

**UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
CENTRO DE FILOSOFIA E CIÊNCIAS HUMANAS
FACULDADE DE EDUCAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM EDUCAÇÃO**

**ESCOLA PÚBLICA E PRIVADA:
O IMPACTO DA REDE ESCOLAR NA APRENDIZAGEM E DESIGUALDADE
EDUCACIONAL**

Daniel Lopes de Castro

Rio de Janeiro

2021

CIP - Catalogação na Publicação

L355e Lopes de Castro, Daniel
 Escola pública e privada: o impacto da rede
 escolar na aprendizagem e desigualdade educacional
 / Daniel Lopes de Castro. -- Rio de Janeiro, 2021.
 149 f.

 Orientador: Tiago Lisboa Bartholo.
 Dissertação (mestrado) - Universidade Federal do
Rio de Janeiro, Faculdade de Educação, Programa de
Pós-Graduação em Educação, 2021.

 1. Educação Infantil. 2. Desigualdade
Educacional. 3. Público e privado. 4. Modelo
multinível hierárquico. 5. Longitudinal. I. Lisboa
Bartholo, Tiago, orient. II. Título.

Daniel Lopes de Castro

**ESCOLA PÚBLICA E PRIVADA:
O IMPACTO DA REDE ESCOLAR NA APRENDIZAGEM E DESIGUALDADE
EDUCACIONAL**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Educação como requisito parcial à obtenção do título de Mestre em Educação pela Universidade Federal do Rio de Janeiro.

Orientador: Prof. Dr. Tiago Lisboa Bartholo

Rio de Janeiro

2021

Cont. Ata de Defesa de Dissertação do mestrando Daniel Lopes de Castro– 30-07-2021

Tiago Bartholo

Prof(a). Dr(a). Tiago Lisboa Bartholo (UFRJ)

Prof(a). Dr(a). Mariane Campelo Koslinski (UFRJ)

Prof(a). Dr(a). Eduardo Ribeiro da Silva (UERJ)

Daniel Lopes de Castro- candidato(a)

Solange Rosa de Araujo (Secretária do PPGE)

Em memória de Esmeralda, minha amada avó

Em memória de Jader, meu amado avô

Em memória de Romilda, minha amada tia e madrinha

Em memória de Rubens, meu amado tio

AGRADECIMENTOS

É de minha prática que os agradecimentos sejam as últimas linhas a serem escrita em meus textos. Em uma jornada, são tantas as pessoas que cruzam nossos caminhos, que é preciso chegar até o destino final para compreender aqueles que caminharam conosco, os que nos levantaram quando caímos e, mesmo, os que nos apresentaram estradas menos sinuosas. O que, certamente, ninguém esperava era que no caminho havia uma pedra chamada pandemia. É, havia uma pandemia no meio do caminho.

Por diversas vezes, a escrita dessa dissertação foi paralisada e quase não chegou ao seu destino. A ida de tantas pessoas amadas, que dão sentido às lutas diárias, que são símbolo dos sonhos sonhados, foi e continua sendo um grande desafio nessa caminhada acadêmica. Por isso, chegar até aqui não se justifica pela busca da titulação, mas pelo horizonte que dá sentido ao que vivo: a Educação.

Meu mais profundo agradecimento aos meus pais, Alécio e Regina, e aos meus irmãos, Gabriel e Carol. Suas histórias, esforço e exemplo são parte do que sou em minha busca por compreender as desigualdades e como combatê-las. Em vocês aprendi um senso profundo de empatia, a importância do compartilhar e de viver pelo que é ético e justo. Vocês são o maior exemplo que nenhuma outra pessoa foi capaz de ser em minha vida. Essa tempestade só foi possível atravessar com vocês.

Meu mais profundo agradecimento a Alice, minha amada. O companheirismo que temos traz leveza à vida, torna os momentos difíceis em momentos mais suaves e que sempre geram belas risadas num futuro não distante. Compartilhar essa estrada contigo é parte do meu orgulho e felicidade. Você é motivo de essas páginas não estarem em branco. Amo tu, meu bem.

Meu mais profundo agradecimento a Tiago Bartholo, meu orientador. Nessa jornada acadêmica, tenho grande orgulho de ser acompanhado há tempos desse profissional que tanto me inspira. A paciência, empatia, acessibilidade, engajamento, conhecimento científico e técnico, apontam para o grande orientador que é. Muito obrigado por essa parceria e tornar essa dissertação um dos maiores aprendizados de minha vida.

Meu mais profundo agradecimento à Universidade Federal do Rio de Janeiro, em especial ao Programa de Pós-graduação em Educação por propiciar essa importante formação. Agradeço à CAPES por financiar esses estudos através da bolsa PROEX, que possibilitou maior engajamento na pesquisa e estabilidade para concluir o mestrado. Agradeço ao LaPOpE/UFRJ por ser há anos meu espaço científico e de pesquisa, possibilitando trabalhar com tantos brilhantes pesquisadores que me inspiram.

Aos que estiveram presentes nesse caminho e andaram comigo nessa tempestade, meu mais profundo agradecimento.

RESUMO

A dissertação descreve padrões de desigualdade no desenvolvimento cognitivo de uma amostra de crianças matriculadas na rede pública e privada no primeiro ano da pré-escola – período inicial da escolarização obrigatória. O estudo utiliza dados com características longitudinais coletados no início e no final do ano letivo com um teste adaptativo que gera medidas para o desenvolvimento em linguagem e matemática. A pesquisa tem dois objetivos principais: a) estimar o tamanho da desigualdade observada no primeiro ano da escolarização obrigatória no desenvolvimento cognitivo de crianças matriculadas em instituições públicas e privadas e b) o impacto de ofertas pública e privada de ensino no desenvolvimento cognitivo ao longo do primeiro ano da pré-escola. A dissertação se utiliza de dados longitudinais inéditos coletados entre 2017 e 2019 em uma grande cidade do Sudeste por pesquisadores do Laboratório de Pesquisa em Oportunidades Educacionais da UFRJ. A revisão bibliográfica realizada na dissertação sugere que esse é o primeiro estudo em larga escala que mede as desigualdades de aprendizagem na pré-escola e estima o impacto das redes pública e privada no aprendizado. A amostra possui 2712 crianças matriculadas em 46 escolas da rede pública e 32 escolas privadas estratificadas pelo valor da mensalidade cobrada. Análises com modelos hierárquicos de dois níveis (aluno e escola) foram realizadas para estimar fatores associados ao desenvolvimento no início da escolarização obrigatória (onda 1 da coleta de dados) e o aprendizado ao longo do primeiro ano da pré-escola. Os resultados das análises sugerem que: i) há grande desigualdade no desenvolvimento cognitivo entre alunos matriculados na rede pública e privada de ensino no início da escolarização obrigatória. Os dados descritivos sugerem uma diferença de 9 a 15 meses de desenvolvimento médio em linguagem e matemática entre crianças da rede pública e privada.; ii) ao final do primeiro ano na escola, menos da metade das crianças matriculadas na rede pública apresenta conhecimentos em matemática e linguagem equivalentes a mediana da rede privada; iii) os modelos longitudinais sugerem que as desigualdades de aprendizagem se ampliam no primeiro ano de escolarização obrigatória quando comparamos crianças matriculadas na rede pública e privada. As limitações da amostra e outras ameaças são detalhadas no estudo e demandam cautela na interpretação dos dados. Os resultados sugerem grande estratificação do sistema escolar desde o início da escolarização obrigatória e sua ampliação ao longo do primeiro ano da pré-escola.

Palavras-chave: educação infantil; desigualdade educacional; público e privado; modelo multinível hierárquico; Longitudinal.

ABSTRACT

The Master's Thesis describes patterns of inequality in the cognitive development of a sample of children enrolled in public and private schools in the first year of pre-school - the initial period of compulsory schooling in Brazil. The study uses data with longitudinal characteristics collected at the beginning and end of the school year with an adaptive test that generates measures for development in language and mathematics. The research has two main objectives: a) to estimate the size of the inequality observed in the first year of compulsory schooling in the cognitive development of children enrolled in public and private schools and b) the impact of public and private school offerings on cognitive development throughout the first year of pre-school. The Thesis is based on unprecedented longitudinal data collected between 2017 and 2019 in a large city in the Southeast region of Brazil by researchers from the Research Laboratory on Educational Opportunities at UFRJ. The literature review suggests that this is the first large-scale study measuring learning inequalities in Pre-K in Brazil and estimating the impact of public and private networks on learning. The sample includes 2712 children enrolled in 46 public and 32 private schools stratified by tuition. Analyses with two-level (student and school) hierarchical models were conducted to estimate factors associated with development at the beginning of compulsory schooling (wave 1 of the data collection) and learning throughout the first year of Pre-K. The results suggest that: i) there is large inequality in cognitive development between students enrolled in public and private schools at the beginning of compulsory schooling. Descriptive data suggest a difference of 9 to 15 months in median development in language and mathematics between public and private children; ii) by the end of the first year in school, less than half of the children enrolled in public schools have cognitive development in mathematics and language equivalent to the median in private schools; iii) longitudinal models suggest that inequalities in learning increase in the first year of compulsory schooling when comparing children enrolled in public and private schools. Sample limitations and other threats are detailed in the study and require caution in interpreting the data. The results suggest a deep stratification of the school system from the beginning of compulsory schooling and its widening through the first Pre-K year.

Keywords: early childhood education; educational inequality; public and private; hierarchical multilevel model; Longitudinal.

