do conhecimento, que possuem uma estrutura um pouco diferente daquela contida nos textos narrativos aos quais eles estavam mais habituados. Para mensurar as habilidades de leitura acadêmica dos estudantes foram utilizados dois instrumentos. O primeiro, chamado Avaliação da Linguagem Acadêmica (ALA)⁵, é composto por nove atividades que se dividem em diferentes tarefas para medir habilidades como compreensão de estruturas complexas, conexão lógica de ideias, acompanhamento de cadeias de referência, organização global

⁴ A amostra selecionada é composta por 12 escolas em Várzea Paulista e 11 em Francisco Morato, sendo que em cada rede municipal foram selecionadas 49 turmas.

⁵ O ALA é instrumento criado por um grupo de pesquisadores da Faculdade de Educação da Universidade de Harvard e foi traduzido para o português pelo Labedu. É um instrumento desenhado para aplicação em estudantes de 4º a 9º ano do ensino fundamental.

de textos, interpretação de textos e identificação de registro acadêmico (não-vocabulário) e uma tarefa com foco em mensurar a compreensão do vocabulário metalinguístico (vocabulário). O ALA também inclui uma tarefa de decodificação, cujo objetivo é mensurar a capacidade do estudante em reconhecer letras e palavras.

O segundo instrumento aplicado é a Escala Internacional de Progresso em Leitura (PIRLS, na sigla em inglês). Composto por dois textos, sendo um deles um texto informativo e o outro um texto narrativo, este instrumento mensura o nível de proficiência em leitura mapeando o desenvolvimento de quatro grandes dimensões: (i) análise e avaliação de conteúdo; (ii) capacidade de realizar inferências diretas; (iii) capacidade de localizar e retirar informações explícitas e (iv) interpretar e integrar ideias e informações. Uma das vantagens do PIRLS é que se trata de um instrumento de avaliação padronizado, cuja pontuação pode ser transformada em uma escala comparável entre diferentes contextos e ao longo do tempo. No Quadro 1 a seguir são apresentadas as principais características de cada instrumento:

Quadro 1- Comparativo entre os instrumentos

Características	Instrumentos	
	ALA	PIRLS
Desenvolvimento	Pesquisadores da Faculdade de Educação de Harvard	IEA ⁶
Objetivo	Mensurar o desenvolvimento da linguagem acadêmica	Avaliar a proficiência em leitura
Composição	9 tarefas e decodificação de palavras	1 texto Informativo e 1 texto narrativo
Correção por TRI	Não	Sim
Dimensões	-	Possui 4 dimensões

Notas: Elaboração própria.

⁶ Sigla para *International Association for the Evaluation of Educational Achievement*.

3. Metodologia

Para que se possa mensurar o impacto de um programa é necessário comparar os indivíduos que passaram pela intervenção com um cenário alternativo, o contrafactual, em que ele não tivesse sido afetado pelo programa. Obviamente, tal cenário é uma construção meramente teórica que não pode ser observada na prática. O que é possível fazer é encontrar um grupo de comparação que seja suficientemente parecido com o grupo de tratamento, de modo que este grupo contrafactual possa emular o que haveria acontecido com os indivíduos afetados pelo programa no cenário alternativo em que estes não houvessem sido tratados.

Neste estudo, os estudantes do 4º ano do ensino fundamental no município de Várzea Paulista formam o grupo de tratamento e receberam as ações do programa. Para compor o grupo de comparação (controle) foram utilizadas escolas do município vizinho de Francisco Morato, pareadas previamente com base em características como número de turmas e quantidade de professores. A proximidade geográfica entre as cidades nos permite supor que ambas compartilhem de vários elementos em comum, como por exemplo fatores econômicos locais, demográficos e culturais.

Uma vez selecionado o grupo de controle é preciso estabelecer um método apropriado para estimar os efeitos do programa. A simples comparação de médias entre os dois grupos poderia levar a conclusões equivocadas, uma vez que existem vários possíveis vieses nesta comparação⁷. Neste estudo foi utilizado método de diferenças em diferenças para estimar o impacto do AET sobre o desempenho dos estudantes que passaram pelo programa. Essencialmente, o método se baseia na ideia de comparar as diferenças de médias, entre os grupos de tratamento e controle, no desempenho mensurado em dois momentos distintos: antes e após a implementação do programa.

⁷ Por exemplo, viés de seleção amostral ou um possível viés causado pela omissão de variáveis não observadas na estimação de um modelo de regressão linear.

Sob a hipótese de que, na ausência do programa, ambos os grupos teriam apresentado a mesma tendência de evolução no desempenho entre os dois momentos observados, o resultado obtido nesta dupla subtração é o efeito médio do programa sobre os indivíduos que foram tratados. O método de diferenças em diferenças é amplamente utilizado⁸ na literatura de avaliação de impacto, especialmente em contextos em que não é possível fazer uso de uma abordagem experimental⁹ para designar os grupos de tratamento e controle.

4. Dados e características gerais da amostra

A coleta de dados foi realizada em dois momentos distintos, linha de base e acompanhamento, no período de 04/04/2022 a 19/04/2022 e 16/11/2022 a 01/12/2022, respectivamente. Em ambos os períodos a pesquisa abrangeu 23 escolas, sendo 12 pertencentes a Várzea Paulista (grupo de tratamento) e 11 a Francisco Morato (grupo de controle). Apesar do número de escolas se diferenciar entre as duas cidades, em cada uma delas foram selecionadas 49 turmas. Os estudantes responderam nesses dois períodos a três questionários: ALA, PIRLS e um questionário contextual.

Considerando os dois momentos de coleta de dados, foram obtidas informações válidas de 2.240 estudantes (1.120 em cada etapa), com respostas para as tarefas do ALA e também para os itens do questionário contextual. A Tabela 1 ilustra as principais características destes estudantes, em cada um dos municípios. Há uma proporção um pouco maior de meninas e de estudantes que se identificam como brancos ou amarelos em Várzea Paulista do que em Francisco Morato. Além disto, a proporção de estudantes

⁸ Este método permite remover vários dos possíveis vieses que podem confundir a identificação dos efeitos do programa numa simples comparação de média. Em particular, elementos que poderiam afetar de maneira diferente os dois grupos (por exemplo, fatores institucionais como a qualidade da uma gestão municipal), mas que não variam ao longo do tempo são eliminados, permitindo assim uma melhor estimativa do impacto do programa.

⁹ Numa abordagem experimental, a designação do tratamento se dá de maneira aleatória, geralmente através de algum mecanismo de sorteio para determinar quais indivíduos estarão em cada grupo.

que declaram haver em casa algum beneficiário do Bolsa Família é maior em Francisco Morato do que em Várzea Paulista. Não há diferença estatisticamente significativa¹⁰ entre os grupos no que diz respeito à taxa de alfabetização das mães, que atinge a quase totalidade dos estudantes em ambos os municípios. Por fim, há uma pequena diferença na proporção de estudantes que afirmam já ter visto os pais lendo em algum momento, sendo a proporção ligeiramente menor em Várzea Paulista do que em Francisco Morato.

Tabela 1 - Características da Amostra do ALA por grupo

Características da amostra	Tratamento (Várzea Paulista) N = 1.234	Controle (Francisco Morato) N = 882	Diferença (tratamento - controle)
% meninas	55,3%	50,3%	5,0% *
% brancos / amarelos	51,6%	43,8%	7,8% *
% bolsa família ¹	56,9%	68,5%	-11,6% *
% mães alfabetizadas	98,3%	97,6%	0,7%
% exemplo de leitura ²	68,3%	68,9%	-0,5% *

Notas: * indica que a diferença é estatisticamente significativa a um nível de significância de 5%.

Apesar de a aplicação de ambos os instrumentos ter acontecido no mesmo dia, não foi possível obter respostas válidas de todos os mesmos estudantes que responderam ao ALA¹¹. O total de estudantes que apresentaram respostas válidas para o PIRLS foi de 2.116, considerando os dois períodos (sendo 1.055 em cada etapa). A Tabela 2 mostra que a amostra de estudantes que apresentou respostas válidas para o PIRLS não é muito diferente daquela que também respondeu ao ALA. Em todas as

¹⁰ Isto significa que, apesar de se observar uma diferença nas proporções médias entre os dois municípios, não é possível afirmar dentro de um grau de confiança estatística que a diferença entre ambos os grupos não é mero acaso.

¹¹ A coleta de ambos os instrumentos aconteceu no mesmo dia, porém a aplicação do ALA foi realizada primeiro. Após um breve intervalo, os estudantes responderam ao PIRLS. Possíveis problemas técnicos na transmissão dos dados ou mesmo a falta de resposta intencional por alguns estudantes fizeram com que houvesse uma pequena diferença no número de respostas válidas para cada instrumento.

dimensões comparadas, as proporções observadas em cada grupo são muito semelhantes em ambas as tabelas apresentadas, o que indica que não há prejuízo em comparar os resultados de ambos os instrumentos.

Tabela 2 – Características da Amostra do PIRLS

Características da amostra	Tratamento (Várzea Paulista) N = 1.234	Controle (Francisco Morato) N = 882	Diferença (tratamento - controle)
% meninas	54,1%	49,8%	4,3% *
% brancos / amarelos	51,9%	44,0%	7,9% *
% bolsa família ¹	56,2%	69,2%	-13,0% *
% mães alfabetizadas	98,4%	98,2%	0,2%
% exemplo de leitura ²	68,1%	68,9%	0,7%

Notas * indica que a diferença é estatisticamente significativa a um nível de significância de 5%.

Um importante ponto de atenção em relação a estas características é que se trata de informação obtida através do autorrelato dos próprios estudantes. Para crianças nesta faixa etária pode ser particularmente desafiador responder corretamente a tais informações. Por exemplo, em relação a existência de algum beneficiário do Bolsa Família no domicílio, aproximadamente 36% dos estudantes não souberam responder ou deixaram em branco. Isto significa que há um potencial erro de medida nestas variáveis que precisa ser considerado e analisado com cuidado.

