

**ACOMPANHAMENTO LONGITUDINAL DO DESEMPENHO ESCOLAR DE ALUNOS DA  
REDE PÚBLICA DE ENSINO FUNDAMENTAL DO RECIFE**

**RECIFE, 2021**

**COORDENAÇÃO:**

Isabel Pessoa de Arruda Raposo – Núcleo de Estudos em Estatísticas Sociais (NEES/  
FUNDAJ)

Michela Barreto Camboim Gonçalves – Núcleo de Estudos em Estatísticas Sociais  
(NEES/ FUNDAJ)

**PESQUISADORES:**

Luis Henrique Romani de Campos – (NEES/ FUNDAJ)

Morvan de Mello Moreira - Centro de Estudos de Cultura, Identidade e Memória  
(CECIM/ FUNDAJ)

Patricia Bandeira de Melo – (NEES/ FUNDAJ)

Magda Caldas Neto - Coordenação Geral de Estudos Econômicos e Populacionais  
(CGEP/ FUNDAJ)

Ivone Medeiros - (CGEP/ FUNDAJ)

**COLABORADORES EXTERNOS:**

André Maia – Fundacentro/ BA

Cristiano Ferraz – UFPE

Diego Firmino da Costa Silva – UFRPE

Maurício Haas Bueno – PPGPsi/UFPE e Instituto Brasileiro de Avaliação Psicológica –  
IBAP

Raul da Mota Silveira Neto – PIMES/ UFPE

Tatiane de Almeida Menezes – PIMES/ UFPE

## SUMÁRIO

APRESENTAÇÃO.....	4
1 INFORMAÇÕES GERAIS SOBRE O CAMPO DE 2013.....	5
2 AMOSTRA DE 2013 .....	6
2.1 Distribuição das escolas por RPAs.....	8
3 BASES DE DADOS E DICIONÁRIOS DE 2013 .....	9
4 INFORMAÇÕES GERAIS SOBRE OS CAMPOS DE 2017 E 2018.....	10
5 AMOSTRA DE 2017 E 2018.....	11
6 BASES DE DADOS E DICIONÁRIOS DE 2017 E 2018 .....	16
REFERÊNCIAS .....	17

## APRESENTAÇÃO

A pesquisa **Acompanhamento Longitudinal do Desempenho Escolar de Alunos da Rede Pública de Ensino Fundamental do Recife** é um projeto desenvolvido pelo Núcleo de Estudos em Estatísticas Sociais da Diretoria de Pesquisas da Fundação Joaquim Nabuco que tem como principal objetivo medir o desempenho acadêmico dos estudantes, sob duas óticas. Na primeira, a partir da ótica do aluno, a pesquisa analisa como as características sociais, econômicas e culturais do estudante e sua família afetam seu desempenho acadêmico. A segunda ótica se baseia no estudo do impacto e importância das políticas educacionais e insumos escolares para a performance acadêmica dos estudantes.

Iniciada em 2013, a pesquisa coletou dados para os anos de 2013, 2017 e 2018 fazendo um acompanhamento longitudinal com mais de 8 mil alunos de 120 escolas públicas do Recife ao longo dos 6º e 7º anos do ensino fundamental, selecionados a partir de uma amostra estratificada definida com base no número de matrículas dos alunos nessas séries e desempenho na Prova Brasil. São investigados estudantes de uma mesma turma, em uma mesma escola ao longo dessas duas séries, o que permite um melhor controle da influência do ambiente da turma e da escola. Na primeira rodada de 2013, foram pesquisados alunos do 6º ano do ensino fundamental de escolas públicas do Recife. Nesse ano, o acompanhamento longitudinal do aluno se deu apenas nas avaliações de matemática aplicadas ao mesmo educando no início e final do ano letivo, enquanto as informações de questionários para alunos, pais ou responsáveis, professores e diretores foram coletadas em um único momento no ano de 2013. Os microdados e todo material auxiliar para essa fase da pesquisa já se encontram publicados e disponíveis no site da Fundaj. Em 2017 e 2018, foi pesquisado um novo grupo de estudantes do 6º ano (2017) e 7º (2018) de escolas públicas do EF do Recife. Nessas rodadas, o acompanhamento longitudinal do aluno, seus pais ou responsáveis, professores e diretores ocorreu tanto para as avaliações de português e matemática, quanto para os questionários administrados. Para essa segunda etapa da pesquisa, os microdados ainda serão publicados no site da Fundaj.