LISTA DE ILUSTRAÇÕES

Figura 1: Distribuição de desempenho em Matemática na 1ª onda considerando dados observados e faltantes para as variáveis de NSE, AAC e frequência à creche.....	39
Figura 2: Distribuição de desempenho em Linguagem na 1ª onda considerando dados observados e faltantes para as variáveis de NSE, AAC e frequência à creche.....	40
Figura 3: Distribuição de AAC para escolas da amostra (observado e imputado)	44
Figura 4: Distribuição de NSE para escolas da amostra (observado e imputado)	45
Figura 5: Distribuição do desenvolvimento em Linguagem e Matemática por rede.....	56
Figura 6: Distribuição do desenvolvimento em Linguagem e Matemática por estrato.....	57

LISTA DE QUADROS

Quadro 1: Descrição das variáveis analisadas na dissertação	27
Quadro 2: Resumo das variáveis que compõem as análises multinível e <i>effect size</i>	53

LISTA DE TABELAS

Tabela 1: Composição amostral após redistribuição das escolas conveniadas.....	23
Tabela 2: Amostra do estudo: intencionada e alcança (1ª Onda)	24
Tabela 3: Amostra de crianças participantes na 1ª e 2ª ondas.....	25
Tabela 4: Taxa de resposta dos questionários de responsáveis (apenas pré-I).....	31
Tabela 5: Dados faltantes e observados.....	37
Tabela 6: Descritiva do resultado de imputação para as variáveis de AAC, NSE e frequência à creche (observados e imputados).....	42
Tabela 7: Interpretação de Effect Size em meses de aprendizagem.....	51
Tabela 8: Effect Size em Linguagem e Matemática no início do ano (1ª onda) por rede e estrato	59
Tabela 9: Estatísticas descritivas para Matemática 1ª onda.....	61
Tabela 10: Estatísticas descritivas para Linguagem 1ª onda	62
Tabela 11: Estatísticas descritivas para Matemática 2ª onda.....	63
Tabela 12: Estatísticas descritivas para linguagem 2ª onda.....	64
Tabela 13: Estatísticas descritivas dos valores de AAC, NSE e frequência à creche por estrato escolar	65
Tabela 14: análise de estratificação no início do 1º ano da pré-escola por rede - Regressão multinível para desenvolvimento de matemática 1ª onda:	70
Tabela 15: análise de estratificação no início do 1º ano da pré-escola por rede – Effect Size para desenvolvimento de matemática 1ª onda.....	71
Tabela 16: análise de estratificação no início do 1º ano da pré-escola por rede - Regressão multinível para desenvolvimento de linguagem 1ª onda	71
Tabela 17: análise de estratificação no início do 1º ano da pré-escola por rede – Effect Size para desenvolvimento de linguagem 1ª onda	72
Tabela 18: análise de estratificação no início do 1º ano da pré-escola por estrato - Regressão multinível para desenvolvimento de matemática 1ª onda.....	75
Tabela 19: análise de estratificação no início do 1º ano da pré-escola por estrato – Effect Size para desenvolvimento de matemática 1ª onda	76

Tabela 20: análise de estratificação no início do 1º ano da pré-escola por estrato - Regressão multinível para desenvolvimento de linguagem 1ª onda	76
Tabela 21: análise de estratificação no início do 1º ano da pré-escola por estrato – Effect Size para desenvolvimento de linguagem 1ª onda	77
Tabela 22: análise de aprendizagem no 1º ano da pré-escola por rede - Regressão multinível para desenvolvimento de linguagem 2ª onda	82
Tabela 23: análise de aprendizagem no 1º ano da pré-escola por rede – Effect Size para desenvolvimento de linguagem 2ª onda.....	83
Tabela 24: análise de aprendizagem no 1º ano da pré-escola por estrato - Regressão multinível para desenvolvimento de linguagem 2ª onda	87
Tabela 25: análise de aprendizagem no 1º ano da pré-escola por estrato – Effect Size para desenvolvimento de linguagem 2ª onda.....	89
Tabela 26: análise de aprendizagem no 1º ano da pré-escola por rede - Regressão multinível para desenvolvimento de matemática 2ª onda.....	92
Tabela 27: análise de aprendizagem no 1º ano da pré-escola por rede – Effect Size para desenvolvimento de matemática 2ª onda.....	93
Tabela 28: análise de aprendizagem no 1º ano da pré-escola por estrato - Regressão multinível para desenvolvimento de matemática 2ª onda.....	96
Tabela 29: análise de aprendizagem no 1º ano da pré-escola por estrato – Effect Size para desenvolvimento de matemática 2ª onda.....	97

SUMÁRIO

INTRODUÇÃO	1
1.1. OBJETIVOS	8
2. DESIGUALDADES ENTRE AS REDES PÚBLICA E PRIVADA: UM OLHAR PARA EDUCAÇÃO INFANTIL E ESTUDOS LONGITUDINAIS	9
3. METODOLOGIA.....	17
3.1. Desenho e Amostra.....	17
3.2. Instrumentos	27
3.2.2. Instrumento cognitivo: PIPS.....	31
3.2.3. Indicadores em escala Rasch: matemática, linguagem, Nível Socioeconômico (NSE) e Ambiente de Aprendizagem em Casa (AAC)	34
3.3. Imputação de dados faltantes.....	36
3.5. Análise dos dados.....	47
4. RESULTADOS	54
4.1. Descrição dos resultados para desenvolvimento em linguagem e matemática no início e final do ano letivo do pré-I (1ª e 2ª onda).....	55
4.2. Fatores associados ao desenvolvimento cognitivo e desigualdade educacional no ingresso da pré-escola.....	69
4.3. Fatores associados ao aprendizado em linguagem no primeiro ano da pré-escola	81
4.4. Fatores associados ao aprendizado em matemática no primeiro ano da pré-escola	91
CONSIDERAÇÕES FINAIS.....	99
BIBLIOGRAFIA	114
ANEXOS	122

INTRODUÇÃO

O campo de estudos sobre as oportunidades educacionais no Brasil apresenta três grandes temas de interesse. O primeiro trata do acesso aos diferentes níveis de ensino e sua relação com perfil do alunado. A grande expansão observada no ensino básico desde o início da década de 1990 (VELOSO et al., 2009) gerou mudanças graduais no foco dos estudos sobre acesso. Por um lado, um maior interesse na transição para o ensino superior (Honorato, 2020; Oliveira; Honorato, 2020) e na outra ponta estudos mais focados na cobertura da educação infantil (FMCSV, 2020). O segundo tema analisa a permanência dos estudantes em diferentes níveis de ensino, em especial no ensino básico. Os estudos sobre transição escolar e sua relação com o perfil socioeconômico das famílias são exemplos de esforços analíticos para compreender como as oportunidades educacionais são distribuídas. O terceiro tema trata da qualidade do ensino ofertado e sua relação com a equidade dos sistemas educacionais. A discussão sobre o que é qualidade na educação marca de forma importante esse tema. No Brasil, a construção de sistemas de avaliação externa com a elaboração de indicadores sobre aprendizado, fluxo e qualidade da infraestrutura nortearam parte do debate e políticas públicas focadas na melhoria da qualidade e equidade dos sistemas educacionais.

No contexto educacional brasileiro, os estudos sobre as desigualdades de oportunidades educacionais têm como um de seus objetos de interesse a forte estratificação do sistema educacional, em especial quando analisamos a oferta de ensino público e privado (aqui tratamos de escolas privadas com fins lucrativos e cobrança de mensalidades). Essa forte estratificação do sistema cria um cenário que separa, desde o início da escolarização obrigatória, crianças oriundas de famílias com perfil socioeconômico diferentes (BARTHOLO; COSTA, 2018; PICANÇO; MORAIS, 2016; COSTA; BARTHOLO, 2014; ALVES; SOARES, 2007). Essa separação dos alunos baseada em características associadas com o perfil socioeconômico das famílias cria um sistema escolar com altos níveis de segregação, em especial quando incorpora nas análises as matrículas das redes pública e privada. (BARTHOLO; COSTA, 2018)

Nas sociedades modernas, a escola é percebida como uma instituição central que deveria prover, pelo menos na teoria, oportunidades iguais para todos, gerando mobilidade social e melhoria nas condições de vida da população (CURY, 2002; RAWLS, 1972). Evidências de

diferentes países, incluindo o Brasil, sugerem que sistemas educacionais que concentram alunos com características específicas (aqui chamados de alunos em desvantagem potencial) em determinadas escolas pode influenciar a forma como eles são tratados, a qualidade do ensino e a aspiração para os níveis subsequentes de educação. Os dados sugerem que sistemas educacionais mais segregados ou com políticas que intencionalmente agrupam alunos em desvantagem potencial possuem maior associação entre os níveis de aprendizagem e o perfil socioeconômico dos alunos, ou seja, alunos com perfil socioeconômico mais desfavorável apresentam na média desempenho inferior (HAAHR et al., 2005; EUROPEAN GROUP, 2005; JENKS et al., 1972; ROSENTHAL; JACOBSON, 1968; BARTHOLO; COSTA; 2016; BARTHOLO et al., 2020). Há ainda um argumento menos presente, mas igualmente importante, que trata da ideia de justiça e equidade de um sistema educacional que separa (ou que pelo menos permite que isso seja feito) com base no perfil socioeconômico da população (CRAHAY; BAYE, 2013).

Essas questões levantam o debate sobre justiça social e da igualdade de acesso aos bens públicos, ou no caso específico da dissertação, acesso a uma educação de qualidade, em uma sociedade democrática (RAWLS, 1972). A ideia de igualdade de oportunidades no início da escolarização obrigatória é chave na dissertação. Rawls argumenta que nas sociedades modernas, os indivíduos deveriam não apenas ter o “direito às oportunidades”, mas terem de fato um “direito efetivo ou garantido”.

A dissertação estima o impacto no aprendizado da criança considerando a matrícula em uma escola da rede pública ou privada no primeiro ano da escolarização obrigatória. Afinal de contas, faz diferença estudar em uma escola da rede pública ou privada? Muitas pesquisas utilizam dados transversais de avaliações padronizadas como a Prova Brasil ou mesmo o PISA para tentar responder essa pergunta. Por vezes, jornalistas apresentam ranking de escolas considerando a nota nessas avaliações e confundem efeito da composição social dos alunos com efeito escola. A dissertação busca analisar fatores associados às desigualdades de aprendizagem no momento inicial da escolarização obrigatória utilizando dados com características longitudinais. Mais do que isso, o estudo mede, de forma inédita, o tamanho das desigualdades educacionais em uma amostra diversificada de crianças matriculadas nas redes pública e privada desde a expansão da Educação Básica obrigatória para crianças a partir dos quatro anos de idade.