5. Resultados

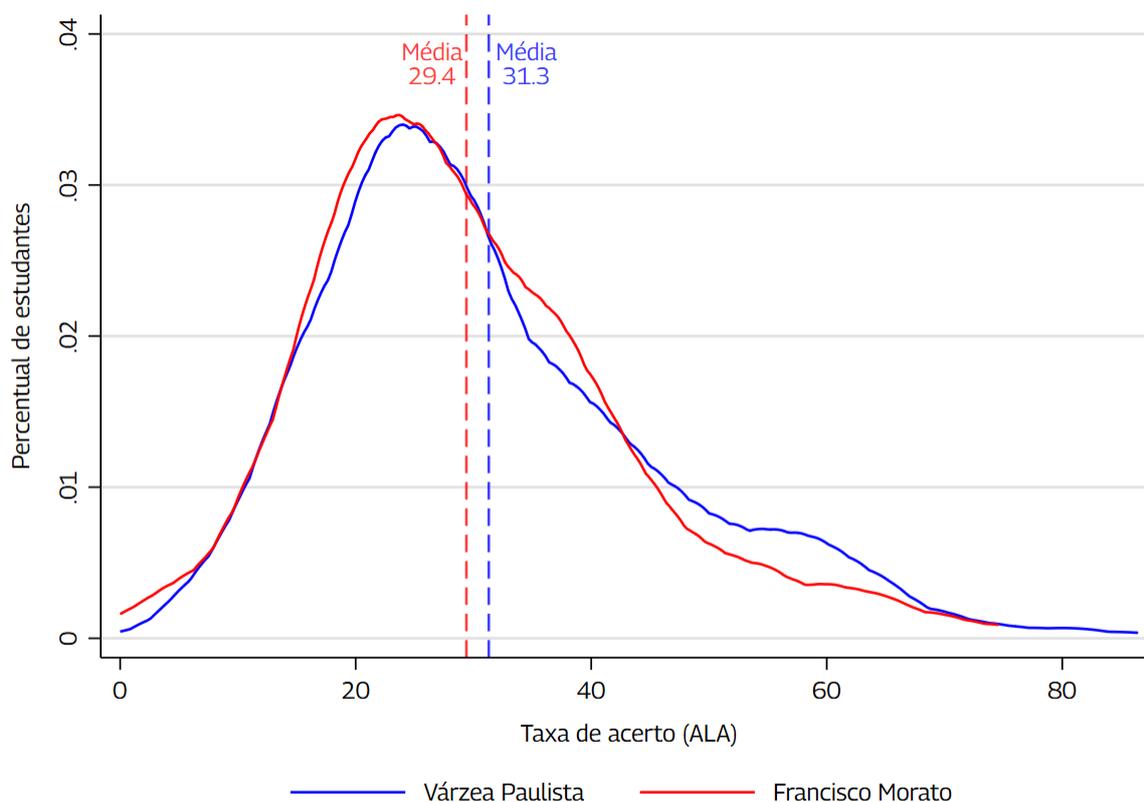
5.1. Desempenho no ALA

Como mencionado anteriormente, para que seja possível identificar o efeito do programa, é preciso que o grupo de controle seja uma boa base de comparação para o grupo de tratamento. Se antes do programa ser implementado o desempenho dos

estudantes em ambos os grupos for similar e atendida a hipótese de identificação, parece razoável imaginar que a diferença observada entre eles após o programa se deve à intervenção e não a outros fatores. A Figura 1 abaixo ajuda a ilustrar o fato de que, antes do início do AET, o nível de proficiência mensurado pelo ALA era de fato bastante similar entre os estudantes em ambos os grupos. A figura ilustra a distribuição da taxa de acertos nos itens do ALA em cada um dos municípios. A taxa de acertos, medida como percentual do total de acertos possíveis, nos itens do ALA foi de 29,4% entre os estudantes de Francisco Morato, enquanto entre os alunos de Várzea Paulista a taxa média foi de 31,3%. Não apenas a média foi parecida, como toda a distribuição de proficiência é bastante similar entre os estudantes de ambos os municípios. Apesar de parecer haver uma concentração um pouco maior de estudantes com taxas de acerto mais alta em Várzea Paulista (a partir de uma taxa de acerto superior a 45%), um teste estatístico¹² desta diferença não consegue rejeitar a hipótese de que ambas as distribuições sejam de fato iguais.

¹² Conhecido na literatura como teste de Kolmogorov-Smirnov, este permite comparar duas distribuições de variáveis aleatórias e testar sua igualdade. A partir de um certo valor crítico o teste permite afirmar que, com algum grau de confiança estatística, ambas as distribuições representam populações diferentes.

Figura 1 – Distribuição da taxa de acerto no ALA na etapa de linha de base, por município



Notas Resultado do teste de Kolmogorov-Smirnov: p-valor = 0,178 (superior ao valor crítico para um nível de significância de 5%).

Uma vez que apresentamos evidências que sugerem que ambos os grupos eram suficientemente similares antes do início do programa, podemos agora apresentar os resultados da estimação do efeito médio do AET sobre os estudantes que passaram pela intervenção. A Tabela 3 mostra que a implementação do programa levou a um aumento na proficiência mensurada pelo ALA, em média, de 2,3 pontos percentuais (p.p.) para os estudantes do grupo de tratamento. Isto equivale a cerca de 5,6% da média observada entre os estudantes de Várzea Paulista antes do início do programa.

Tabela 3 – Efeito do AET sobre o desempenho médio no ALA

	ALA (Total)	Vocabulário	Não-vocabulário	Decodificação de palavras
Efeito do AET	2.30** (0.88)	1.45 (1.32)	2.59*** (0.92)	0.78 (0.74)
N	2.240	2.240	2.240	1.924
R ²	0.078	0.050	0.076	0.061

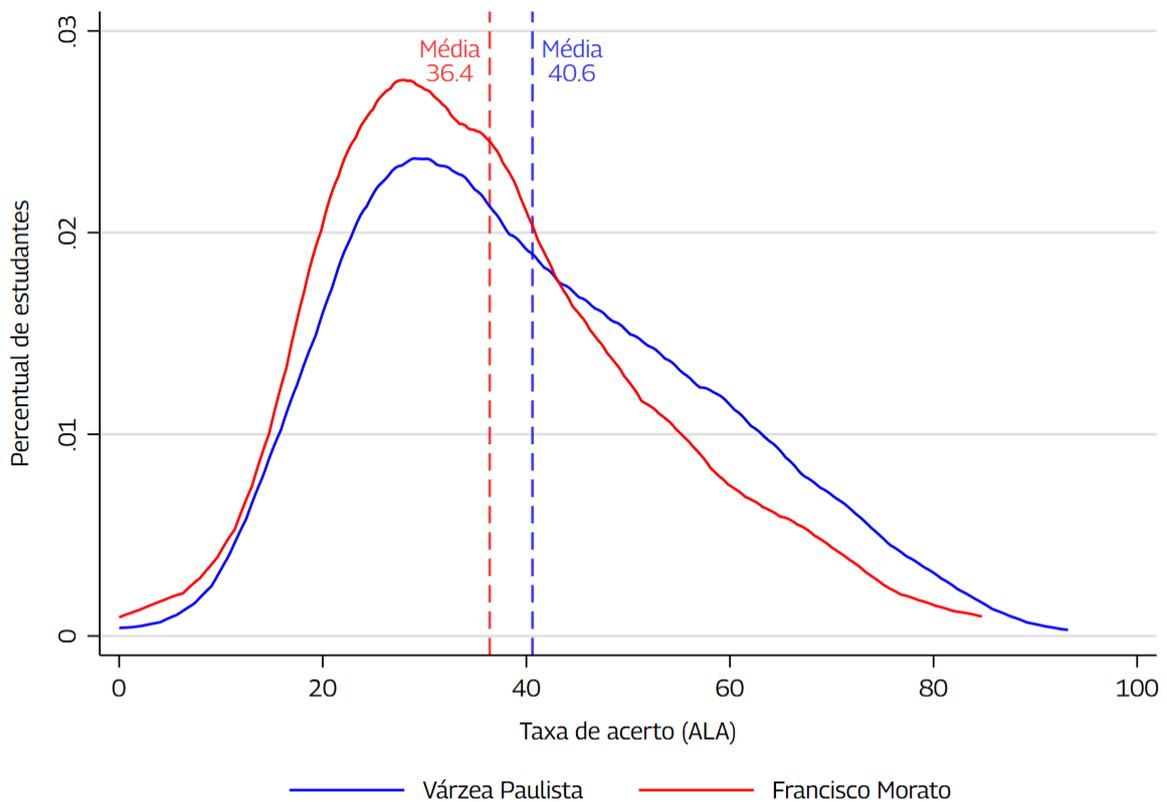
Notas: Pontuação medida em taxas de acerto. Erros-padrão agrupados por escola-turma em parênteses.

* Significante a 10%; ** Significante a 5%; ***Significante a 1%

É interessante notar que o impacto foi observado, essencialmente, em relação às tarefas que envolvem habilidades do tipo não-vocabulário, o que ajuda a reforçar a hipótese de que os efeitos observados se devem de fato ao programa e não a outros fatores que poderiam ter ocorrido ao mesmo tempo. Em relação aos itens do tipo vocabulário e à tarefa de decodificação de palavras, apesar de as estimativas pontuais serem positivas, estas não são estatisticamente significantes. Em outras palavras, não é possível afirmar com algum grau de confiança estatística que o efeito observado nestes desfechos seja diferente de zero.

Outro ponto relevante é o fato de que o programa parece ter gerado um impacto não apenas na média, mas em toda a distribuição de proficiência medida pelo ALA. A Figura 2 mostra que, após a intervenção do AET, a probabilidade de se encontrar um estudante com baixa proficiência em Várzea Paulista é menor do que em Francisco Morato, para qualquer nível de taxa de acerto observada. Além disto, a aplicação do teste de Kolmogorov-Smirnov nos resultados observados durante a etapa de acompanhamento permite afirmar, com algum grau de confiança estatística, que as distribuições são de fato diferentes.