O caráter longitudinal da pesquisa com dados em painel confere a esta proposta de trabalho uma característica inédita no Brasil, e que do ponto de vista metodológico, permite efetivamente isolar o efeito das políticas educacionais, aos atributos relacionados à família, ambiente social e cultural dos alunos e das escolas. O estudo

destaca-se também por aferir um conjunto de informações inovadoras em surveys educacionais de larga escala no Brasil, tais como a mensuração de habilidades não cognitivas ou socioemocionais, o levantamento da rede de amizades do aluno em sala de aula, o georreferenciamento dos endereços residenciais dos alunos e das escolas e a coleta de informações sobre a saúde dos alunos e seus pais. Em um módulo qualitativo da pesquisa (campos de 2017 e 2018), é investigada a influência do capital cultural familiar sobre os resultados escolares através de entrevistas com subamostras de distintos perfis de capital cultural familiar do aluno.

Além da apresentação da pesquisa, esse documento traz informações gerais sobre os campos de 2013, 2017 e 2018 e a metodologia da amostra aplicada para esse ano.

## **1 INFORMAÇÕES GERAIS SOBRE O CAMPO DE 2013**

Este capítulo tem como objetivo apresentar uma síntese comentada da pesquisa de Acompanhamento Longitudinal do Desempenho Escolar dos Alunos da Rede Pública de Ensino Fundamental do Recife realizada no ano de 2013. Nesta fase, a pesquisa contou com questionários aplicados aos alunos, aos seus pais ou responsáveis e às escolas (na pessoa do professor de matemática e diretor), bem como com a realização de dois testes de matemática, realizados no início e final do ano letivo.

A Tabela 1 traz o perfil geral da amostra investigada em 2013. Ao final da pesquisa, nossa amostra contemplou cerca de 3.568 alunos<sup>1</sup>, de 146 turmas do 6º ano do ensino fundamental de 120 escolas da rede pública do Recife. Os estudantes pesquisados se submeteram a duas provas de matemática aplicadas no início e final do ano letivo de 2013. Essas avaliações foram desenvolvidas pela Fundaj com base nos parâmetros curriculares da educação básica definidos pela Secretaria de Educação e Esportes de Pernambuco [PERNAMBUCO, (2012a), PERNAMBUCO, (2012b)]<sup>2</sup>. Cada prova foi composta por 10 questões (ou itens), com a escala da nota variando de 0 a 100, em que a nota 0 representa o caso do estudante que não acertou nenhuma questão e 100 o caso daquele que acertou todos os itens da prova. A avaliação feita no início do

---

<sup>1</sup> Esse número é obtido com base na quantidade de alunos que fizeram as duas provas de matemática e que tiveram seu questionário e de seus pais devidamente preenchido.

<sup>2</sup> Os documentos “Parâmetros para Educação Básica do Estado de Pernambuco – Parâmetros Curriculares” e “Parâmetros para Educação Básica do Estado de Pernambuco – Parâmetros Curriculares de Matemática para o Ensino Fundamental e Médio” são base para a formulação do SAEPE – Sistema de Avaliação Educacional de Pernambuco. As provas da Fundaj foram elaboradas utilizando-se da Teoria de Resposta ao Item.

ano letivo do 6º ano mensura o conhecimento acumulado do educando até o 5º ano do ensino fundamental, enquanto que a avaliação feita ao final desse ano apura a evolução do seu desempenho escolar e mensura o conhecimento acumulado do aluno ao final do 6º ano.

Ainda com base na Tabela 1, note-se que em média os estudantes se saíram melhor na primeira prova que na segunda, apesar disso ocorreu um acréscimo de aprendizado em 10% comparando as duas avaliações realizadas, e isso se explica pelo peso maior de variações positivas registradas. Percebe-se também que as escolas maiores, aquelas com duas turmas do 6º ano do ensino fundamental, apresentaram desempenho um pouco inferior àquelas escolas menores, com apenas uma turma.

**Tabela 1: Estatísticas Amostrais**

Variáveis	Nº obs.	Média	Desvio Padrão	Min.	Max.
Nº de escolas	120	-	-	-	-
Nº de turmas	146	-	-	-	-
Nº de alunos	3.568	-	-	-	-
Nota 1*	3.570	40,76	15,99	0	95
Nota 2*	3.570	39,81	15,61	0	95
Varição da Nota*	3.563	0,10	0,66	-1	10
Nota 1**	2.928	41,05	16,02	0	95
Nota 2**	2.928	40,34	15,57	0	95
Varição da Nota**	2.922	0,11	0,68	-1	10
Nota 1***	642	39,45	15,80	0	95
Nota 2***	642	37,40	15,57	0	95
Varição da Nota***	641	0,06	0,59	-1	6

Fonte: Elaboração própria.