As desigualdades educacionais podem se manifestar de diferentes formas e, por vezes, pesquisadores analisam a distribuição de recursos, transições escolares, aprendizado ou acesso a currículo mais acadêmico (JENKS, 1972; SILVA, 2003; RIBEIRO, 2011). Possivelmente uma das teses mais amplamente testadas no campo da Sociologia da Educação seja que o desempenho acadêmico está associado com o perfil socioeconômico da família do estudante. Dados coletados em diferentes sistemas educacionais corroboraram parcialmente os achados pioneiros de Coleman (1966), que destacam as características familiares como importante preditor dos desfechos escolares (SIRIN, 2005).

O “Relatório Coleman” foi parcialmente contestado por estudos subsequentes que indicaram que, para além do papel relevante das famílias, as escolas também podiam contribuir e explicar parte da variação observada no aprendizado dos alunos (BROOKE; SOARES, 2008). Os estudos sobre eficácia escolar ressaltavam, a partir de estudos empíricos, que escola e política educacional importavam e que era necessário observar a qualidade da oferta escolar e como essa qualidade era distribuída considerando os diferentes grupos sociais.

No Brasil há um número ainda pequeno de estudos que analisam as desigualdades educacionais no início da escolarização obrigatória (BARTHOLO et al; 2020; KOSLINSKI; BARTHOLO, 2020). Em outras palavras, há uma lacuna na bibliografia da Educação Básica, em especial a Educação Infantil, que procura descrever os padrões de desigualdades observados no início da vida escolar. Poucos estudos buscaram descrever desigualdades relacionada a qualidade da oferta e/ou o desenvolvimento dos alunos. Na breve revisão de literatura realizada nessa dissertação não encontramos nenhum estudo que descreve e compara as desigualdades de aprendizagem entre a rede pública e privada no período da pré-escola – etapa inicial da Educação Básica obrigatória. É sobre essa lacuna na literatura educacional brasileira que a dissertação de mestrado busca dar uma contribuição. De forma clara, o estudo busca descrever as desigualdades educacionais no desenvolvimento cognitivo de uma amostra de crianças matriculadas na rede pública e privada no início do primeiro ano da pré-escola (4 anos de idade) e acompanhar o aprendizado ao longo do ano letivo. A amostra é de uma metrópole da região sudeste e a característica longitudinal dos dados permitirá associar o tipo de oferta escolar com o

aprendizado no primeiro ano da escolarização obrigatória. O estudo apresenta duas questões principais:

a) qual é o tamanho da desigualdade observada no primeiro ano da escolarização obrigatória no desenvolvimento cognitivo de crianças matriculadas em instituições públicas e privadas?

b) qual é o impacto de ofertas pública e privada de ensino no desenvolvimento cognitivo ao longo do primeiro ano de escolarização obrigatória?

Diante da falta de pesquisas no contexto brasileiro que lancem mão de dados de larga escala sobre aprendizagem na Educação Infantil, pesquisadores do Laboratório de Pesquisa em Oportunidades Educacionais (LaPOpE/UFRJ) realizaram um estudo longitudinal com o objetivo principal de identificar fatores escolares e extraescolares associados ao desenvolvimento dos dois primeiros anos na pré-escola. O estudo já coletou dados em duas cidades brasileiras, uma na região sudeste e outra no nordeste, tendo a participação de 6390 crianças ao total. Os dados apresentados na dissertação são específicos do município localizado na região sudeste do Brasil.

A amostra possui 3268 crianças matriculadas na rede pública e privada para o primeiro ano da pré-escola. As informações sobre o desenvolvimento cognitivo foram obtidas com o instrumento PIPS (*Performance Indicator in Primary Schools*), que produz escalas para o desempenho em linguagem e matemática. Trata-se de um teste adaptativo com aplicação individual (clínica) com duração média de 15 minutos por criança. Os dados contextuais das famílias, escolas e turmas foram obtidos através de aplicação de questionário para os responsáveis das crianças, professores e diretores das escolas estudadas. Na presente investigação será focalizado nas informações de familiares enquanto principais variáveis explicativas. Para além das descrições e interpretações pedagógicas dos resultados, são realizadas análises de regressão hierárquica de dois níveis, sendo o primeiro o nível da criança e, no segundo nível, a escola.

Nesse sentido, a dissertação busca contribuir com o contínuo avanço nos estudos de desigualdade educacional no Brasil alcançado nas últimas décadas. Esse progresso e produção de evidências no cenário nacional foi proveniente da maior disponibilidade de dados, a partir da

consolidação dos sistemas de avaliação em larga escala e da publicação anual do Censo Escolar. Deste modo, foi possível descrever o desempenho dos alunos ao longo do tempo, assim como, mudanças na qualidade da infraestrutura, número de matrículas em diferentes etapas de ensino e perfil discente. Análises multivariadas possibilitaram estimar os fatores associados ao desempenho dos alunos em diferentes etapas do ensino e o peso da origem socioeconômica das famílias nas diferentes transições (ALVES, 2020; SOARES; DELGADO, 2016; ALVES; SOARES; XAVIER, 2016).

Esses foram avanços importantes na compreensão do fenômeno das desigualdades educacionais e ajudaram a qualificar o debate sobre como desenhar programas educacionais com objetivo de mitigar as desigualdades observadas. No entanto, é possível observar três limitações importantes na bibliografia nacional. A primeira é a falta de dados sobre as desigualdades educacionais observadas no início do processo obrigatório de escolarização que, desde 2009, inicia-se na pré-escola (4 anos de idade). Não há no Brasil, até o presente momento, avaliações em larga escala produzidas pelo governo federal ou governos locais sobre o desenvolvimento/aprendizado das crianças na pré-escola. A segunda limitação é que poucos estudos incluem em suas análises as matrículas da rede privada de ensino. A ausência de dados representativos para o universo dos alunos matriculados nas diferentes etapas do ensino básico dificulta a compreensão do fenômeno das desigualdades educacionais e possivelmente subestima o desafio que deve ser enfrentado pelos gestores públicos. Como um exemplo, Bartholo e Costa (2016) calcularam os níveis de segregação escolar para cinco capitais no Brasil e evidenciaram que a inclusão de dados censitários da rede privada ampliava de forma significativa os níveis de segregação escolar calculados. O resultado corrobora a hipótese que há grande estratificação entre as redes de ensino. A última limitação tem relação direta com a escassez de estudos longitudinais (duas ou mais medidas sobre o desenvolvimento dos alunos), que permitam análises mais robustas sobre o fenômeno das desigualdades educacionais e o peso da origem social em diferentes transições escolares.

O estudo pioneiro que olhou para o início da estratificação escolar e sua evolução no Brasil foi o Geres (FRANCO; BROOKE; ALVES, 2008), iniciado em 2005, e que acompanhou crianças matriculadas nas redes pública e privada em cinco cidades do Brasil na antiga primeira série até a quarta série – equivalente ao segundo e quintos anos do ensino fundamental atual. O

estudo permitiu identificar fatores escolares e extraescolares associados ao aprendizado dos alunos entre 7 e 10 anos, mas não foi acompanhado nos últimos 15 anos de pesquisas semelhantes, tampouco foram elaboradas pesquisas robustas de desigualdade educacional que olhassem para o novo período inicial da Educação Básica obrigatória: a Educação Infantil, mais precisamente o primeiro ano da pré-escola (4 anos de idade).

Olhar para a Educação Infantil dentro do país é ainda mais importante diante das diversas evidências internacionais sobre seus impactos positivos de curto e longo prazo na diminuição das desigualdades educacionais e na melhoria das trajetórias escolares. (SHONKOFF; PHILIPS, 2000; HECKMAN, 2006; SYLVA, 2006; DAMIANI et al., 2011). A diminuição das desigualdades educacionais, medida a partir do aprendizado, sugere que a oferta da Educação Infantil beneficia mais as crianças em situação de vulnerabilidade social e deve ser compreendida como importante medida redutora de desigualdades sociais. No entanto, os mesmos estudos sugerem que o efeito moderador dessa etapa ocorre quando a oferta é de qualidade.

Investigar a desigualdade na pré-escola não é, portanto, apenas auxiliar no preenchimento de uma lacuna na bibliografia nacional, mas é, também, lançar luz sobre a inexplorada qualidade de oferta dessa fase da Educação Infantil no país. Sua introdução enquanto parte da Educação Básica obrigatória em 2009, através da Ementa Constitucional Nº 59, fez com que a pré-escola se tornasse a etapa escolar que mais se expandisse na última década no Brasil, ampliando em mais de quinhentas mil vagas entre 2010 e 2019 segundo o Censo Escolar (INEP, 2021). Desde sua obrigatoriedade, a pré-escola vem sendo ampliada e caminha para a universalização ao longo da década de 2020. Essa expansão ocorrida até o momento não foi realizada apenas pelo ensino público, mas, em parte, impulsionada pela rede privada, representando em 2019 23,14% do total matrículas de pré-escola. Desenvolver análises que levem em consideração essa estratificação é importante tanto pelo tamanho da representatividade das instituições privadas quanto por haver indícios de que essa separação público-privada estabeleça oferta de programas de ensino paralelos, propondo-se a destinos e visões pedagógicas diferentes na Educação Infantil e sendo vertente para impactos díspares na aprendizagem ao longo da vida (CAMPOS, FÜLLGRAF; WIGGERS, 2006; KLEIN, 2006; RIBEIRO, 2011). A expansão da oferta de vagas, portanto, não é garantia de qualidade e tampouco de diminuição das desigualdades educacionais.