Figura 2 – Distribuição da taxa de acerto no ALA na etapa de acompanhamento, por município



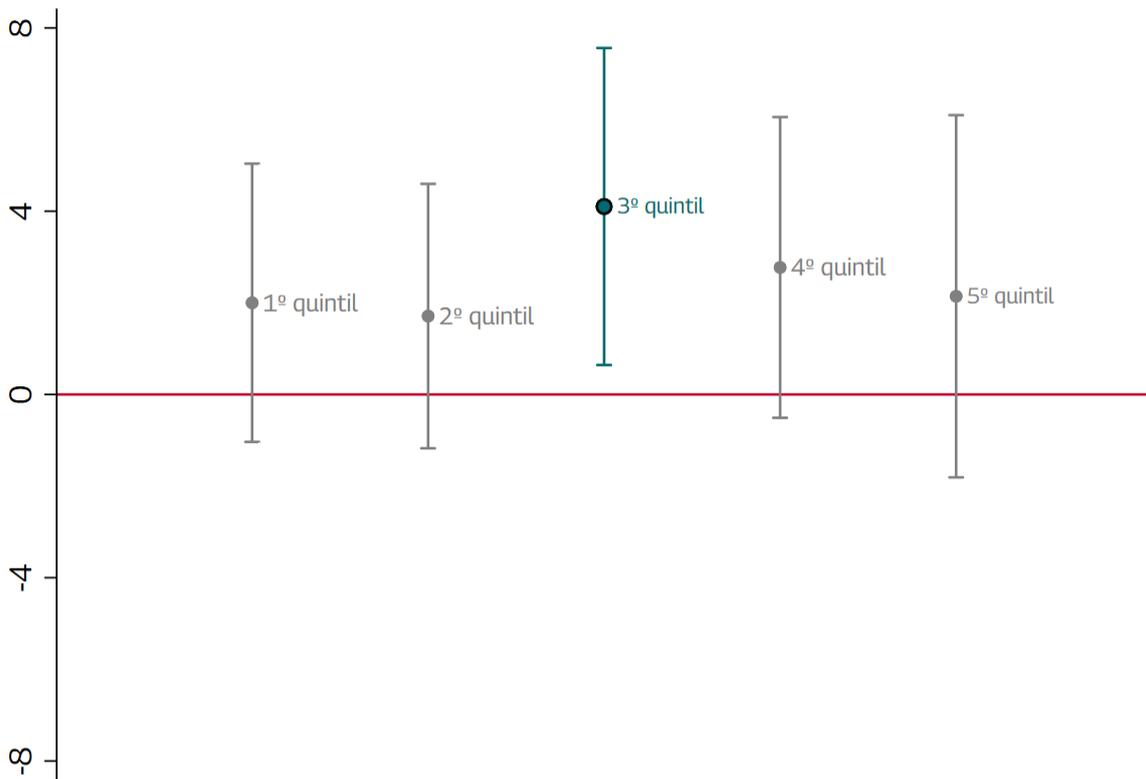
Notas Resultado do teste de Kolmogorov-Smirnov: p-valor = 0,001 (inferior ao valor crítico para um nível de significância de 5%).

Existem vários estudos evidenciando que o fechamento das escolas por causa da pandemia de COVID-19 teve impactos gravíssimos sobre o aprendizado dos estudantes e que nem todos foram afetados da mesma maneira. Neste sentido, torna-se essencial entender se o AET impactou de maneira diferente os estudantes que mais precisavam de apoio para recomposição da aprendizagem. Os resultados observados sugerem que, de fato, os estudantes que apresentavam o menor desempenho inicial foram aqueles que mais se beneficiaram do programa. Para estimar este efeito, os estudantes foram

agrupados em cinco cortes de igual tamanho, chamados quintis, em ordem crescente de desempenho no ALA na linha de base. A Figura 3 mostra o impacto do programa para cada um destes grupos separadamente. Barras totalmente acima da linha vermelha indicam que o efeito do programa para o referido grupo é positivo e estatisticamente significativo, enquanto barras que cruzam a linha vermelha indicam que não se pode rejeitar a hipótese de que o impacto do programa seja nulo. O que esta estimativa mostra é que os estudantes localizados no terceiro quintil, ou seja, aqueles cujo desempenho inicial os colocaria no intervalo entre 40% e 60% das melhores taxas de acerto foram os mais beneficiados pelo programa. Para os demais estudantes, apesar de estimativa pontual ser positiva, não é possível afirmar, dentro de uma margem de confiança estatística, que o efeito seja diferente de zero.

O efeito estimado para os estudantes no terceiro quintil da distribuição de desempenho inicial (4,1p.p.) equivale a cerca de 14,5% da taxa média de acertos observada neste grupo. Para se ter uma ideia melhor do que significa que esta magnitude, se tomássemos o estudante médio do terceiro quintil no grupo de controle e o colocássemos para receber a intervenção do AET, este impacto seria capaz de movê-lo para o quarto quintil, ou seja, no intervalo entre as 60%-80% melhores taxas de acerto.

Figura 3: Efeitos do AET sobre o nível de proficiência no ALA, por quintil de desempenho na linha de base

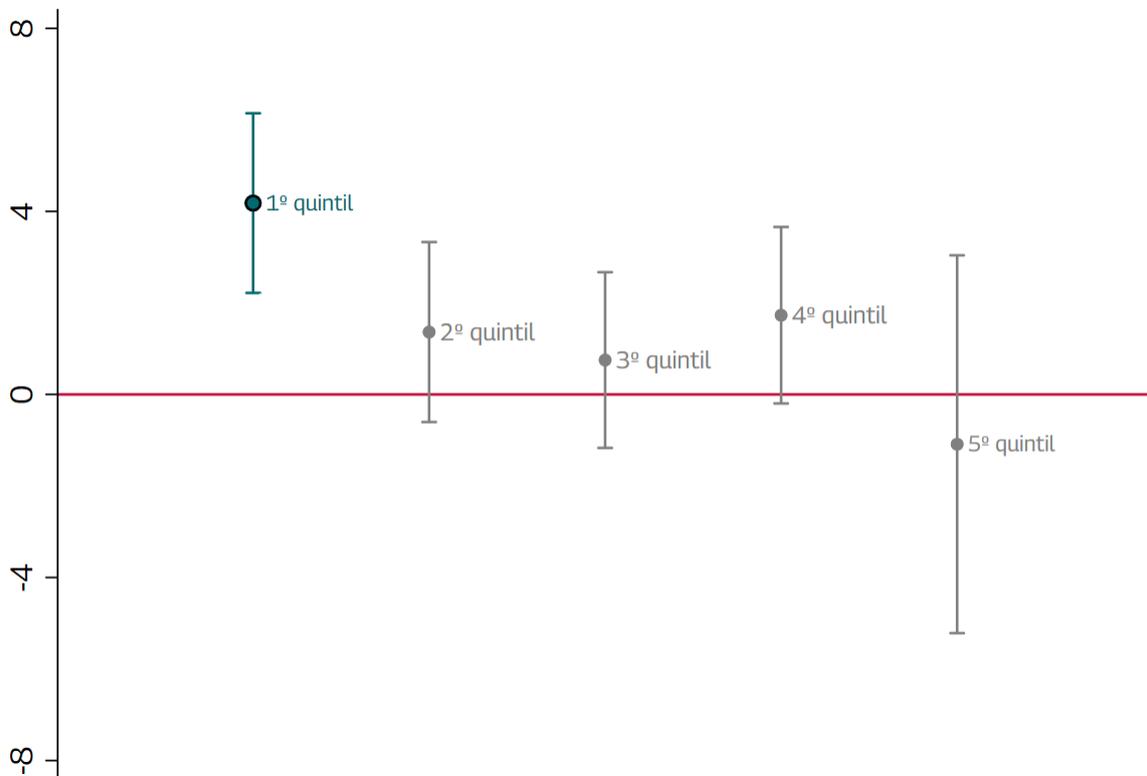


Notas: Pontuação medida em taxas de acerto. Limites das barras indicam intervalo de confiança de 95%. Intervalos acima de zero indicam efeito estatisticamente significativo.

O mesmo exercício foi realizado usando a taxa de acertos na tarefa de decodificação de palavras. Neste caso, o resultado apresentado na Figura 4 mostra que os estudantes mais beneficiados pelo programa foram aqueles localizados no primeiro quintil da distribuição de desempenho, ou seja, aqueles entre as 20% menores taxas de acerto na tarefa de decodificação. Neste caso, o efeito é bastante substancial, especialmente porque o desempenho médio dos estudantes neste grupo é muito baixo. O impacto gerado pelo programa (4,2p.p.) é equivalente a 5,7 vezes a taxa média de acertos dos estudantes no primeiro quintil, na etapa inicial. Fazendo aqui a mesma

analogia apresentada antes, este impacto seria suficiente para mover um estudante do primeiro para o segundo quintil da distribuição de desempenho (ou seja, saído da faixa entre 0-20% para os 20%-40% desempenhos mais baixos).