\*Todas as Escolas. \*\* Apenas escolas com uma turma. \*\*\*Apenas escolas com duas turmas.

## 2 AMOSTRA DE 2013

A amostra para o campo de 2013 foi estratificada conforme as matrículas do 6º ano de escolas públicas (municipais e estaduais) do ensino fundamental do Recife e de suas respectivas notas de matemática na Prova Brasil. O Censo Escolar de 2006 (INEP/MEC, 2006), juntamente com as notas de matemática da Prova Brasil (2005), constituíram a base de dados para a construção das informações sobre as escolas públicas recifenses do ensino fundamental avaliadas nesta pesquisa. Do universo de escolas avaliadas na pesquisa foram excluídas aquelas com menos de dez participantes na série avaliada, assim como não foram consideradas as escolas rurais e aquelas destinadas ao atendimento exclusivo de educandos de comunidades indígenas. Foram ainda eliminadas as escolas com informações indisponíveis ou que apresentaram valores

iguais a zero para os insumos escolares necessários para construção dos estratos amostrais. Após aplicação desses critérios de seleção, a população alvo da pesquisa compreendeu 28.983 alunos do 6º ano matriculados em 148 escolas da rede pública de ensino localizadas nas seis Regiões Político-Administrativas (RPAs) da cidade do Recife.

A determinação dos estratos amostrais se baseou em algoritmos iterativos propostos por Lavallée e Hidioglou (1988), em que os limites dos estratos são estimados de forma a minimizar a variância do estimador empregado num plano de amostragem estratificada. Desse procedimento um total de 17 estratos foram gerados por meio da combinação das notas e matrículas. O plano amostral exigiu que as escolas eleitas fossem selecionadas com probabilidade proporcional aos estratos de matrículas e nota de matemática, por RPAs, conforme consta na população alvo. Sendo assim, uma amostra aleatória simples de alunos foi selecionada e 118 escolas foram então sorteadas para a participação na pesquisa. Adicionalmente duas escolas integraram a amostra com probabilidade 1<sup>3</sup>, totalizando um total de 120 escolas pesquisadas. A Tabela 2 resume os dados da amostra executada, com a distribuição das 120 escolas participantes por estratos e RPAs.

---

<sup>3</sup> As duas escolas públicas que integraram a amostra com probabilidade 1 foram selecionadas a título de controle, uma vez que apresentam características distintas da maior parte dos estabelecimentos públicos de ensino, como por exemplo a aplicação de provas para seleção de alunos.

**Tabela 2: Número de escolas por estrato e RPAs**

<b>Estratos</b>	<b>RPA 1</b>	<b>RPA 2</b>	<b>RPA 3</b>	<b>RPA 4</b>	<b>RPA 5</b>	<b>RPA 6</b>	<b>TOTAL</b>
<b>1.1</b>	0	0	1	1	1	1	4
<b>1.2</b>	0	0	1	0	1	2	4
<b>1.3</b>	1	2	3	0	1	1	8
<b>1.4</b>	0	0	0	1	0	0	1
<b>2.1</b>	4	3	5	1	3	0	16
<b>2.2</b>	0	1	4	2	3	3	13
<b>2.3</b>	2	1	3	1	1	2	10
<b>2.4</b>	0	2	2	0	1	2	7
<b>3.1</b>	2	0	3	1	2	1	9
<b>3.2</b>	1	1	0	2	1	3	8
<b>3.3</b>	0	0	2	3	1	1	7
<b>3.4</b>	0	0	1	0	1	4	6
<b>4.1</b>	0	1	2	0	0	1	4
<b>4.2</b>	0	1	1	5	3	1	11
<b>4.3</b>	1	0	1	0	1	1	4
<b>4.4</b>	0	1	1	0	0	4	6
<b>5</b>	0	0	1	1	0	0	2
<b>TOTAL</b>	<b>11</b>	<b>13</b>	<b>31</b>	<b>18</b>	<b>20</b>	<b>27</b>	<b>120</b>

Fonte: Elaboração própria.