É relevante destacar também que período entre os 0 e 5 anos de idade, fase da Educação Infantil, recebe ainda mais atenção dada sua importância diante da janela de oportunidade para do desenvolvimento cerebral conhecido como Período Sensível, que são momentos da vida com maior capacidade de modificação e maleabilidade do cérebro em resposta às experiências ambientais (SINGER, 1995; KOLB et al., 2013). Com isso, fica potencializado tanto a aprendizagem e desenvolvimento de habilidades quanto sua sensibilidade para efeitos perversos promovidos por situação de vulnerabilidade/risco, expressos em uma maior probabilidade de desenvolvimento de doenças fisiológicas e psicológicas, como problemas cardiovasculares e transtorno de ansiedade e depressão (EVAN; KIM, 2007; FRANKLIN; SAAB; MANSUY, 2012).

A dissertação busca, portanto, contribuir com o preenchimento de lacunas na bibliografia nacional e com o debate da qualidade da Educação Infantil Brasil. O texto divide-se em cinco capítulos. O primeiro capítulo apresenta a questão, justificativa e os objetivos gerais e específicos deste trabalho. O capítulo dois apresenta uma breve revisão bibliográfica sobre o que já sabemos sobre as desigualdades educacionais na pré-escola e, em especial, estudos que tenham incorporado dados das redes pública e privada no cenário nacional. O terceiro capítulo descreve o desenho longitudinal proposto, instrumentos, amostra e modelagem utilizada para estimar os fatores associados ao desenvolvimento infantil no primeiro ano na pré-escola. O quarto capítulo apresenta as análises. Inicialmente são explorados os dados descritivos que permitem caracterizar a amostra. Em seguida são apresentados resultados do modelo multivariado que estima os fatores associados ao desenvolvimento cognitivo das crianças no ingresso da pré-escola. O modelo final da dissertação estima os fatores associados ao aprendizado das crianças ao longo do primeiro ano na pré-escola. O último capítulo apresenta as principais conclusões do estudo e debate as implicações dos resultados para a política educacional na primeira infância, além de expor propostas para trabalhos futuros com os dados elaborados nesta dissertação.

1.1. OBJETIVOS

A dissertação apresenta dois objetivos principais: a) estimar o efeito da oferta escolar (rede pública ou privada) no aprendizado em linguagem e matemática no primeiro ano da pré-escola e; b) medir as desigualdades educacionais no início e no final do primeiro ano da pré-escola, comparando o desenvolvimento cognitivo dos alunos matriculados nas redes pública e privada. Além dos dois objetivos principais supracitados, outros quatro objetivos específicos são apresentados em seguida:

1. Descrever o perfil socioeconômico das famílias considerando os diferentes tipos de oferta de pré-escola;
2. Analisar os dados faltantes e realizar a imputação do indicador de nível socioeconômico e ambiente de aprendizagem em casa para as famílias que não responderam ao questionário de responsáveis;
3. Identificar fatores extraescolares associados ao desenvolvimento cognitivo no início da escolarização obrigatória;
4. Descrever as desigualdades no desenvolvimento cognitivo mensuradas no início e no final do primeiro ano da pré-escolar entre as crianças matriculadas em escolas da rede pública e privada;

2. DESIGUALDADES ENTRE AS REDES PÚBLICA E PRIVADA: UM OLHAR PARA EDUCAÇÃO INFANTIL E ESTUDOS LONGITUDINAIS

Há poucos estudos na literatura nacional que analisam o fenômeno das desigualdades educacionais entre as redes de ensino pública e privada, principalmente quando o recorte é a pré-escola – anos iniciais da Educação Básica obrigatória. Uma vez que não há uma política nacional de acompanhamento de aprendizagem na Educação Infantil, como há para o Ensino Fundamental e Médio através do Saeb, a maior parte dos estudos para essa faixa etária se utilizam apenas de dados observacionais ou não se balizam por desenhos e modelagem estatística robustas. Em suma, praticamente todos os estudos brasileiros publicados até o momento sobre a qualidade na pré-escola e que observam as diferenças nas ofertas públicas e privadas, apresentam apenas análises descritivas e correlacionas, elementos insuficientes para apontar impacto das redes nas desigualdades educacionais e aprendizagem.

Kramer e Kappel (2000) utilizaram dados da PNAD para comparar condições materiais de creches e pré-escolas públicas e privadas. Os resultados indicam que as instituições privadas apresentam condições estruturais melhores, medidas através de a quantidade de insumos como livros, televisões e computadores. Simões e Lima (2016) vão na mesma direção e, ao analisar dados de provenientes do Censo Escolar da Educação Básica de 2012, evidenciam que as instituições privadas detêm uma proporção de quase o dobro de oferta de parque infantil e banheiro adequado quando comparado com as instituições públicas. Apesar de não discriminarem as características dessa oferta, as autoras apontam esses resultados como um indício de desigualdade na qualidade de oferta da Educação Infantil.

Indo além dos dados descritivos, Campos et al. (2018) se utilizam de instrumento específico para medir qualidade do ambiente ofertado na Educação Infantil, não chegando ao mesmo resultado do que os estudos anteriores. Ao investigar 147 instituições públicas, privadas e conveniadas de Educação Infantil em seis capitais brasileiras, as autoras reforçam que há grande heterogeneidade dentro de cada tipo de oferta, o que dificulta conclusões generalistas: “devido à grande heterogeneidade de condições observadas em cada um desses segmentos [público e privado], qualquer generalização, com base na amostra definida para esta pesquisa, se torna arriscada” (CAMPOS et al., 2011, p.38).

No que tange medidas cognitivas para indicação de qualidade de oferta, algumas pesquisas coletaram dados sobre o desenvolvimento das crianças durante a Educação Infantil com amostras que permitem comparações preliminares entre as redes pública e privada. Por exemplo, Pedras et al. (2006) analisaram o desenvolvimento do vocabulário em uma amostra com 86 crianças matriculadas na pré-escola (idades entre 4 e 5 anos). A amostra por conveniência apresenta limitações quando pensamos a sua representatividade e os resultados devem ser interpretados com cautela. Em linhas gerais, os autores destacam que as crianças matriculadas na rede privada apresentaram resultados médios superiores comparado com as crianças da rede pública. No entanto, os autores reforçam que, apesar de os alunos matriculados na rede pública apresentarem resultados na média menores, os valores ainda sugerem que estão no nível considerado “adequado” para os parâmetros da avaliação utilizada.

O estudo realizado por Amorin et al. (2019) com dados com características longitudinais apresenta achados que em grande medida corroboram o estudo de Pedras et al. (2006). A pesquisa apresenta uma amostra de 389 crianças de 5 anos que vivem na região nordeste do Brasil. Através de análise descritiva, sem controle para perfil socioeconômico das famílias, sugere que as crianças matriculadas na rede privada apresentam na média resultados superiores para leitura e escrita no início do segundo ano da pré-escola. O mesmo estudo ao medir as mesmas crianças no final do segundo ano da pré-escola identificou que a distância no desenvolvimento das crianças entre as redes pública e privada havia aumentado, sugerindo um aprendizado mais acelerado na rede privada de ensino. Cabe reforçar mais uma vez, que os estudos citados apresentam amostras por conveniência, não representativas, e, portanto, os resultados não podem ser extrapolados para a população de crianças matriculadas nas redes pública e privada.

Os resultados apresentam um padrão já identificado em outros estudos internacionais e nacionais que sugere uma relação entre o perfil socioeconômico das famílias e o desenvolvimento discente (BOUDON, 1981; COLEMAN et al., 1966; CRAHAY; BAYE, 2013; SYLVA, 2006). A rede privada de ensino no Brasil tem perfil distinto do que é observado por vezes em outros países, uma vez que a maior parte das escolas privadas é com fins lucrativos e cobra mensalidades das famílias. Segundo dados de 2018 da UIS/UNESCO, a porcentagem de matrículas em escolas privadas para *primary education* no Brasil é de 18.2%, próximo dos 20,5%

para a América Latina, mas distante dos países de renda alta que, em média, detém 13% apenas das vagas na rede privada, enquanto a grande maioria do atendimento é público. Essa fase é o que corresponde ao principal início da Educação Básica obrigatória no mundo e, aqui é representada pelo Ensino Fundamental I (6 a 11 anos).

Ainda que não se trate da Educação Infantil, Engel de Abreu et al (2015) publicaram relatório produzido por acadêmicos da Universidade de Luxemburgo, USP, UFBA e UNIFESP, mas não revisado por pares, um estudo em que investiga a influência de características socioeconômicas no desempenho cognitivo em matemática, leitura e linguagem. Para isso, compararam resultados de avaliação cognitiva em uma amostra intencional de 355 crianças de escolas públicas e privadas de São Paulo e da Bahia em idade escolar do 1º e 2º anos do Ensino Fundamental. Dada às características do estudo, que visava “determinar relação cognitiva e nível socioeconômico”, crianças que tiveram problemas no processo de natalidade, como mães que usavam drogas, nascimento prematuro, atraso no desenvolvimento entre outros elementos ou, ainda, com bolsa de estudo ou que eram bilíngues, foram excluídas da amostra. A própria amostra intencional pouco heterogênea, além da exclusão de perfis, compromete a validade externa do estudo, isto é, a capacidade de generalização dos resultados para outros contextos (CANO, 2004).

Utilizando de modelagem de equações estruturais, os autores estimaram o desenvolvimento cognitivo e funções executivas e relatam que as características econômicas da família explicam 30% da variação nos modelos que estimam função executiva e 50% da variação no desenvolvimento da linguagem. Para além, os autores indicam que há evidências de que estudar em escola privada causa uma grande vantagem no desempenho. Apontam, ainda, que não identificaram diferenças entre escolas de uma mesma rede, ainda que houvesse variação socioeconômica interna à rede privada e pública. Ou seja, não foi observada variação de resultados entre escolas de uma mesma rede, indicando não existir estratificação interna nas redes quanto aos resultados.