Figura 4: Efeitos do AET sobre o nível de proficiência na tarefa de decodificação de palavras do ALA, por quintil de desempenho na linha de base



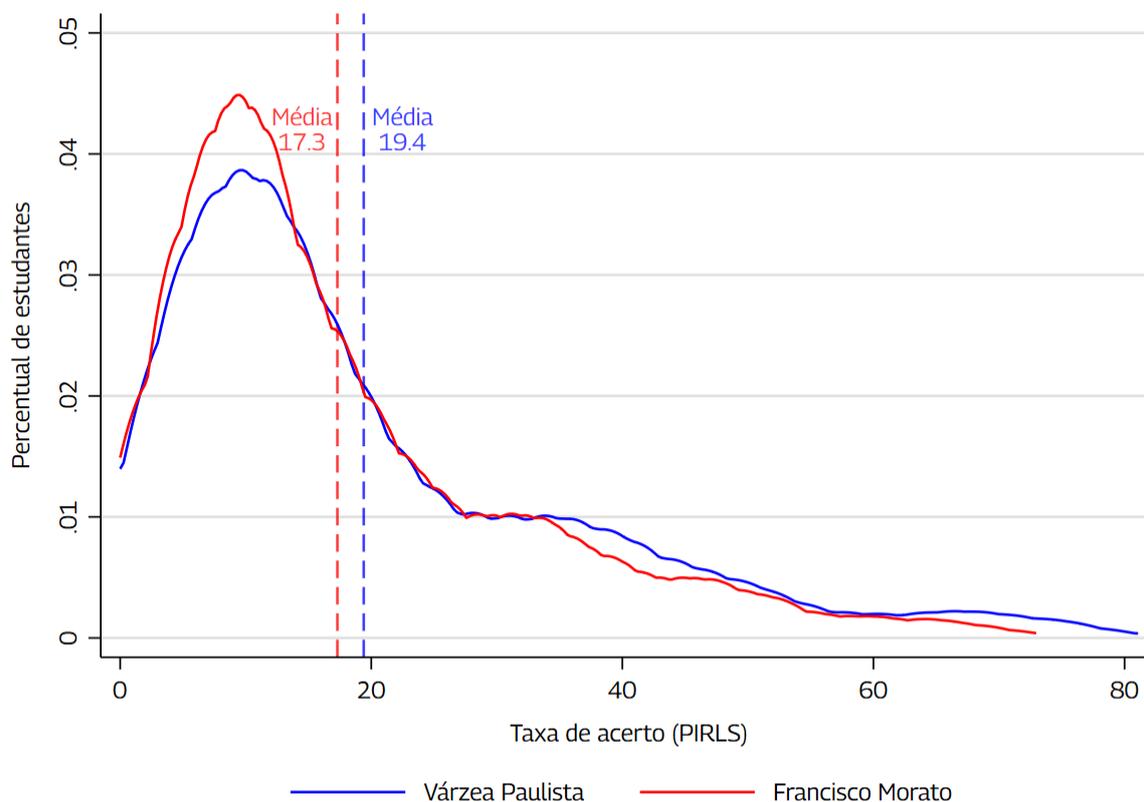
Notas: Pontuação medida em taxas de acerto. Limites das barras indicam intervalo de confiança de 95%. Intervalos acima de zero indicam efeito estatisticamente significativo.

5.2. Desempenho no PIRLS

O desempenho mensurado pelo ALA representa uma habilidade muito específica: o nível de proficiência do estudante na linguagem acadêmica. No entanto, uma das premissas do AET é de que melhorar o desempenho nesta dimensão é uma das maneiras de ajudar o estudante a desenvolver sua habilidade de leitura e compreensão de todo tipo de texto.

Para medir o impacto do programa sobre o desenvolvimento da capacidade leitora mais ampla dos estudantes, apresentamos as mesmas análises anteriores usando como variável de desfecho o nível de proficiência medido pelo PIRLS. A Figura 5 mostra a comparação das distribuições das taxas de acerto no PIRLS, em cada município, na etapa inicial da linha de base. Assim como no caso do ALA, a figura sugere que também nesta dimensão os estudantes do município de Francisco Morato parecem ser um bom grupo de comparação para os estudantes de Várzea Paulista, uma vez que a distribuição das notas é similar em ambos os municípios. Neste caso, percebe-se que as distribuições de notas estão ainda mais próximas. Para qualquer nível de taxa de acerto, a concentração de estudantes é muito parecida em ambos os grupos. Mais uma vez, o teste de Kolmogorov-Smirnov aplicado não permite rejeitar a hipótese de que as distribuições sejam iguais.

Figura 5 – Distribuição da taxa de acerto no PIRLS na etapa de linha de base, por município



Notas Resultado do teste de Kolmogorov-Smirnov: p-valor = 0,423 (superior ao valor crítico para um nível de significância de 5%).

A estimação do efeito médio do programa sobre o desempenho no PIRLS sugere que o AET de fato teve um impacto sobre o desenvolvimento da capacidade de leitura mais ampla dos estudantes. A Tabela 4 mostra que o programa promoveu uma melhora no desempenho dos estudantes de Várzea Paulista de 3,5p.p., equivalente a 18,2% da sua taxa de acertos média na linha de base.

Quando analisamos separadamente por tipo de texto, vemos que o impacto maior foi observado sobre o desempenho no texto narrativo (5,6p.p.), enquanto que no

texto informativo apesar de uma estimativa pontual positiva, não podemos rejeitar a hipótese de que não tenha havido qualquer efeito. Uma hipótese que podemos levantar para explicar esta diferença é de que os textos narrativos possuem uma estrutura mais familiar às crianças nesta faixa etária, o que tornaria mais fácil para que o programa ajudasse a desenvolver a capacidade de leitura deste tipo de texto.

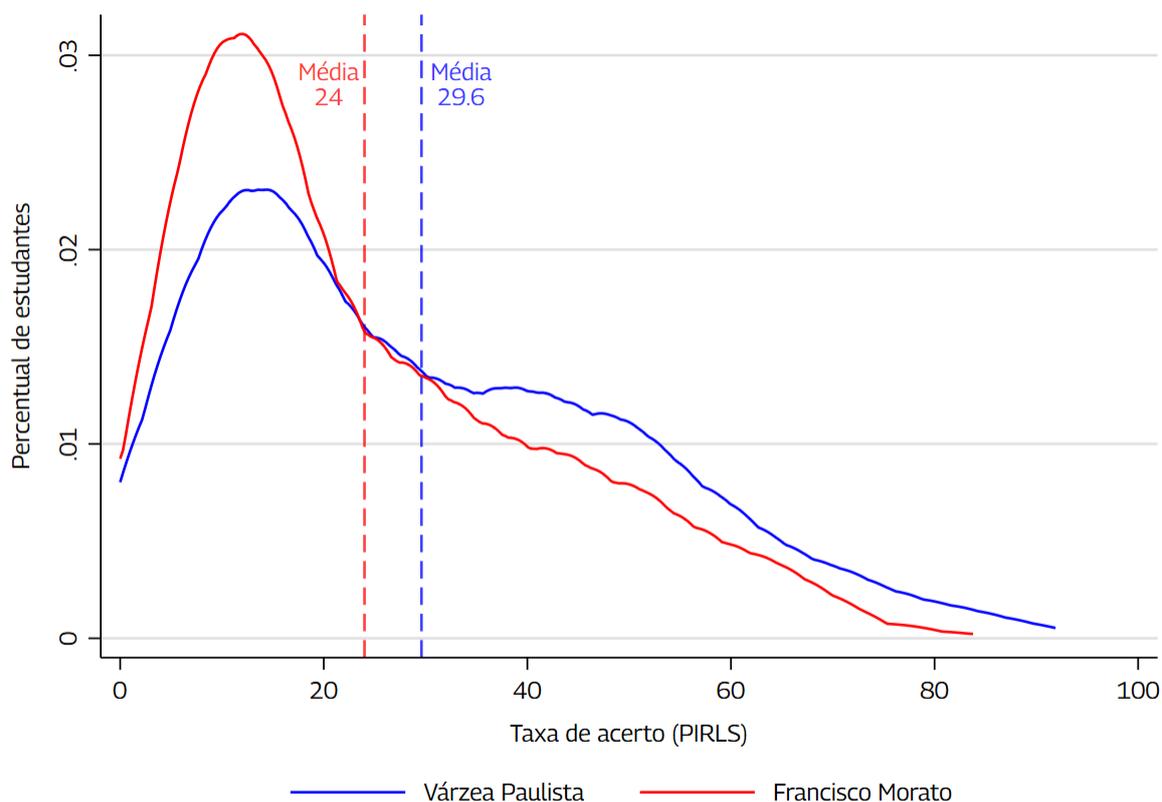
Tabela 4 - Efeito do AET sobre o desempenho médio no PIRLS

	(1)	(2)	(3)
	PIRLS (Total)	Texto Informativo	Texto Narrativo
Efeito do AET	3.54*** (1.15)	1.40 (0.98)	5.57*** (1.76)
N	2.110	2.110	2.110
R ²	0.072	0.045	0.071

Notas: Pontuação medida em taxas de acerto. Erros-padrão agrupados por escola-turma em parênteses. * Significante a 10%; ** Significante a 5%; ***Significante a 1%

Para além do efeito médio do programa, podemos ver na Figura 5 que também no caso do PIRLS toda a distribuição de desempenho parece ter sido impactada pelo programa. Comparando com a Figura 4, que mostra distribuição inicial, nota-se que há um deslocamento de ambas as distribuições para a direita. No entanto, a distribuição da taxa de acertos entre os estudantes de Várzea Paulista fica muito mais “descolada” na etapa posterior, indicando em Várzea Paulista há uma maior probabilidade de se encontrar um estudante de melhor desempenho do que em Francisco Morato, para qualquer nível de proficiência.

Figura 6 – Distribuição da taxa de acerto no PIRLS na etapa de acompanhamento, por município



Notas Resultado do teste de Kolmogorov-Smirnov: p-valor = 0,0002 (inferior ao valor crítico para um nível de significância de 5%).

O PIRLS também permite avaliar quatro habilidades específicas no desenvolvimento da proficiência em leitura: (i) análise de conteúdo; (ii) inferência direta; (iii) localização de informações; e (iv) interpretação de informações. A Tabela 5 apresenta o impacto do AET sobre o desempenho em cada uma destas habilidades, separadamente para cada tipo de texto. No Painel A vemos que no texto informativo apenas as dimensões de análise de conteúdo e capacidade de interpretar informações apresentaram um impacto positivo e estatisticamente significativo do programa (1,7p.p. e 2,4p.p., respectivamente). Já no texto narrativo, cujas estimativas são apresentadas

no Painel B, vemos que em todas as dimensões percebe-se um impacto do AET sobre o desempenho dos estudantes. A capacidade de localizar informações no texto foi a habilidade onde se observou o maior efeito (9,6p.p.), seguida pela capacidade de analisar conteúdo (6,4p.p.), capacidade de fazer inferências diretas (4,5p.p.) e capacidade de interpretar informações (3,4p.p.).