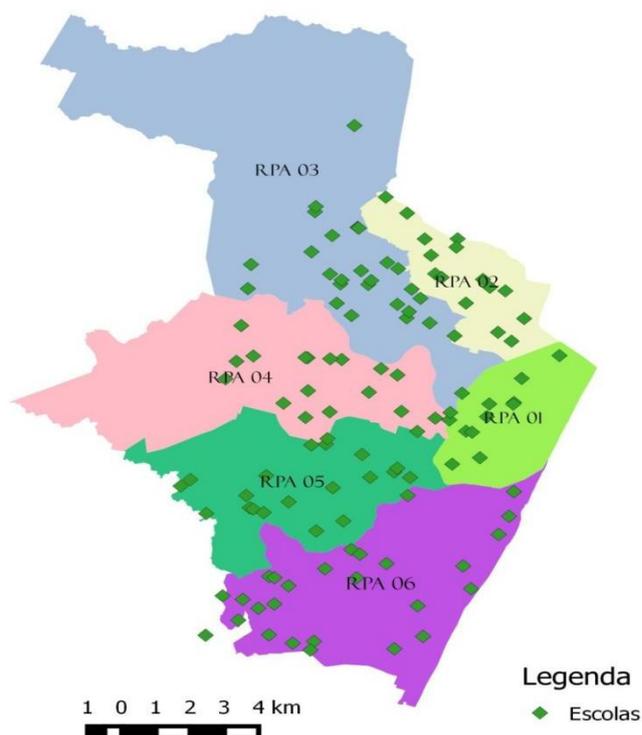
Notas: Os limites dos estratos por nota de matemática na Prova Brasil (2005) se distribuem da seguinte forma: estrato 1: 140 – 172,10; estrato 2: >172,10 – 182,66; estrato 3: >182,66 – 193,01; estrato 4: >193,01 – 213,84. O segundo nível do estrato corresponde ao número de matrículas: em que estrato 1: 15 - 135; estrato 2: >135 – 195; estrato 3: >195 – 281; estrato 4: >281 – 604. Os números dos estratos estão de acordo com a ordem crescente de nota e de matriculados, significa que o estrato 1.1 é aquele formado pelas escolas com nota até 172,10 e até 135 alunos matriculados na série de interesse. O estrato 1.2 corresponde às escolas com nota até 172,10 e 136 a 195 alunos matriculados. Assim sucessivamente, até o estrato 4.4 que consiste das escolas com nota acima de 193,01 e com mais de 281 matriculados. O estrato 5 foi considerado o estrato certo em que as escolas entraram na amostra com probabilidade 1.

Em cada uma das 120 escolas selecionadas, sorteou-se uma turma de alunos de 6º ano para participar da pesquisa e, em 26 dessas escolas, duas turmas foram sorteadas devido ao elevado número de matrículas. Como resultado, a pesquisa contou com 4.191 alunos pesquisados num total de 146 turmas. Além disso, os responsáveis por esses alunos, seus professores de matemática e diretores também foram pesquisados, conforme já mencionado.

## **2.1 Distribuição das escolas por RPAs**

As escolas investigadas distribuíram-se ao longo de toda a Cidade do Recife, como mostra a Figura 1.

**Figura 1: Localização das escolas por RPA**



Fonte: Elaboração própria.

Aproximadamente 18% das escolas estão localizadas em aglomerados subnormais, como se pode observar na Tabela 3. Conforme o IBGE (2010), essas áreas são caracterizadas por possuírem vias irregulares de circulação e pelo tamanho e forma dos lotes e/ou carência de serviços públicos essenciais (como coleta de lixo, rede de esgoto, rede de água, energia elétrica e iluminação pública) e são comumente chamadas de favela, comunidade, grotão, vila, mocambo, dentre outros.

**Tabela 3: Proporção de escolas situadas em setores subnormais**

	RPA1	RPA2	RPA3	RPA4	RPA5	RPA6	Total
<b>Subnormal</b>	9,09%	37,50%	21,43%	11,11%	0,00%	23,08%	17,65%

Fonte: Elaboração própria.

### 3 BASES DE DADOS E DICIONÁRIOS DE 2013

Um total de quatro bases de dados foram geradas pela pesquisa no ano de 2013. A primeira delas se refere aos microdados oriundos dos questionários aplicados juntos aos alunos da amostra, a qual também inclui as notas de matemática desses estudantes nas provas aplicadas pela Fundaj no início e final do ano letivo de 2013. A segunda base

apresenta os microdados dos questionários aplicados junto aos pais ou responsáveis pelos alunos investigados na amostra. A terceira base traz microdados dos questionários administrados com os professores de matemática da turma participante da amostra e a quarta base, os microdados das entrevistas feitas junto aos diretores das escolas.

Os microdados estão em formato excel e para cada base há um dicionário correspondente. Todos eles trazem a explicação da variável identificadora que deverá ser usada para juntar essas quatro diferentes bases.