Três são os principais elementos que fragilizam os achados dos estudos supracitadas, tanto no que tange à validade interna dos estudos, que indica a qualidade da evidência causal, quanto à validade externa, que apresenta a capacidade factível de generalização dos resultados para além da amostra estudada, isto é, o universo para qual se busca generalizar os resultados

(CANO, 2004): A) falta de dados longitudinais sobre desenvolvimento cognitivo da coorte estudada, B) amostras não representativas da população e C) a falta de tratamento apropriado nos dados através de modelos controlados e hierárquicos que auxiliem diferenciar efeito-escola e efeito de composição discente.

O primeiro elemento se expressa no desenho dos estudos analisados que são, principalmente, transversais/seccionais. Analisar aprendizagem depende de um desenho de estudo conhecido como “longitudinal”, em que se compara o desempenho atual de um indivíduo com seu desempenho anterior, interpretando o aprendizado como a diferença entre esses dois momentos. A existência de uma única medida de desempenho cognitivo (desenhos transversais) impede, portanto, qualquer tentativa de analisar impacto do tipo de rede sobre a aprendizagem, uma vez que a medida de desempenho seria um agregado da aprendizagem ao longo do tempo, enquanto as medidas escolares, como insumos e processos, são coletadas em um único momento, impossibilitando identificar e realizar inferências confiáveis sobre o efeito das escolas (FRANCO; BROOKE; ALVES, 2008; GOLDSTEIN, 1997). Nesse sentido, em contraposição aos desenhos transversais de pesquisa, a aprendizagem só é captada por desenhos longitudinais através de medidas consecutivas do desempenho de uma mesma coorte de alunos, permitindo analisar o valor agregado pelas instituições e como características individuais e contextuais influenciam nos desempenhos dos estudantes ao longo do tempo (BONAMINO; OLIVEIRA, 2013; LEE, 2010; REARDON; RAUDENBUSH, 2009; SOARES et al., 2017). Brooke et al. (2014) sintetizam a fragilidade dos desenhos transversais frente aos longitudinais para estudos de efeito-escola:

Pela falta de conexão com as origens da aprendizagem, provocada pelas incertezas sobre o ponto de partida dos alunos e das contribuições específicas do ambiente de aprendizagem, os pesquisadores só se permitem falar de “fatores associados” e, raramente, comprometem-se a indicar causas e efeitos. Brooke et al. (2014), P. 79.

O mesmo já era sintetizado por Goldstein em 1997, como pode ser visto abaixo em tradução livre:

Todo o desempenho anterior, história social e circunstâncias de um aluno podem ser relevantes, e as medições feitas em um único momento são inadequadas. Do ponto de vista da coleta de dados, isso levanta sérios problemas práticos, uma vez que requer que as medições sejam feitas em crianças por longos períodos de tempo. Goldstein (1997). Pg. 375.

E podemos complementar através de Soares et al. (2017):

Ao se avaliar a qualidade da escola apenas por medidas ao final de uma etapa escolar, pode-se estar desconsiderando substancialmente o esforço realizado pela escola vis-à-vis as variáveis de contexto que influem nos resultados finais. Soares et. al (2017), Pg. 82-83.

Sendo assim, desenhos longitudinais são indispensáveis para avaliar o impacto de fatores intra e extraescolares no aprendizado dos estudantes. Ou seja, é necessário lançar mão de estudos que acompanham os indivíduos ao longo do tempo para que seja possível avaliar impacto tanto de insumos e perfis familiares quanto de redes escolares (pública ou privada) sobre a aprendizagem dos alunos.

O segundo ponto, elemento principal da validade externa, é quanto à representatividade das amostras. Na grande maioria dos estudos observados não eram representativos das redes estudadas ou sequer buscavam estabelecer amostras com perfis heterogêneos das redes, mesmo que sem garantia de representatividade. Desse modo, qualquer tentativa de generalização dos resultados se torna frágil, uma vez que o estudo realizado pode estar apenas captando grupos específicos, não representando, assim, o comportamento da população a que se pretende inferenciar as evidências (CANO, 2004).

Por último, trata-se da construção de modelos matemáticos capazes de diferenciar A) efeito composicional do perfil de alunos matriculados na escola do B) efeito dos processos e insumos escolares sobre a aprendizagem (efeito-escola). Raudenbush e Willms (1995) defendem que, para realizar essa discriminação, é necessário estruturar ajustes estatísticos através de medidas individuais e contextuais dos estudantes e de insumos e processos escolares. São exemplos de variáveis que deveriam compor os modelos: o sexo do estudante, nível socioeconômico e diversas medidas cognitivas comparativas de cada estudante ao longo do

tempo, assim como medidas de controle sobre características intraescolares, como práticas em sala de aula, tamanho de turma, formação dos professores e liderança da coordenação escolar. Aliado a essa composição de variáveis, Goldstein (1997) aponta para a necessidade do uso de modelos estatísticos multiníveis/hierárquicos, no intuito de isolar de forma mais robusta a influência dos diferentes agrupamentos que os estudantes estão inseridos e que podem influenciar na aprendizagem, como, por exemplo, turma, escola e município.

Os poucos artigos encontrados que apresentam o impacto da rede pública e privada na aprendizagem na pré-escola são, em sua maioria, de metodologia descritiva ou correlacional. Aqueles que se lançam em modelagem inferencial detêm poucas variáveis de controle, não se utilizam de modelagem multinível, têm amostras não representativas e com desenho transversal, fragilizando as evidências.

A falta de pesquisas com desenhos longitudinais não é específica da Educação Infantil, mas do campo educacional como um todo no Brasil. Em uma investigação breve sobre pesquisas longitudinais na educação na produção nacional, Giraldi e Sigolo (2016) identificaram em uma busca rápida por palavra-chave de estudos longitudinais 16 artigos entre 1999 e 2014 que lidam com o campo da educação. As autoras indicaram que grande parte deles era desenvolvido por pesquisadores de forma independente, sem grande escala e por um curto período temporal, com exceção de um: o estudo longitudinal da geração escolar 2005, conhecido, também, como Projeto Geres.

Para o cenário nacional, o Geres é a única grande fonte de estudo longitudinal produzida até início dos anos 2010 e que possibilita identificar diversos fatores que geram efeito sobre a aprendizagem dos alunos, como o nível socioeconômico das famílias, rede escolar (pública e privada), insumos e processos escolares, assim como raça e sexo dos estudantes. Executado em cinco cidades de grande porte do país, com uma amostra representativa para escolas públicas e privadas dos municípios, o projeto acompanhou mais de 21.000 crianças que estavam matriculadas em turmas de 2º ano do Ensino Fundamental em 2005, coletando informações cognitivas de Língua Portuguesa e Matemática dessa mesma coorte até 2008. Ademais, foram levantados dados contextuais junto das famílias das crianças e de insumos e processos com as escolas (FRANCO; BROOKE; ALVES, 2008).

Ainda que não olhe para a Educação Infantil, é preciso ressaltar que o Projeto Geres foi realizado quando o Ensino Fundamental era o início da Educação Básica obrigatória no país. Desse modo, os resultados obtidos pelos autores que se debruçaram sobre esses dados podem apontar direções que auxiliem compreender os resultados da presente dissertação, que também olha para o início da Educação Básica obrigatória, mas que hoje se expressa no primeiro ano da pré-escola.

Ao analisar os dados dos quatro anos de execução da pesquisa, entre 2005 e 2008, Brooke et al. (2014) buscaram identificar diferentes tipos de padrões no comportamento da aprendizagem de Língua Portuguesa e de Matemática, levando em consideração o quanto fatores de rede (pública e privada), raça/cor (preto e não preto) e nível socioeconômico influenciam os desempenhos e as aprendizagens. Os autores identificaram que o nível socioeconômico do aluno, o nível socioeconômico médio da escola e estudar em rede privada não ampliam a desigualdade na aprendizagem em Língua Portuguesa, o que significa que as desigualdades observadas no início da escolarização se mantêm constante ao longo dos anos quando observadas por essas variáveis. Nesse sentido, o que se está evidenciando é que, na média, estudar na rede privada não indica que há uma melhora na aprendizagem em Língua Portuguesa quando comparada com a rede pública para os anos iniciais do Ensino Fundamental. Por outro lado, ser do gênero masculino ou ser preto se apresentam como fatores que influenciam nos resultados em Língua Portuguesa, indicando que os estudantes com essas características não apenas iniciaram a pesquisa com resultados significativamente menores em comparação com meninas ou não pretos como, também, esses grupos tendem a sofrer com uma menor aprendizagem nessa dimensão do conhecimento, ampliando a desigualdade educacional ao longo dos anos quando comparados aos seus pares.

Para Matemática, o comportamento da influência das variáveis é diferente. As análises indicam que ter nível socioeconômico alto ou estar estudando na rede privada ou, ainda, estar classificado como não preto são características que geram, na média, vantagem na aprendizagem de matemática ao longo dos anos, ampliando a desigualdade observada no início da pesquisa.

Achados na mesma direção com os dados provenientes do Projeto Geres também são apresentados por Procópio, Silva e Chein (2015). Ainda que não coloquem nas análises controles de rede pública e privada, apresentam que o nível socioeconômico dos alunos se expressa

enquanto o fator de maior contribuição para a ampliação das desigualdades ao longo dos anos, principalmente para conhecimentos de matemática. Já, no que tange raça e gênero, os autores enfatizam que há pouca ou nenhuma influência de ser masculino ou preto na desigualdade de início, contudo, ao passar dos anos, esses começam a ser identificados como fatores de desvantagem potencial de aprendizagem.