Tabela 5 – Efeito do AET sobre o desempenho médio no PIRLS

	(1)	(2)	(3)	(4)
Painel A: Texto informativo				
	Análise de conteúdo	Inferência direta	Localizar informações	Interpretar informações
Efeito do AET	1.69* (0.93)	-0.94 (1.78)	2.42 (2.18)	2.41*** (0.88)
N	2.110	2.110	2.110	2.110
R ²	0.006	0.027	0.029	0.027
Painel B: Texto narrativo				
Efeito do AET	6.39*** (2.41)	4.47** (2.12)	9.58*** (2.57)	3.40* (1.76)
N	2.110	2.110	2.110	2.110
R ²	0.026	0.061	0.062	0.058

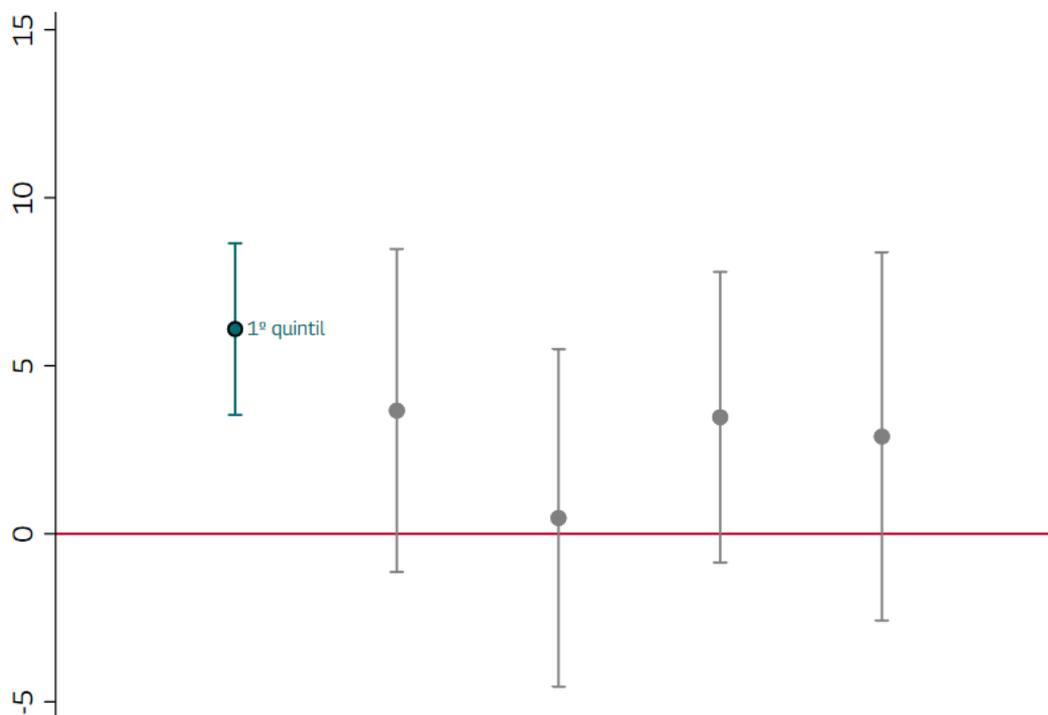
Notas: Pontuação medida em taxas de acerto. Erros-padrão agrupados por escola-turma em parênteses. * Significante a 10%; ** Significante a 5%; ***Significante a 1%

Da mesma maneira que foi feita com o ALA, repetimos o exercício de analisar o impacto do AET de acordo com o nível de proficiência apresentado na coleta de linha de base. Assim como antes, os estudantes foram agrupados em cinco cortes, ordenados pela taxa de acertos no PIRLS observada na etapa inicial. A Figura 7 mostra que mais uma vez, os estudantes que apresentavam maior dificuldade inicial foram os que mais se beneficiaram das ações do AET. Os estudantes localizados no primeiro quintil da distribuição (os 20% menores desempenhos iniciais) apresentaram um impacto positivo e estatisticamente significativo ao final do programa, da ordem de 6,1p.p. Para os demais

estudantes, apesar de a estimativa pontual ser positiva, não podemos descartar a hipótese de que o programa não tenha produzido nenhum efeito sobre eles.

Para se ter uma ideia da grandeza deste impacto, a taxa média de acertos entre os estudantes do grupo de controle (Francisco Morato) que estavam no primeiro quintil de desempenho, na etapa de linha de base, foi de 5,3p.p. Se tomássemos um indivíduo com este desempenho médio e o colocássemos no grupo de tratamento, o efeito causado pelo programa seria suficiente para movê-lo para o terceiro quintil (acima dos 40% menores desempenhos).

Figura 7: Efeitos do AET sobre o nível de proficiência no PIRLS, por quintil de desempenho na linha de base



Notas: Pontuação medida em taxas de acerto. Limites das barras indicam intervalo de confiança de 95%. Intervalos acima de zero indicam efeito estatisticamente significativo.

5.3. Uma escala alternativa para a medida do PIRLS

Por se tratar de um instrumento padronizado, construído a partir de um banco de itens testado e validado internacionalmente, o PIRLS permite usar a Teoria de Resposta ao Item (TRI) para produzir uma medida de proficiência que é comparável entre os diversos países que o aplicam e também ao longo do tempo. O uso desta medida é bastante interessante, pois permite avaliar o impacto do AET em uma unidade que facilita a comparação dos estudantes de Várzea Paulista e Francisco Morato com estudantes de qualquer outro país que tenha feito a aplicação do instrumento, inclusive o próprio Brasil que fez a primeira aplicação em escala nacional em 2021.

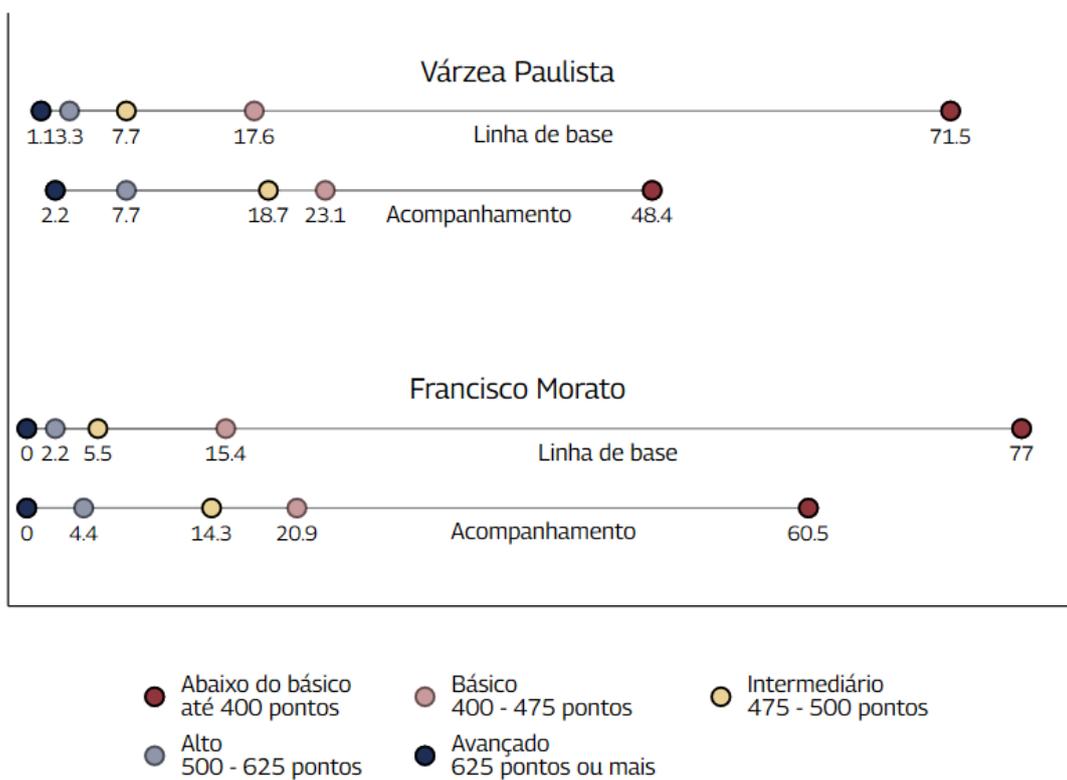
A pontuação do PIRLS traduzida para a escala global produz uma medida com média de 500 pontos e desvio-padrão de 100 pontos. Assim, se calcularmos o desempenho médio dos estudantes em cada um dos municípios podemos fazer algumas comparações que ajudem a situar melhor o tamanho do desafio enfrentado e dos resultados apresentados por estas redes.

Por exemplo, na coleta da linha de base a média apurada entre os estudantes de Francisco Morato foi de 350 pontos, enquanto em Várzea Paulista a média foi de 360 pontos. Se estes municípios fossem dois países¹³ comparados àqueles que aplicaram o PIRLS em 2021, teriam ficado à frente apenas da África do Sul, que obteve uma média de 288 pontos e ocupou a 43ª posição. Já na etapa seguinte, ao final do programa, a média dos estudantes de Francisco Morato foi de 387 pontos, pouco acima do desempenho da Jordânia (381 pontos) que ocupava a 41ª posição. Em Várzea Paulista, a média obtida na etapa de acompanhamento foi de 413 pontos, desempenho equivalente ao do Irã, que ocupava a 40ª posição, mas ainda abaixo da média do Brasil (419 pontos), que ocupava a 39ª posição no *ranking* de países que aplicaram o PIRLS.

¹³ Obviamente esta é uma generalização meramente ilustrativa e que, ainda assim precisa ser avaliada com cautela. Em Várzea Paulista, todas as escolas da rede municipal participaram da coleta de dados, enquanto em Francisco Morato foram selecionadas escolas comparáveis às de Várzea Paulista. Ainda assim, como um simples exercício comparativo a ideia permanece válida.

O PIRLS também permite construir uma escala para classificação do desempenho dos estudantes, de acordo com a pontuação obtida na medida ajustada pela TRI. A escala agrupa os estudantes em cinco categorias: (i) abaixo do básico; (ii) básico; (iii) intermediário; (iv) alto; e (v) avançado. A Figura 8 apresenta o percentual de estudantes de acordo com esta escala em cada um dos municípios, antes e depois da implementação do programa.

Figura 8: Distribuição do desempenho dos estudantes na escala PIRLS



Notas: Elaboração própria. A figura mostra o percentual de estudantes em cada categoria da escala, por município e etapa da coleta de dados.