#### 4 INFORMAÇÕES GERAIS SOBRE OS CAMPOS DE 2017 E 2018

Este capítulo tem como objetivo apresentar uma síntese comentada da pesquisa de Acompanhamento Longitudinal do Desempenho Escolar dos Alunos da Rede Pública de Ensino Fundamental do Recife realizada nos anos de 2017 e 2018. Nesta fase, a pesquisa contou com questionários aplicados aos alunos, aos seus pais ou responsáveis e às escolas (professores de português, matemática e diretor), bem como com a realização de testes de português e matemática, realizados no início e final de cada ano letivo.

A Tabela 4 traz o perfil geral da amostra investigada em 2017 e 2018. Nossa amostra contemplou cerca de 6.444 alunos<sup>4</sup> de 87 escolas<sup>5</sup> 167 turmas de ensino fundamental da rede pública do Recife. Também participaram da pesquisa, 206 professores de português, 200 professores de matemática e 102 diretores de escolas.

**Tabela 4: Estatísticas Amostrais**

Variáveis	Nº obs.
Nº de escolas	87
Nº de turmas	167
Nº de alunos total	6.444
Nº de responsáveis total	6.444
Nº de alunos 2017	3.274
Nº de alunos 2018	3.170
Nº de alunos novatos 2018	1.140
Nº de professores de português	206
Nº de professores de matemática	200
Nº de diretores	102

Fonte: Elaboração própria.

<sup>4</sup> Esse número é obtido com base na quantidade de alunos que participaram de todas as etapas da coleta de dados, fizeram as oito provas e cujos responsáveis preencheram seu questionário sociodemográfico e assinaram termo de consentimento permitindo a participação do aluno na pesquisa, ou seja, número de alunos com informações não missing.

<sup>5</sup> Uma escola participante deixou de oferecer ensino fundamental em 2018 saindo da amostra. O desenho original da pesquisa contemplava a participação de 88 escolas e 169 turmas.

Os alunos se submeteram a provas de português e matemática no início e final do ano letivo de 2017 e 2018. Todas as oito provas (sendo quatro de português e quatro de matemática) foram elaboradas pela equipe da Fundaj com base nos documentos Parâmetros para Educação Básica do Estado de Pernambuco, Padrões de Desempenho Estudantil em Língua Portuguesa e Matemática disponibilizados pela Secretaria Estadual de Educação e Esportes [PERNAMBUCO, (2014a), PERNAMBUCO, (2014b)]<sup>6</sup>. Cada prova foi composta por 10 questões (ou itens), com a escala da nota variando de 0 a 100, em que a nota 0 representa o caso do estudante que não acertou nenhuma questão e 100 o caso daquele que acertou todos os itens da prova. As questões possuem pontuações distintas que variam conforme o grau de dificuldade do item.

As estatísticas descritivas das notas aferidas pelo estudante nas provas aplicadas se encontram na Tabela 5. Note-se que, em geral, os estudantes se saíram melhor na prova de português que matemática. No 6º ano tiveram nota em torno de 43 e 47 (sendo nota mínima zero e nota máxima 100). No 7º ano, exceto pela primeira prova de português que aparentemente foi mais fácil que as demais, a nota média foi mais baixa em ambas as provas.

**Tabela 5: Estatísticas Amostrais**

Variáveis	Nº obs.	Média	Desvio Padrão	Min.	Max.
2017					
Nota 1 Português 6º ano	3.941	47,93	20,52	0	100
Nota 1 Matemática 6º ano	3.941	43,41	20,65	0	100
Nota 2 Português 6º ano	3.274	46,69	20,13	0	100
Nota 2 Matemática 6º ano	3.274	44,52	21,64	0	100
2018					
Nota 1 Português 7º ano	3.799	53,19	20,38	0	100
Nota 1 Matemática 7º ano	3.799	32,04	20,07	0	100
Nota 2 Português 7º ano	3.170	36,81	18,70	0	100
Nota 2 Matemática 7º ano	3.170	36,33	20,01	0	100

Fonte: elaboração própria.

## 5 AMOSTRA DE 2017 E 2018

Esta seção discute os procedimentos adotados para extrair a amostra para os campos de 2017 e 2018 da pesquisa. São apresentados os estudos realizados para o

<sup>6</sup> Os documentos Parâmetros para Educação Básica do Estado de Pernambuco, Padrões de Desempenho Estudantil em Língua Portuguesa e Matemática são base para a formulação do SAEPE – Sistema de Avaliação Educacional de Pernambuco. As provas da Fundaj foram elaboradas utilizando-se da Teoria de Resposta ao Item.

desenho da amostra, bem como a descrição de como utilizar os pesos em processos de estimação.