Os resultados descritos reforçam a necessidade de novos estudos, em especial com desenho longitudinal e amostras representativas, que permitam avançar na compreensão do fenômeno das desigualdades educacionais no início da escolarização obrigatória. Um ponto chave é introduzir na seleção da amostra e, conseqüentemente nas análises, o máximo de heterogeneidade que existe na oferta da rede privada. Quando falamos da rede privada, deve-se considerar que há grandes diferenças na qualidade da infraestrutura e ambiente de aprendizagem, assim como, no perfil discente. Isso pode ser constatado considerando a grande variedade no valor das mensalidades cobradas. A mensalidade pode ser interpretada como uma proxy do perfil socioeconômico das famílias que matriculam seus filhos. Novas pesquisas, com amostras representativas para a rede privada de ensino devem buscar identificar se há clara relação entre o perfil discente, qualidade da infraestrutura e qualidade do ambiente de aprendizagem nas escolas.

Um outro aspecto importante que merece aprofundamento são as propostas pedagógicas ofertadas pelas diferentes redes de ensino. Campos, Füllgraf e Wiggers (2006) compararam diversas propostas escolares da Educação Infantil de instituições públicas e privadas e constataram que as escolas privadas se destacavam por terem, proporcionalmente, mais propostas e documentos pedagógicos do que as instituições públicas. Contudo, indicam que “essas propostas se caracterizavam por certa falta de clareza e pela excessiva ênfase nas áreas de conhecimento e no desenvolvimento cognitivo (...)” (Pág. 23). Os achados sugerem que, para além de aspectos relacionados à infraestrutura e perfil discente, haja abordagens pedagógicas diferentes entre as redes públicas e privadas que podem auxiliar explicar resultados cognitivos diferentes.

3. METODOLOGIA

O presente capítulo apresenta o desenho, os instrumentos, a amostra, os critérios para a imputação dos dados e a modelagem estatística utilizada no estudo. O primeiro subcapítulo busca descrever o desenho da pesquisa e a seleção da amostra, considerando as especificidades e limites para a comparação dos dados para a rede pública e privada. Em seguida são apresentados os instrumentos utilizados na coleta de dados e as principais estratégias para garantir a qualidade das medidas. O terceiro subcapítulo lança luz na construção de variáveis latentes através de Rasch, técnica que pode ser compreendida como um subconjunto da Teoria de Resposta ao Item (TRI). São detalhados a elaboração das variáveis de desempenho em matemática e linguagem, principais variáveis dependentes e independentes do estudo, e os indicadores de Nível Socioeconômico (NSE) e Ambiente de Aprendizagem em Casa (AAC), utilizados como variáveis de controle. O quarto subcapítulo trata dos dados faltantes (*missings*), detalhando-os para cada estrato de escola (pública, privada baixa, média e alta mensalidade) e explicitando o processo de múltipla imputação utilizado para lidar com tal problemática. Por fim, são apresentadas as técnicas de análise dos dados, explicitando a construção do modelo de regressão hierárquica de dois níveis e a transformação dos coeficientes em *effect-sizes* no intuito de uma facilitada interpretação dos resultados.

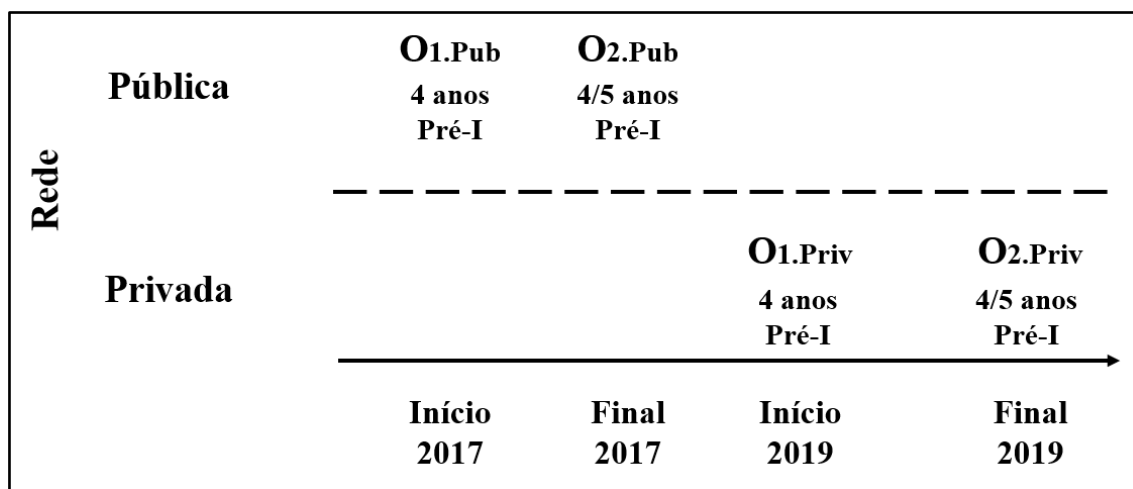
3.1. Desenho e Amostra

A dissertação faz um recorte do estudo longitudinal desenvolvido pelo Laboratório de Pesquisa em Oportunidades Educacionais da Universidade Federal do Rio de Janeiro (LaPOpE/UFRJ). O projeto tem como principais objetivos identificar fatores escolares e extraescolares associados com o desenvolvimento das crianças matriculadas na pré-escola (BARTHOLO et al., 2020; KOSLINSK; BARTHOLO; 2020; AGUIAR; BARTHOLO, 2020; SANTOS, 2020).

A presente dissertação utiliza dados com características longitudinais com duas medidas sobre o desenvolvimento cognitivo dos alunos (matemática e linguagem), uma no início e outra ao final ano letivo do 1º ano da pré-escola, em outras palavras, no primeiro ano de escolarização obrigatória. São compostos dois grupos de comparação: escolas municipais e escolas privadas, sendo que a amostra das escolas privadas foi selecionada considerando três estratos de acordo

com a mensalidade cobrada: baixa, média e alta mensalidade. Abaixo, na Figura 1, é apresentado o desenho utilizado para análise da presente dissertação. Ressalta-se que o estudo completo executado pelo LaPOpE/UFRJ envolvia, também, estudantes do 2º ano da pré-escola, contudo a presente dissertação apenas focaliza no Pré-I.

Figura 1 – Desenho de pesquisa da Dissertação - recorte do projeto LaPOpE/UFRJ



Fonte: Produção própria do autor.

O desenho do estudo é inovador para o cenário nacional ao permitir descrever o que as crianças sabem e são capazes de fazer quando ingressam na pré-escola e como elas se desenvolvem ao longo do primeiro ano de escolarização obrigatória. Não há para essa etapa de ensino avaliações padronizadas aplicadas pelo Inep ou mesmo governos estaduais ou municipais. A maior parte dos estudos que investiga a qualidade da oferta na Educação Infantil coleta medidas sobre a infraestrutura ou qualidade do ambiente/ sala de aula. Os dados apresentados irão permitir a descrição detalhada do desenvolvimento das crianças e associar esse desenvolvimento com características escolares e extraescolares, comparando diferentes tipos de oferta na Educação Infantil.

A importância da medida inicial sobre o desenvolvimento infantil no início do processo obrigatório de escolarização é chave para medir de forma adequada o efeito da escola sobre a aprendizagem e desigualdade de desempenho. Assim como já debatido na revisão de literatura, o

uso de desenho longitudinal com uma medida inicial - denominada linha de base - permite realizar análises de valor agregado, o que o torna os estudos longitudinais metodologicamente mais robustos do que os desenhos transversais para inferência causal do efeito-escola (BONAMINO; OLIVEIRA, 2013; LEE, 2010; SOARES et al., 2017).

Na dissertação utilizaremos duas denominações para diferenciar as escolas: redes e estratos. Todas as vezes que for citado *rede*, está se referindo sobre a separação em instituições públicas e privadas. Quando for citado *estrato*, refere-se a uma decomposição da rede, sendo que as instituições públicas permanecerão denominadas de pública, mas as instituições privadas serão divididas em três grupos, sendo a) baixa mensalidade (até R\$400), b) média mensalidade (R\$401 – R\$1000) e c) alta mensalidade (acima de R\$1000). Para rede, portanto, estamos nos referindo à comparação pública e privada como um agregado e, para estrato, a comparação pública (agregada) e instituições privadas de baixa, média e alta mensalidade. Essa diferenciação permitirá verificar o efeito médio das instituições privadas agregadas e sua estratificação interna em diferentes mensalidades, buscando compreender uma desigualdade da própria rede privada. Ressalta-se que, nos modelos de regressão a serem apresentados nos resultados, a categoria de base de comparação será sempre a rede pública de ensino. Essa escolha pode ser justificada considerando que a rede pública apresenta o maior número de matrículas e instituições na amostra, além de apresentar a maior proporção nacional de matrículas para a faixa etária estudada e, portanto, principal foco de interesse dos estudos educacionais.

Os dados foram coletados em dois anos distintos, sendo 2017 as escolas públicas e, em 2019, as escolas privadas, composta de uma amostra de instituições privadas conveniadas com poder público e as privadas sem convênio e que cobram mensalidade (baixa, média e alta mensalidade). A coleta em dois anos distintos é uma limitação do desenho do estudo e que introduz uma possível ameaça à validade interna conhecida como ameaça histórica. É razoável assumir que entre 2017 e 2019 mudanças ocorreram nas escolas que podem estar associadas com a principal variável dependente do modelo. Acreditamos que a aprovação e gradual implementação da BNCC para a Educação Infantil pode ser um evento que altera, mesmo que de forma parcial, o currículo e práticas docentes na pré-escola. Se isso for verdade, o risco é, ao comparar diferentes redes com intervalo médio de 2 anos, atribuir uma diferença entre as redes que na verdade tem relação com outras mudanças que ocorrem no mesmo intervalo de tempo e

que podem explicar, mesmo que parcialmente, os resultados descritos. Iremos retomar esse argumento na discussão dos resultados apontando as limitações/ameaças do estudo e cuidados com as interpretações.