Os dados apresentados até aqui sintetizam os resultados observados em cada um dos municípios, nas duas etapas de coleta de dados. No entanto, a questão mais relevante a entender é a contribuição do programa para estes resultados. A Tabela 6 apresenta a estimação do impacto do AET sobre o desempenho dos estudantes no PIRLS, mensurado com o ajuste pela TRI. Em linhas gerais, os efeitos vão na mesma direção vista quando a proficiência é medida em taxas de acerto, apesar de uma magnitude reduzida.

O impacto do AET sobre o desempenho geral dos estudantes de Várzea Paulista no PIRLS é de 16,1 pontos, equivalente a cerca de 4,5% da média que eles obtiveram antes do início do programa. Não foi encontrado impacto significativo sobre o desempenho no texto informativo, apenas no texto narrativo (23,1 pontos). Uma possível explicação para a diferença na magnitude entre as duas medidas, a taxa de acertos como percentual de acertos possíveis e o escore ajustado via TRI, é de que no primeiro caso todos os itens avaliados têm o mesmo peso em termos de medida de proficiência. Já no escore TRI, um dos parâmetros utilizados na calibração dos itens é a sua dificuldade. Neste caso, um estudante que acerto um item mais “difícil” demonstraria maior proficiência do que um estudante que acerta um item mais “fácil”.

Tabela 6 - Efeito do AET sobre o desempenho médio no PIRLS (TRI)

	(1)	(2)	(3)
	PIRLS (Total)	Texto Informativo	Texto Narrativo
Efeito do AET	16.09** (6.47)	0.88 (4.29)	23.10*** (7.68)
N	2.110	2.110	2.110
R ²	0.073	0.043	0.069

Notas: Pontuação medida pelo escore ajustado via TRI. Erros-padrão agrupados por escola-turma em parênteses. * Significante a 10%; ** Significante a 5%; ***Significante a 1%

A Tabela 7 mostra que quando analisamos o impacto sobre o desempenho dos estudantes em cada uma das dimensões, separadamente por tipo de texto, as mesmas conclusões anteriores permanecem. No texto informativo, apenas as dimensões capacidade de interpretar informações (11,24 pontos) e de analisar conteúdos (1,23 pontos) apresentaram impactos significativos. Já no texto narrativo, encontramos impactos do programa sobre todas as dimensões avaliadas.

Tabela 7 - Efeito do AET sobre o desempenho médio no PIRLS (TRI)

	(1)	(2)	(3)	(4)
Painel A: Texto informativo				
	Análise de conteúdo	Inferência direta	Localizar informações	Interpretar informações
Efeito do AET	1.23* (0.68)	-5.24 (4.36)	2.56 (4.79)	11.24*** (3.47)
N	2.110	2.110	2.110	2.110
R ²	0.006	0.023	0.004	0.040
Painel B: Texto narrativo				
Efeito do AET	12.74*** (4.61)	12.75** (6.31)	19.98*** (5.48)	9.87** (4.50)
N	2.110	2.110	2.110	2.110
R ²	0.027	0.057	0.063	0.058

Notas: Pontuação medida em escore ajustado via TRI. Erros-padrão agrupados por escola-turma em parênteses. * Significante a 10%; ** Significante a 5%; ***Significante a 1%

5.4. Interações com a tarefa de decodificação de palavras

Uma hipótese adicional que foi testada é a de que as atividades do AET, ao realizarem uma imersão na leitura de textos, poderia ajudar a desenvolver as habilidades mais básicas de decodificação de palavras, ajudando os estudantes a compreenderem melhor as tarefas e textos presentes nos instrumentos. Neste caso, seria de se esperar que os estudantes que mostraram desempenho mais alto nesta tarefa também apresentariam um impacto do programa ainda mais forte.

Para testar esta hipótese, utilizamos o método descrito por Holm & Breen (2023) para análise de mediação em estratégias de diferenças em diferenças. Essencialmente, o método aplicado aqui consiste em definir uma variável indicadora que assume o valor um quando o estudante apresentou um desempenho na tarefa de decodificação de palavras, na etapa posterior à intervenção do AET, que o coloca acima da mediana da taxa de acertos¹⁴. Esta variável indicadora é interagida com as demais variáveis do modelo, de modo que uma determinada combinação¹⁵ dos coeficientes estimados resulta na parcela do efeito do programa que é mediada pelo desempenho na tarefa de decodificação.

A Tabela 8 apresenta os resultados da estimação do efeito do AET mediado pela interação com o desempenho na tarefa de decodificação de palavras. O Painel A mostra que não foi observado nenhum impacto estatisticamente significativo no desempenho mensurado pelo ALA. No entanto, como mostra o Painel B, há um impacto de 5,1p.p. na taxa de acertos geral no PIRLS. Interessante que, neste caso, encontramos um impacto no desempenho dos estudantes em ambos os textos, o que sugere que a habilidade de decodificar palavras é um importante mecanismo para o desenvolvimento da capacidade leitora em geral.

¹⁴ Em outras palavras, quando o estudante tem um desempenho que o coloca na metade mais alta da distribuição da taxa de acertos nesta tarefa.

¹⁵ Para mais detalhes sobre a estimação destes modelos, ver Holm & Breen (2023).

Tabela 8 – Efeito do AET mediado pelo desempenho na tarefa de decodificação de palavras

	(1)	(2)	(3)
Painel A: Desempenho no ALA			
	ALA (Total)	Vocabulário	Não-vocabulário
Efeito mediado	3.16 (2.22)	3.19 (3.47)	3.15 (2.09)
N	1.664	1.664	1.664
R ²	0.244	0.157	0.237
Painel B: Desempenho no PIRLS			
	PIRLS (Total)	Texto Informativo	Texto Narrativo
Efeito mediado	5.09*** (1.93)	3.85** (1.55)	6.26** (2.83)
N	1.642	1.642	1.642
R ²	0.229	0.149	0.218

Notas: Pontuação medida em escore ajustado via TRI. Erros-padrão agrupados por escola-turma em parênteses. * Significante a 10%; ** Significante a 5%; ***Significante a 1%

5.5. Comentários adicionais

Até agora foram apresentados os principais resultados da avaliação de impacto do AET sobre o desempenho acadêmico dos estudantes de Várzea Paulista. Uma interpretação mais ingênua destes resultados seria de que a magnitude dos efeitos encontrados pode parecer pequena. No entanto, o efeito médio do programa equivale a cerca de 14% de um desvio-padrão na proficiência dos estudantes medida pelo ALA, ou de cerca de 21,5% de um desvio-padrão na proficiência medida pelo PIRLS, o que está bem alinhado com aquilo que se encontra na literatura especializada em relação a intervenções comparáveis ao AET¹⁶.

Os mesmos instrumentos coletados aqui também foram aplicados entre os estudantes do município de Caieiras (SP) em 2020 e 2021, numa primeira tentativa de avaliar o programa, que acabou não sendo implementado por conta do fechamento das escolas por causa da pandemia. Os estudantes ali foram avaliados no início do 4º ano

¹⁶ Vide Machin e McNally (2008), Muralidharan e Sundararaman (2011), Das et al. (2010) e Banerjee et al. (2007).

do ensino fundamental e depois novamente quase ao final do 5º ano. Usando a taxa de acertos do ALA, vemos que o impacto do AET sobre o desempenho dos estudantes de Várzea Paulista equivale a cerca de 70% do progresso dos estudantes de Caieiras. Evidentemente, esta comparação precisa ser olhada com cautela, porque os contextos são bastante distintos. No caso de Caieiras, os estudantes passaram praticamente todo o período compreendido entre as duas aplicações fora da escola. Em certo sentido, a evolução medida ali captura um pouco do que seria um “aprendizado natural”, fora da escola. Mesmo assim, é preciso levar em conta que estamos falando de um programa implementado durante cerca de sete meses impactando comparado a uma evolução natural de cerca de 19 meses.

Uma outra maneira de interpretar o efeito do programa é em termos de tempo-equivalente de aprendizagem. Se tomarmos a evolução dos estudantes de Francisco Morato entre as coletas de linha de base e acompanhamento na medida do ALA (7,05p.p.) como uma boa aproximação do que teria acontecido com os estudantes de Várzea Paulista na ausência do programa e considerarmos que o tempo de intervenção entre as duas etapas de coleta de dados foi de cerca de 7 meses, podemos fazer uma estimativa simplificada de que o impacto que AET proporcionou aos estudantes de Várzea Paulista (2,3p.p.) equivale a um ganho de 2,3 meses¹⁷ de aprendizagem. No caso da proficiência medida pela taxa de acertos no PIRLS, o impacto do AET equivale a cerca de 3,7 meses de aprendizagem ou 3,0 meses quando é usada a medida do PIRLS ajustada por TRI.

¹⁷ Esta é apenas uma estimativa muito simplificada, assumindo que na ausência do programa a evolução dos estudantes de Várzea Paulista seria equivalente à dos estudantes de Francisco Morato. Supondo que esta evolução pudesse ser convertida em um avanço linear de 1p.p. por mês (7,05p.p./7meses), teríamos o ganho aproximado de 2,3 meses.

6. Considerações finais

O presente relatório apresentou uma avaliação do impacto do programa Aprender a Estudar Textos (AET), uma intervenção voltada aos estudantes de 4º e 5º anos do ensino fundamental que tem como objetivo apoiar e subsidiar os professores para que seus estudantes consigam desenvolver as habilidades necessárias para os desafios da linguagem acadêmica.

Implementado em 12 escolas do município de Várzea Paulista (SP) ao longo de cerca de sete meses, o programa apresentou efeitos positivos sobre o progresso dos estudantes em duas medidas de proficiência distintas: o ALA, um instrumento voltado à mensuração das habilidades de leitura acadêmica; e o PIRLS, um instrumento padronizado que mede a competência leitora geral.

Para realizar a avaliação de impacto foram utilizados como grupo de controle os estudantes de escolas da rede municipal de Francisco Morato, município vizinho a Várzea Paulista, que não passaram pela intervenção do AET.

Os resultados encontrados mostram que o programa aumentou o desempenho dos estudantes do grupo de tratamento entre 14% e 21,5% de um desvio-padrão ou o equivalente a algo entre 2,3 e 3,7 meses de aprendizagem, dependendo da medida utilizada (ALA ou PIRLS). Além disto, o impacto do programa é maior para os estudantes de menor desempenho na coleta inicial de linha de base.