Tendo em vista que no campo de 2013 houve uma amostra aleatória a tendência natural para o campo de 2017/2018 seria simplesmente repetir a mesma amostra. Isto, inclusive, teria a vantagem de gerar dados longitudinais para as escolas com os três anos do campo<sup>7</sup>. Contudo, dado que a rede estadual está transferindo para a rede municipal o ensino até o nono ano, é de se supor que muitas escolas visitadas em 2013 não possuísem mais turmas de sexto e sétimo anos e novas escolas tivessem surgido. Por isto, o planejamento do campo do biênio 2017/2018 inicia pela análise da amostra de 2013 para o novo desenho da rede.

No campo realizado em 2013 o desenho da amostra foi tal que a probabilidade de uma escola ser sorteada fosse proporcional ao seu tamanho e ao seu desempenho no IDEB, considerando as Regiões Político Administrativas (RPAs) do Recife. Esta opção facilitou os processos de estimação, mas ao mesmo tempo limitou algumas das possibilidades de análise. Para poder ampliar a possibilidade de análises, adotou-se no campo de 2017/2018 um extrato certo ampliado, incluindo escolas com interesse especial da gestão. A presença destas escolas no banco de dados permite que sejam tentados modelos usando as mesmas como contrafactuais, já que sua composição de alunos, professores e ambientes de estudo são bastante diferentes do restante.

Para avaliar como se comportaria a amostra selecionada no primeiro campo no atual quadro de alunos da rede pública de ensino do Recife elaborou-se um conjunto de tabelas onde a hipótese inicial é que as escolas que foram visitadas no primeiro campo teriam a totalidade das turmas participantes do segundo campo. A Tabela 6 apresenta o total de turmas e de alunos matriculados distribuídos nas RPAs que é o critério de conglomeração utilizado no primeiro campo e que poderia se repetir no segundo. As colunas “Primeira Amostra” é fruto da simulação da repetição da amostra do campo de 2013 para o ano de 2017. As colunas “Fora Amostra” apresentam os quantitativos de alunos que estariam fora da amostra, caso fosse repetida a amostra de 2013. Nota-se que o total de alunos previstos para aplicação de provas e entrevistas é elevado.

---

<sup>7</sup> Afirma-se dados longitudinais para as escolas pois seriam possíveis análises por média obtida pelas escolas nas avaliações. Mas não seriam dados longitudinais para os alunos, pois o conjunto destes é novo para os anos de 2017/2018.

**Tabela 6 – Quantitativos de amostra supondo repetição do sorteio de escolas do primeiro campo**

RPA	Primeira Amostra		Fora Amostra		Censo 2015	
	Total Turmas	Total Alunos	Total Turmas	Total Alunos	Total Turmas	Total Alunos
1	39	1.007	17	550	56	1.557
2	53	1.797	7	219	60	2.016
3	85	2.634	19	634	104	3.268
4	60	1.780	13	339	73	2.119
5	65	1.976	16	522	81	2.498
6	102	3.258	17	549	119	3.807
<b>Total</b>	<b>404</b>	<b>12.452</b>	<b>89</b>	<b>2.813</b>	<b>493</b>	<b>15.265</b>

Fonte: elaboração própria com base no Censo Escolar 2015 e na amostra de 2013.

As frações amostrais hipotéticas estão apresentadas na Tabela 7. Notam-se frações elevadas para todas as RPAs, mas algum desequilíbrio com a RPA 1 apresentando uma fração de 12 pontos percentuais inferior à fração média. Na primeira amostra houve a criação de estratos a partir do número de alunos matriculados em cada escola e do desempenho no IDEB, além da divisão por RPA. Isto gerou um grande número de subdivisões e é o principal fator que fez com que a amostra tivesse tal dimensão.

**Tabela 7 – Frações amostrais supondo a repetição da primeira amostra**

RPA	Primeira Amostra	
	Total turmas	Total Alunos
1	69,6%	64,7%
2	88,3%	89,1%
3	81,7%	80,6%
4	82,2%	84,0%
5	80,2%	79,1%
6	85,7%	85,6%
<b>Total</b>	<b>81,9%</b>	<b>81,6%</b>

Fonte: Elaboração própria

Assim, optou-se por adotar procedimento de amostragem distinto do usado em 2013. O primeiro passo foi a elaboração de cadastro de todas as escolas públicas que possuíam alunos no 6º e 7º anos no Município do Recife, com base no Censo Educacional de 2015. Também foram identificadas quais escolas pertenceram à amostra de 2013, os endereços das escolas, o número turmas e de alunos para os anos em questão e o desempenho no IDEB de 2015. Ao tabular o endereço também foi

identificado em qual Região Político Administrativa (RPA) a escola estava localizada. A Tabela 8 apresenta a distribuição de escolas, turmas e alunos por série e por RPA.