Originalmente a dissertação iria incluir nas análises uma amostra de escolas conveniadas que também foi selecionada para participar da coleta de dados em 2019. Contudo, para as análises apresentadas nessa dissertação, foi decidido realocar algumas instituições originalmente categorizadas como conveniada para a categoria de escolas privadas que cobram mensalidade. Isso se deve porque foi verificado que algumas instituições conveniadas cobravam mensalidades para os alunos da pré-escola e atendiam alunos cujo as famílias tinham alta escolaridade e renda, comportando-se de forma mais parecida com as outras escolas privadas do que com instituições que recebem subsídios públicos para atender - pressupõe-se – perfis discentes majoritariamente de baixo nível socioeconômico. As conveniadas que detêm o comportamento de atender baixos níveis socioeconômicos e que praticamente não cobram mensalidade foram excluídas da análise nesse estudo. Com isso, a dissertação irá comparar escolas públicas, ou seja, que não cobram mensalidade, com instituições privadas que cobram diferentes valores de mensalidade.

Defendo que para estudar as conveniadas como um todo precisaria compreender sua natureza jurídica no município pesquisado, ou seja, compreender se os alunos matriculados em instituições conveniadas deveriam estar ou não isentos de pagar mensalidade ou, ainda, se haveria algum valor limite, por exemplo. Isso é importante pois a cobrança de mensalidade deve ser vista como um fator que estratifica as escolas de acordo com o perfil socioeconômico das famílias. Contudo, essa investigação não foi possível de ser realizado a tempo para a finalização do presente estudo. Dado o comportamento heterogêneo dessa categoria, portanto, foi preferível separar algumas escolas conveniadas com características típicas das escolas privadas sem convênio da amostra e, assim, aumentar o tamanho amostral das instituições privadas que cobram mensalidade. Esta é uma das investigações futuras a serem realizadas e que pode vir a contribuir para melhorias nas evidências produzidas nesta dissertação.

Apresenta-se em seguida o como as amostras de escolas públicas, conveniadas e privadas com mensalidade foram escolhidas, assim como a redistribuição de algumas escolas conveniadas em privadas com mensalidade. Deve-se ter em mente, porém, que todas as análises que serão

apresentadas ao longo da dissertação dizem respeito apenas a escolas públicas e privadas com cobrança de mensalidade, incluindo as reclassificadas.

As amostras das redes pública, conveniada e privada apresentam características distintas. A rede pública detém uma amostra por conglomerado, estratificada por diferentes regiões administrativas da cidade e é representativa para a rede pública de ensino do município investigado. A pesquisa foi autorizada pela Secretaria Municipal de Educação e, a partir desse ponto, a equipe de pesquisadores da UFRJ entrou em contato com as escolas sorteadas para participar do estudo para explicar a pesquisa e obter as autorizações individuais dos respondentes para participação no estudo. Ao longo de todo o processo que levou algumas semanas de preparação, a equipe de coordenação de campo contactava as instituições públicas sorteadas e buscava esclarecer qualquer dúvida sobre o estudo, negociando melhores estratégias de entrada na escola com o mínimo de impacto na rotina. Esse protocolo gerou grande aceitação do estudo por parte das escolas participantes e atores escolares, como diretoras e professoras das escolas públicas sorteadas. O contato regular com as escolas aliado a estratégia de devolutivas pedagógicas regulares foi importante para maximizar as taxas de resposta do estudo, assim como, devolver para a rede estudada e escolas os principais resultados da pesquisa de forma rápida.

A rede conveniada, isto é, instituições privadas que recebem aporte público para manter algumas ou todas as vagas com desconto ou bolsas integrais, também apresenta uma amostra por conglomerado. Ao total foram sorteadas 11 instituições escolares conveniadas que ofertam a etapa da pré-escola, contudo, após o início da coleta de dados a equipe de pesquisa constatou que algumas das escolas conveniadas cobravam mensalidade da grande maioria das famílias. Algumas famílias recebiam bolsa integral, mas essa informação não foi coletada junto as famílias e/ou escola. Dessa forma, não é possível saber se as crianças que participaram do estudo eram bolsistas ou as famílias efetivamente pagavam mensalidades escolares. Nas análises apresentadas, um total de seis escolas conveniadas serão tratadas como escolas da rede privada e alocadas no seu respectivo estrato de acordo com a mensalidade cobrada das famílias, sendo 2 como escolas privadas de mensalidade média e 4 de mensalidade alta. As cinco instituições que não foram realocadas foram removidas das análises.

As escolas privadas, por sua vez, foram selecionadas a partir de uma amostra por conglomerado de conveniência, não probabilística, e estratificada pelo valor da mensalidade cobrada: a) escolas com mensalidade de baixo valor (até R\$400,00); b) escolas com mensalidade de valor médio (de R\$401,00 a R\$1000,00); c) escolas com mensalidade de alto valor (mais de R\$1000,00). A amostra das escolas privadas apresenta características específicas, em especial quando comparamos a rede pública.

Inicialmente a equipe de pesquisa pensou em realizar uma seleção aleatória das escolas privadas a partir de sorteio. Isso foi feito, mas muitas escolas não responderam ao primeiro contato da equipe de pesquisa – feito por carta, e-mail ou ligação telefônica. A negociação para a aprovação da pesquisa com as escolas privadas é feita de forma individual com cada estabelecimento de ensino, o que tornou o processo demorado e mais complexo. Muitas escolas nunca responderam os contatos iniciais feitos pela equipe da UFRJ e, diante do cenário desafiador, foi tomada a decisão de substituir as escolas sorteadas por outras escolas com perfil semelhante (considerando o valor da mensalidade). Esse processo de substituição foi intenso e diante da não resposta de diversas instituições, algumas escolas foram abordadas a partir de contato prévio da equipe de pesquisa com algum membro da comunidade escolar. Diante do exposto, é importante enfatizar que a amostra das escolas da rede privada não é representativa para esse tipo de oferta para o município estudado, mas buscou-se manter a característica de uma grande heterogeneidade no perfil discente observado na rede privada a partir do valor cobrado nas mensalidades. Apesar da não representatividade garantida, trata-se, possivelmente, da melhor amostra produzida nos anos recentes para estudar a rede privada com atendimento à pré-escola. Os resultados apresentados que comparam a rede pública e privada devem ser interpretados com cautela considerando as duas principais ameaças descritas até o momento nesse capítulo de métodos. A primeira referente a ameaça histórica – dados para a rede pública foram coletados em 2017 e para a rede privada em 2019. A segunda ameaça tem relação com a limitação da amostra das escolas privadas, que demanda cuidados no processo de generalização dos resultados, assim como, em comparação com a rede pública estudada na mesma cidade. Não obstante, é ressaltado que os resultados apresentados não encontram paralelo nacional na revisão de bibliografia realizada nessa dissertação considerando a faixa etária analisada.

Uma vez a escola aceitando participar do estudo, todas as crianças matriculadas em turmas do 1º ano da pré-escola (pré-I) compunham a amostra potencial da pesquisa. Todos esses protocolos foram aprovados pelo Comitê de Ética em Pesquisa e a coordenação do estudo respeitava procedimentos adicionais de autorização que cada instituição solicitasse, como o envio de carta para a casa das crianças explicitando todo o trabalho realizado, além de ficar à disposição das famílias para esclarecimento de dúvidas. Reforçamos, ainda, que todas as crianças foram perguntadas em gravação de áudio se desejavam realizar a avaliação cognitiva, apresentando de antemão o material e explicitando o processo que realizariam, dando a elas a total liberdade de parar a avaliação a qualquer momento caso participassem e que, caso não quisessem, não precisavam realizar a avaliação. Raras foram as negativas das crianças e, dentre aquelas que negaram, era comum que posteriormente solicitassem para realizar a atividade proposta pelo pesquisador diante do caráter lúdico da avaliação e, também, pelo incentivo de ver outras crianças realizando a atividade.

Nas tabelas abaixo, podemos ver a realocação das escolas conveniadas em escolas privadas com cobrança de mensalidade (Tabela 1) e, conforme o estrato, o total de crianças matriculadas em turmas de 1º ano da pré-escola (amostra intencionada) e aquelas que efetivamente participaram do estudo (amostra alcançada).

Tabela 1: Composição amostral após redistribuição das escolas conveniadas

Rede	Estrato	N Escolas inicialmente	Redistribuição conveniadas	N Escolas final (após redistribuição)
PÚBLICA	Pública	46	0	46
PRIVADA	Conveniadas	11	-6	Removidas da amostra
	Baixa mensalidade (até R\$400)	5	0	5
	Média mensalidade (R\$401 – R\$1000)	8	+2	10

Rede	Estrato	N Escolas inicialmente	Redistribuição conveniadas	N Escolas final (após redistribuição)
	Alta mensalidade	10	+4	14

Fonte: Elaboração própria do autor.

Tabela 2: Amostra do estudo: intencionada e alcança (1ª Onda)

Rede	Estrato	Escolas	N crianças matriculadas Pré-I	N crianças participantes Pré-I	% crianças participantes
PÚBLICA	Pública	46	3965	2710	68,34%
PRIVADA	Baixa mensalidade (até R\$400)	5	144	140	97,22%
	Média mensalidade (R\$401 – R\$1000)	10	203	138	67,98%
	Alta mensalidade (acima de R\$1000)	14	411	274	66,67%
Total		75	4723	3262	69,06%

Fonte: Elaboração própria do autor.

A tabela 1 apresenta que das 11 escolas conveniadas coletadas originalmente no estudo, 6 foram realocadas para outros estratos, sendo 2 para o grupo de escolas com média mensalidade (R\$401-R\$1000) e 4 instituições para escolas de alta mensalidade (acima de R\$1001). As cinco instituições que permaneceram classificadas como conveniadas foram removidas da amostra. A tabela 2 foi construída levando em consideração a quantidade de crianças matriculadas e alcançada do estudo após a redistribuição das escolas conveniadas. A rede pública compõe a maior parte da amostra alcançada do estudo, representando 83% (2716 de 3268 matrículas). A porcentagem de crianças alcançadas em cada estrato é muito parecida, sendo próxima de 70%, com exceção das escolas de baixa mensalidade em que praticamente todas as crianças da amostra intencionada participaram.