Em suma, o AET se mostrou uma intervenção efetiva para o desenvolvimento da habilidade de leitura dos estudantes, tanto na linguagem acadêmica como na leitura geral. Num contexto em que se fazem necessários grandes esforços para recomposição das aprendizagens após o período da pandemia, um programa com tais características parece ser uma boa alternativa à disposição dos formuladores de políticas públicas.

REFERÊNCIAS

ABEBERESE, B, A; KUMLER, J, T; LINDEN, L, L. Improving reading skills by encouraging children to read in school: a randomized evaluation of the Sa Aklat Sisikat reading program in the Philippines. **National Bureau of Economic Research: Working paper 17185**. Cambridge, jun. 2011.

BANERJEE, A; COLE, S; DUFLO, E. and Linden, L. Remedying Education: Evidence from Two Randomized Experiments in India. **Quarterly Journal of Economics**. Vol. 122, pp. 1235- 1264, 2007

BERRY, D; O'Connor, E. Behavioral risk, teacher-child relationships, and social skill development across middle childhood: A child-by-environment analysis of change. **Journal of Applied Developmental Psychology**, 31(1), 1-14. 2010

BOF, Alvana Maria; BASSO, Flavia Viana; DOS SANTOS, Robson. Impactos da Pandemia na Alfabetização das Crianças Brasileiras. **Cadernos de Estudos e Pesquisas em Políticas Educacionais**, v. 7, 2022.

DAS, J et. AL. School Inputs, Household Substitution, and Test Scores, **Working Paper**. Department of Economics. The University of California at San Diego, 2011.

HOLM, Anders; BREEN, Richard. Causal Mediation in Panel Data-Estimation Based on Difference in Differences. **Working Paper**. 2023.

MACHIN, S; MCNALLY, S. The Literacy Hour. **Journal of Public Economics**. Vol. 92, pp. 1441- 62, 2009.

Mullis, I. V., Martin, M. O., Foy, P., & Hooper, M. (2017). International results in Reading. Chestnut Hill, Mass.: **TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, and International Association for the Evaluation of Educational Achievement**.

MURALIDHARAN, K; SUNDARARAMAN, V. Teacher Performance Pay: Experimental Evidence from India. **Journal of Political Economy**, vol.119, pp. 39-77, 2011.

PIZATO, E; MARTURANP, E; FONTAINE, A. Trajetórias de habilidades sociais e problemas de comportamento no ensino fundamental: influência da educação infantil. **Psicologia: Reflexão e Crítica**. 27. 189-197. 2014.

APÊNDICE

1. Atrito na amostra

Considerados os dois períodos de coleta de dados, foram obtidas 3.314 observações em ambos os municípios. No entanto, como é comum em coletas de dados em larga escala, nem sempre é possível encontrar os mesmos indivíduos em ambas as etapas. Em particular, num contexto de reabertura das escolas após a pandemia, este desafio se torna ainda maior.

Esta perda de informação entre os períodos de coleta é conhecida como atrito amostral. O problema do atrito é especialmente relevante se as razões pelas quais o indivíduo sai da amostra forem correlacionadas com o mecanismo que determina aqueles que receberão ou não a intervenção. No caso do AET, a designação dos grupos de tratamento e controle foi feita através de uma negociação entre os implementadores do programa e as redes municipais. É difícil imaginar que esta designação tenha algo a ver com as razões pelas quais os estudantes saíram da amostra.

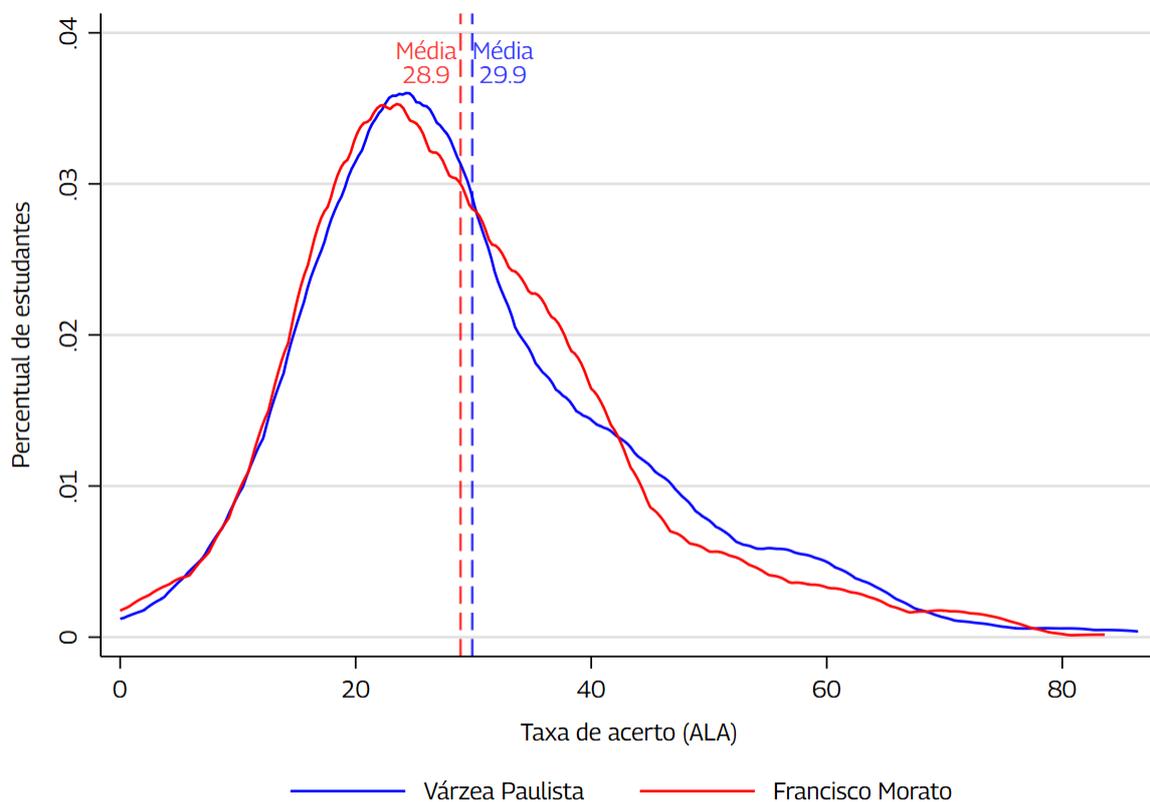
Ainda assim, ao longo de todo o relatório optamos por utilizar um painel balanceado de estudantes, ou seja, apenas as observações dos estudantes que participaram dos dois momentos de coleta. Esta estratégia permite evitar alguns dos possíveis vieses que seriam introduzidos no caso de usarmos todas as observações coletadas. No entanto, o custo desta alternativa é que os efeitos observados podem ser interpretados apenas como um efeito local, apenas sobre os estudantes que permaneceram na amostra.

Uma maneira de testar a robustez dos resultados é estimar os mesmos impactos usando todas as observações disponíveis, independente do atrito amostral. Se os resultados obtidos forem similares, teríamos uma indicação de que a estratégia adotada

de usar um painel balanceado pode ser uma boa alternativa para estimar os efeitos do programa.

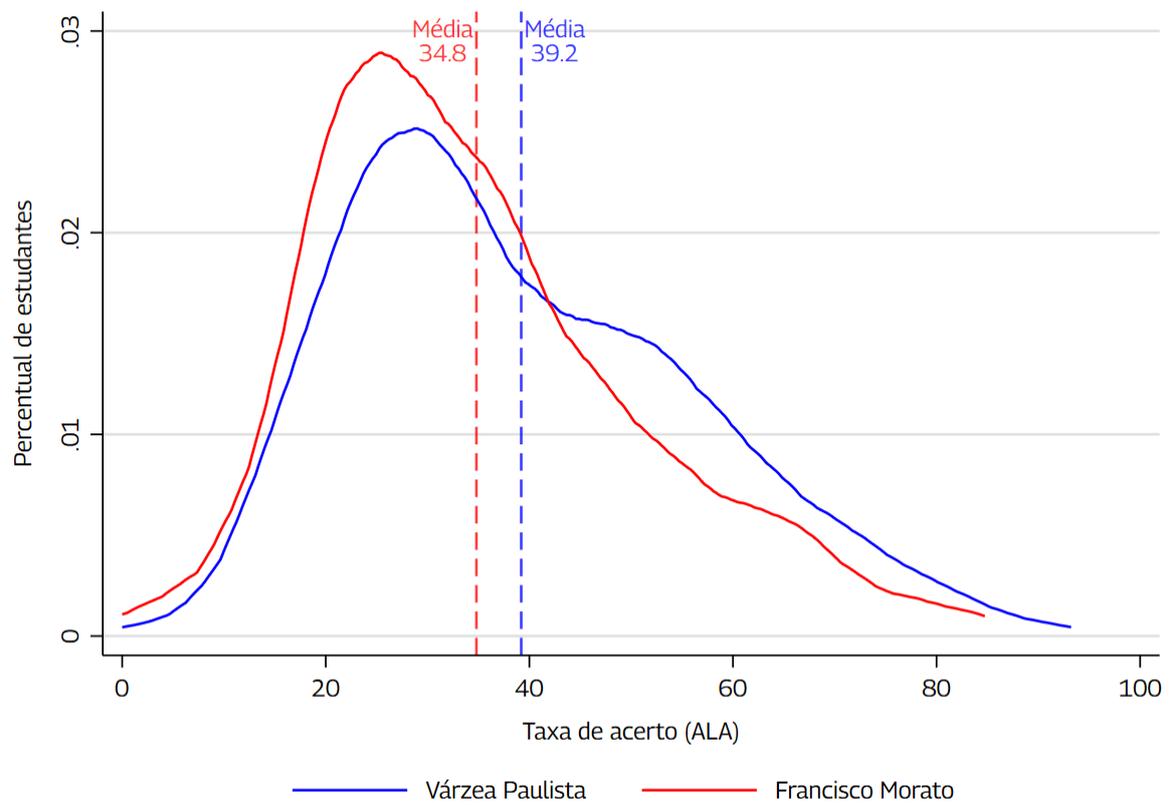
Em primeiro lugar, mostramos aqui que as distribuições de taxas de acerto no ALA e no PIRLS, quando consideramos todas as observações, não são muito diferentes daquelas observadas na amostra restrita do painel balanceado. De fato, as Figuras A1-A4 são muito similares às aquelas apresentadas anteriormente. O teste de Kolmogorov-Smirnov mostra que as distribuições não são estatisticamente diferentes nas coletas de linha de base, mas são diferentes nas coletas de acompanhamento.