No quesito número de turmas de sexto ano por escola a amplitude<sup>8</sup> é de 10 salas, sendo 0 o menor valor. Já para as turmas de sexto ano a amplitude é de 9 salas, sendo 0 o menor valor. O valor zero ocorre porque no Censo de 2015 existem escolas com sexto ano sem a presença do sétimo e/ou o caso contrário. Já o número de alunos variou de 0 a 307 em ambas as séries, sendo o número médio de alunos no sexto ano de 111 por escola e de 109 para o sétimo ano. As amplitudes revelam que coexistem escolas com tamanhos bem diferentes.

A última coluna da Tabela 8 apresenta a média do IDEB das escolas por RPA<sup>9</sup>. O menor valor de 2015 foi de 2,7 e o maior de 8,5, o que mostra uma grande heterogeneidade de qualidade da educação ofertada pelos estabelecimentos de ensino públicos no Recife.

Da tabela se destaca que a RPA 3 apresenta a maior participação de escolas e a RPA 6 a maior participação de salas e alunos. A RPA 3, então tem as menores médias de sala por escola enquanto que as RPAs 1 e 6 apresentam as maiores médias de sala por escola. A amplitude de alunos por turma é pequena sendo a menor média na RPA 1, de 28 alunos do sexto ano e a maior média na RPA 2, também no sexto ano. Nota-se que no quesito rendimento a distribuição espacial também parece fazer efeito.

**Tabela 8 – Distribuição de Escolas e Alunos dos 6 e 7 Anos do Ensino Público por RPA – Recife, 2015**

RPA	Número de escolas	Total de salas 6 ano	Total de alunos 6 ano	Total de salas 7 ano	Total de alunos 7 ano	Média do IDEB <sup>10</sup>
1	13	56	1.557	50	1.445	4,0
2	18	60	2.016	64	2.111	3,7
3	33	104	3.268	99	3.129	3,8
4	21	73	2.119	66	2.016	4,4
5	23	81	2.498	77	2.462	4,1
6	29	119	3.807	116	3.770	4,1
Total	137	493	15.265	472	14.933	4,0

Fonte: elaboração própria com base em dados do Censo Educacional, 2015, Inep.

<sup>8</sup> Refere-se à diferença entre ao número de turmas da maior escola com o número de turmas da menor escola.

<sup>9</sup> Média não ponderada pelo número de alunos por escola.

<sup>10</sup> Nota do Ideb de 2015 para 4ª série (5º ano).

A importância de se destacar a RPA no desenho amostral está no fato da população não ser homogeneamente distribuída no espaço, o que leva a que as escolas e vagas reflitam esta heterogeneidade. Além disto, a renda per capita e indicadores sociais, como violência, também são sabidamente heterogêneos no espaço.

Uma vez de posse destas análises e do cadastro efetuou-se simulação de qual seria o tamanho de uma Amostra Aleatória Simples sem reposição (AASs) se a variável pretendida fosse o número de alunos na escola, ou o IDEB de 2015, encontrando como resultado 131 escolas, praticamente um censo, o que comprova a necessidade de algum procedimento de estratificação. Os testes realizados utilizando-se como estratificação apenas a RPA, ou apenas o IDEB ou apenas o tamanho da escola não se mostraram promissores porque existe grande heterogeneidade mesmo dentro dos estratos, o que levou a considerar a abordagem de 2013, que tratou estas três dimensões simultaneamente como estratégia de ação.

Para o biênio de 2017/2018 optou-se por diminuir o número de estratos. Desta feita são três os fatores considerados na primeira fase do desenho amostral, a saber: tamanho da escola; desempenho prévio no IDEB; e RPA à qual pertence. A redução foi obtida considerando o número de matriculados no sexto ano e a nota do IDEB em uma única variável. Em uma primeira simulação esta redução do número de matriculados no sexto ano e do IDEB obtido em 2015 em uma única dimensão foi feita com o uso da técnica de componentes principais. O componente obtido responde por 58% da variância encontrada. Das 131 escolas presentes no Censo Escolar de 2015 apenas 123 apresentaram notas do IDEB, fazendo com que algumas escolas tenham sido excluídas da montagem do fator resumo. Uma vez realizada construção da variável acessória, a mesma foi utilizada para estabelecer grupos utilizando análise de cluster K-means. As escolas que não apresentaram notas do IDEB foram tratadas todas como um único grupo. Como já afirmado algumas escolas foram reclassificadas para um estrato cuja probabilidade de ser sorteada seria 1. Assim ao final restou uma matriz de 6 x 5 estratos<sup>11</sup> para o procedimento de sorteio.

O total de escolas sorteadas foi 88, com uma regra adicional de que nas escolas em que houvesse duas ou mais turmas seriam cobertas apenas duas turmas. Assim, a **unidade primária amostral é a escola**. Uma vez de posse das escolas sorteadas as Secretarias de Educação forneceram o número de salas para 2017, quando se procedeu

---

<sup>11</sup> Ou seja, são 30 possíveis estratos.

ao sorteio das salas nas quais seriam aplicadas as provas e entrevistas. A sala é tratada, então como um cluster. Uma vez sorteada a sala, todos os alunos da mesma são convidados a participar da pesquisa.

Como na pesquisa são aplicadas provas e questionários a alunos e questionários aos pais, uma das unidades de análise pode ser considerada aluno. Outra unidade de análise pode ser considerada a turma, uma vez que são aplicados questionários aos professores de matemática e português para as turmas sorteadas. A terceira unidade de análise é a escola, pois foram aplicados questionários para os diretores e elaboradas fichas de observação da infraestrutura escolar. Por tratar-se, então, de uma amostra complexa é preciso considerar os pesos amostrais nos procedimentos de estimação. O peso amostral nada mais é do que o inverso da probabilidade de uma escola ser sorteada ou o inverso da probabilidade de uma turma ser sorteada.

No arquivo peso escola consta o código de identificação da escola, o peso da escola, o peso dos alunos para o ano de 2017 e o peso dos alunos para o ano de 2018. Se o objetivo for fazer análises cuja unidade for a escola, deve-se utilizar o peso da escola. Para as demais análises deve-se utilizar o peso do aluno. Há diferença entre os pesos de 2017 e 2018 para os alunos porque o número de salas foi alterado, o que afeta as probabilidades. Note-se que a escolha do peso para os alunos deve ser com base na variável de interesse. Por exemplo, se está sendo realizada uma regressão cuja variável dependente é uma nota de prova de 2018, o peso a ser utilizado é o de 2018, mesmo que existam variáveis de controle coletadas em 2017.

Para aqueles que utilizarem o pacote R para fazer as estimativas o pacote aconselhável é o ‘survey’. Para aqueles que utilizarem o pacote Stata os comandos envolvem svy.

## **6 BASES DE DADOS E DICIONÁRIOS DE 2017 E 2018**

Um conjunto de seis bases de dados foi gerado para cada ano dos campos de 2017 e 2018. A primeira delas se refere aos microdados oriundos dos questionários aplicados juntos aos alunos da amostra, a qual também inclui as notas de português e matemática desses estudantes nas provas aplicadas pela Fundaj no início e final de cada ano letivo. As segunda e terceira bases apresentam os microdados de dois questionários autoadministrados pelos próprios estudantes em que são levantadas questões sobre

merenda escolar, hábitos alimentares e amizades em sala de aula. A quarta base apresenta os microdados dos questionários aplicados junto aos pais ou responsáveis pelos alunos investigados na amostra. A quinta base traz microdados dos questionários administrados com os professores de português e matemática da turma participante da amostra e a sexta base, os microdados das entrevistas feitas junto aos diretores das escolas.

Os microdados estão em formato excel e para cada base há um dicionário correspondente. Todos eles trazem a explicação da variável identificadora que deverá ser usada para juntar as diferentes bases.

## REFERÊNCIAS

LAVALLÉE, P. E HIDIROGLOU, M. A. **On the stratification of skewed populations.** Survey, 1988.

PERNAMBUCO, SEDUC. **Parâmetros para a Educação Básica do estado de Pernambuco – Parâmetros Curriculares.** Recife: SEDUC, 2012a.

\_\_\_\_\_. **Parâmetros para a Educação Básica do estado de Pernambuco – Parâmetros Curriculares de Matemática para o Ensino Fundamental e Médio.** Recife: SEDUC, 2012b.

PERNAMBUCO, SEDUC. **Parâmetros para a Educação Básica do estado de Pernambuco - Padrões de Desempenho Estudantil em Matemática.** Recife: SEDUC, 2014a.

\_\_\_\_\_. **Parâmetros para a Educação Básica do estado de Pernambuco - Padrões de Desempenho Estudantil em Português.** Recife: SEDUC, 2014b.