Figura A1 – Distribuição da taxa de acerto no ALA na etapa de linha de base, por município (amostra completa)



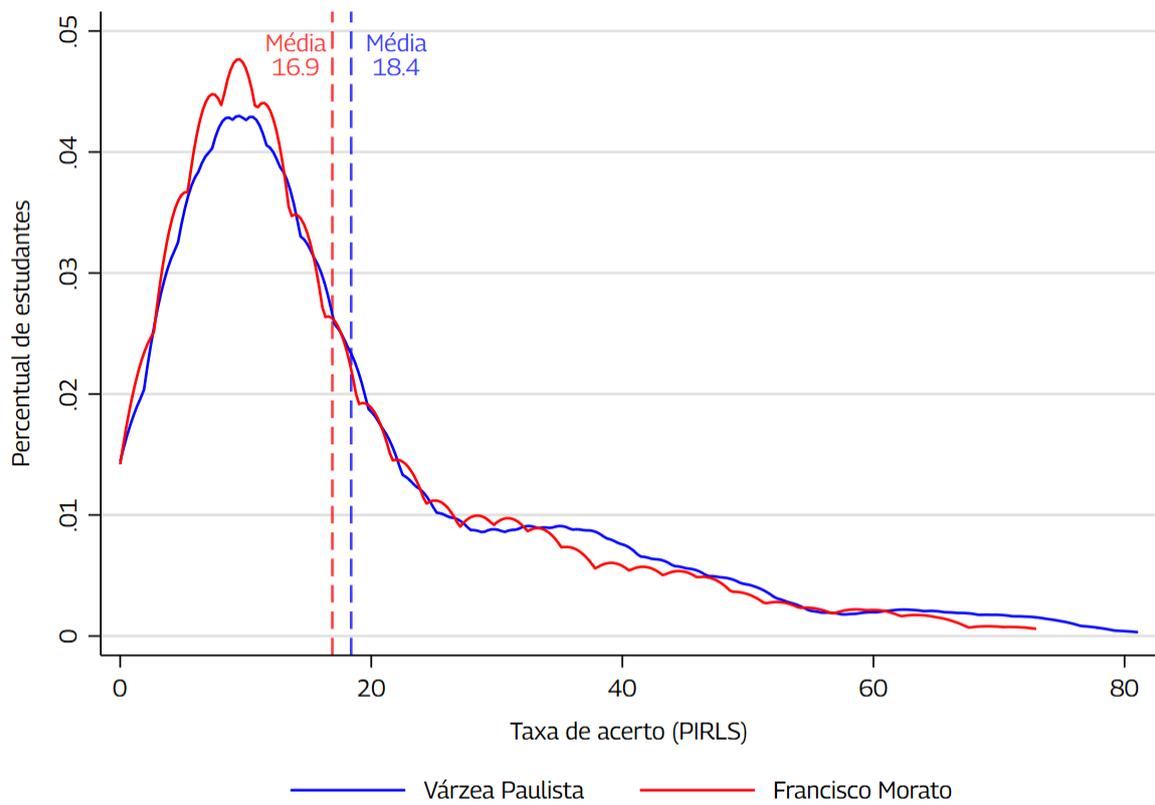
Notas Resultado do teste de Kolmogorov-Smirnov: p-valor = 0,229 (superior ao valor crítico para um nível de significância de 5%).

Figura A2 – Distribuição da taxa de acerto no ALA na etapa de acompanhamento, por município (amostra completa)



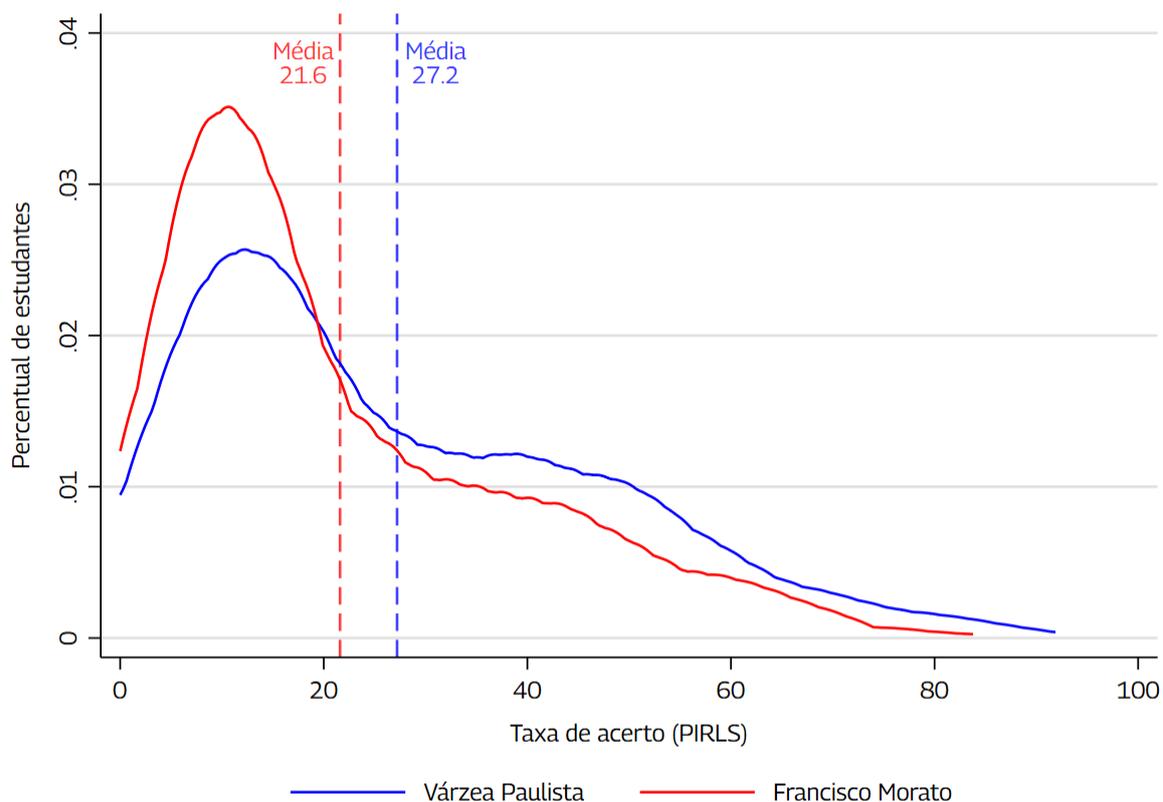
Notas Resultado do teste de Kolmogorov-Smirnov: p-valor = 0,000 (inferior ao valor crítico para um nível de significância de 5%).

Figura A3 – Distribuição da taxa de acerto no PIRLS na etapa de linha de base, por município (amostra completa)



Notas Resultado do teste de Kolmogorov-Smirnov: p-valor = 0,555 (superior ao valor crítico para um nível de significância de 5%).

Figura A4 – Distribuição da taxa de acerto no PIRLS na etapa de acompanhamento, por município (amostra completa)



Notas Resultado do teste de Kolmogorov-Smirnov: p-valor = 0,000 (inferior ao valor crítico para um nível de significância de 5%).

As Tabelas A1-A3 apresentam os resultados principais, utilizando a amostra completa, sobre o nível de proficiência dos estudantes medida pelo ALA e pelo PIRLS. Em linhas gerais, os resultados encontrados são bem próximos daqueles obtidos com a amostra restrita. De fato, o que este exercício sugere é que ao restringir a amostra estamos apresentando uma estimativa mais conservadora do impacto do programa.

Tabela A1 – Efeito do AET sobre o desempenho médio no ALA

	ALA (Total)	Vocabulário	Não-vocabulário	Decodificação de palavras
Efeito do AET	3.39*** (1.01)	2.77** (1.23)	3.60*** (1.04)	1.55** (0.64)
N	3.314	3.314	3.314	3.138
R ²	0.068	0.044	0.065	0.049

Notas: Pontuação medida em taxas de acerto. Erros-padrão agrupados por escola-turma em parênteses. * Significante a 10%; ** Significante a 5%; ***Significante a 1%

Tabela A2 – Efeito do AET sobre o desempenho médio no PIRLS

	(1)	(2)	(3)
	PIRLS (Total)	Texto Informativo	Texto Narrativo
Efeito do AET	4.15*** (1.24)	1.37 (1.01)	6.78*** (1.78)
N	3231	3231	3231
R ²	0.052	0.030	0.053

Notas: Pontuação medida em taxas de acerto. Erros-padrão agrupados por escola-turma em parênteses. * Significante a 10%; ** Significante a 5%; ***Significante a 1%

Tabela A3 – Efeito do AET sobre o desempenho médio no PIRLS

	(1)	(2)	(3)	(4)
Painel A: Texto informativo				
	Análise de conteúdo	Inferência direta	Localizar informações	Interpretar informações
Efeito do AET	1.79** (0.78)	0.33 (1.66)	1.23 (2.03)	2.07** (0.92)
N	3231	3231	3231	3231
R ²	0.006	0.018	0.017	0.019
Painel B: Texto narrativo				
Efeito do AET	7.50*** (2.31)	6.66*** (2.08)	10.13*** (2.48)	3.86** (1.54)
N	3231	3231	3231	3231
R ²	0.022	0.044	0.045	0.044

Notas: Pontuação medida em taxas de acerto. Erros-padrão agrupados por escola-turma em parênteses. * Significante a 10%; ** Significante a 5%; ***Significante a 1%

LEPES

LABORATÓRIO DE ESTUDOS E PESQUISAS
EM EDUCAÇÃO E ECONOMIA SOCIAL

www.lepes.fearp.usp.br

 @LEPESUSP

 @LEPESUSP

 /LEPESUSP

 @LEPESUSP