Tabela 3: Amostra de crianças participantes na 1ª e 2ª ondas

Rede	Estrato	Total de crianças participantes da 1ª onda	Total de crianças da 1ª onda participando na 2ª onda	% das crianças participantes da 1ª onda encontradas na 2ª onda
PÚBLICA	Pública	2710	2212	81,62%
PRIVADA	Baixa mensalidade (até R\$400)	140	129	92,14%
	Média mensalidade (R\$401 – R\$1000)	138	122	88,41%
	Alta mensalidade (acima de R\$1000)	274	249	90,88%
Total		3262	2712	83,14%

Há dois tipos de perda amostral que devemos nos atentar. O primeiro é quanto à amostra intencionada e a amostra alcançada, que pode ser vista na tabela 2 e é exemplificado principalmente pela dificuldade de encontrar a criança na escola nos dias em que os pesquisadores realizavam as visitas ou, principalmente nas escolas privadas, quando o responsável pela criança não encaminhava a autorização de participação na pesquisa – normalmente devido a uma não resposta, ao invés de uma negativa. O segundo tipo se refere à perda amostral longitudinal (atrato), que ocorre quando não foi possível encontrar ao final do ano (2ª onda) a criança participante no início do ano (1ª onda). A tabela 3 apresenta esse tipo de perda que é justificada, também, pela falta da criança nos dias de visita dos pesquisadores, mas, em especial, por sua mudança de instituição, o que dificultava em muito encontrá-la.

A coordenação do estudo mantinha contato constante com as instituições para realizar a pesquisa com as crianças faltosas, enviando pesquisadores às escolas em diversos dias e, com isso, coletar os dados com quem faltou em dias anteriores. Ainda assim e diante desse contexto,

houve uma relativa perda entre a amostra intencionada e amostra participante de 31%, além de não encontrar ao final do ano 16,8% das crianças que participaram da pesquisa no início do ano.

Diante desse cenário de desenho e amostra é preciso explicitar as três principais ameaças descritas no capítulo de métodos:

- A) Ameaça histórica, uma vez que os dados de escolas públicas e privadas foram coletados em épocas diferentes, sendo um em 2017 e outro em 2019. A aprovação da BNCC com ampla divulgação na imprensa e com posterior implementação nas redes pode ser vista como uma ameaça histórica, em especial considerando que há um documento específico para nortear o currículo na educação infantil. Nesse caso, o risco é que a BNCC permita uma melhor estruturação dos currículos, o que em grande medida dialoga com parte dos itens dos testes de linguagem e matemática. Se isso de fato ocorrer, o risco é superestimar os efeitos da rede privada (coletada em 2019) em relação a rede pública (coletada em 2017), diante de possíveis mudanças nos currículos durante esse período.
- B) Viés de seleção na amostra das escolas privadas. A amostra de escolas privadas não é representativa da rede privada do município, porque diante da não resposta de muitas escolas inicialmente sorteadas para participar da pesquisa, a equipe da UFRJ iniciou contato com escolas que havia um contato prévio, respeitando a proporção de escolas nos diferentes estratos. Há ainda uma sobre representação das escolas de alta mensalidade na amostra das escolas privadas, o que pode inflar os resultados da rede privada, uma vez que mais escolas com alunado de perfil socioeconômico alto fazem parte do estudo. Variabilidade de resultados cognitivos e sua forte correlação com características socioeconômicas e sociais nos permitem realizar controles para tentar diminuir a possibilidade de tal ameaça sobre os resultados e sua generalização.
- C) Mortalidade amostral durante a coleta de dados no primeiro ano na escola. Um tipo de efeito comum aos estudos longitudinais, que ocorre quando indivíduos participantes da pesquisa não são encontrados em coleta de dados subsequentes. Essa perda pode não ser aleatória, detendo perfis específicos de crianças e familiares.

A condição ideal para análise de dados de uma pesquisa seria que nenhuma dessas ameaças ocorresse, o que tende a ser incomum em estudos educacionais de larga escala. São ameaças a validade interna e externa que demandam cautela na interpretação dos dados,

em especial na magnitude das diferenças observadas entre as redes de ensino. O mais provável é que a direção final dos resultados esteja correta, mas o tamanho/magnitude das diferenças detenham viés. Explicitar as ameaças, contudo, não remove a importância desse estudo executado pelo LaPOpE/UFRJ, que continua a ser um estudo pioneiro a mensurar: a) as desigualdades no início da pré-escola – primeiro ano da escolarização obrigatória, b) como as desigualdades se transformam ao longo do primeiro ano da pré-escola; c) como as redes de ensino, em especial a distinção entre as redes pública e privada, contribuem para as desigualdades de aprendizado em uma grande metrópole.

3.2. Instrumentos

Os dados utilizados na dissertação se dividem em duas grandes categorias: a) medidas diretas sobre o desenvolvimento cognitivo das crianças e b) dados contextuais sobre as famílias dos estudantes e informações sobre a própria criança. Os dados cognitivos foram coletados utilizando uma versão adaptada para o contexto brasileiro do instrumento PIPS (*Performance Indicator in Primary Schools*), gerando medidas de desenvolvimento cognitivo em linguagem e matemática. Para compor os dados contextuais foram produzidos questionários pela equipe do LaPOpE/UFRJ (BARTHOLO et al., 2019; 2020) e aplicado aos responsáveis dos estudantes, tendo algumas informações complementadas pelas escolas e Secretaria Municipal de Educação (SME), no caso das instituições públicas.

Os dados coletados irão compor as variáveis apresentadas no Quadro 1 abaixo, indicando de quais fontes cada variável foi composta. No capítulo 3.3 é explicitado o como as variáveis dependentes Linguagem e Matemática, assim como as variáveis independentes de Ambiente de Aprendizagem em Casa (AAC) e Nível Socioeconômico (NSE) são construídas.

Quadro 1: Descrição das variáveis analisadas na dissertação

Nome	Tipo	Descrição	Fonte
Tipo de rede	Catégorica	Indica a rede da escola, podendo ser pública ou privada	Escola

Nome	Tipo	Descrição	Fonte
Tipo de estrato	Categórica	Indica o estrato da escola que a criança está matriculada, podendo ser: escola pública e escolas privadas separadas em mensalidade: - Baixa (até R\$400); - Média (R\$401 - R\$1000); - Alta (acima de R\$1000)	Escola
Linguagem 1ª Onda	Contínua	Medida de linguagem calculada em escala Rasch para o início do ano letivo do 1º ano da pré-escola (1ª Onda)	PIPS ¹
Linguagem 2ª Onda	Contínua	Medida de linguagem calculada em escala Rasch para o final do ano letivo do 1º ano da pré-escola (2ª Onda)	PIPS
Matemática 1ª Onda	Contínua	Medida de matemática calculada em escala Rasch para o início do ano letivo do 1º ano da pré-escola (1ª Onda)	PIPS
Matemática 2ª Onda	Contínua	Medida de matemática calculada em escala Rasch para o final do ano letivo do 1º ano da pré-escola (2ª Onda)	PIPS
Idade	Contínua	Idade da criança em meses de vida	Questionário; Escola; SME ²
Sexo	Dummy	Indica o sexo da criança (feminino e masculino)	Questionário; Escola; SME
Nível Socioeconômico (NSE)	Contínua	Indicador de nível socioeconômico da família calculada em escala Rasch	Questionário; SME
NSE Escola	Contínua	Indicador de nível socioeconômico da escola calculada através da média de NSE das famílias das crianças matriculadas	Questionário; SME
Ambiente de Aprendizagem em Casa (AAC)	Contínua	Indicador de Ambiente de Aprendizagem em Casa calculada em escala Rasch	Questionário; SME
Frequência à creche	Dummy	Indica se a criança frequentou a creche	Questionário; SME

1: PIPS: Instrumento cognitivo – *Performance Indicator in Primary Schools*;

2: SME: Banco de dados administrativos da secretaria municipal de educação foi fonte complementar de dados apenas para crianças matriculadas em escolas públicas.

Fonte: Elaboração própria do autor.

3.2.1. Instrumento de dados contextuais: o questionário de responsáveis

O questionário de pais e responsáveis tem como principal função captar informações sobre o contexto familiar da criança, possibilitando estabelecer variáveis de controle sobre o efeito composição discente na escola, relações e oportunidades de aprendizagem dentro do lar e estabelecer controle de maturação biológica através da idade da criança. Foi através dos dados provenientes dos questionários de responsáveis que foi possível construir as variáveis latentes de Nível Socioeconômico (NSE) e Ambiente de Aprendizagem em Casa (AAC), elaboradas a partir de Teoria de Resposta ao Item, como veremos mais à frente.

É amplamente estudado no campo educacional, seja nacional ou internacional, a forte correlação existente entre Nível Socioeconômico (NSE) e sucesso escolar, ganhando desde meados do século XX através dos resultados dos grandes surveys, como o Relatório Coleman em 1966, um traço indispensável dos estudos de larga escala a necessidade de se incluir variáveis de NSE como controle dos modelos matemáticos (BROKE; SOARES, 2008; CRAHAY; BAYE, 2013; HECKMAN, 2006; RIBEIRO, 2011; SYLVA, 2006). Por outro lado, o Ambiente de Aprendizagem em Casa (ACC) é proveniente dos recentes estudos sobre desigualdade e desenvolvimento de crianças pequenas, em que se busca investigar o como as interações e oportunidades educativas no seio familiar e contexto comunitário promovido pelos adultos, ambiente este mais direto e primário de estímulos educativos dos pequenos, influenciam sobre o que as crianças aprendem e seus pontos de partida na entrada da vida escolarizada (TAGGART et al., 2011; SYLVA et al., 2010; SAMMONS et al., 2008). Trata-se de um conceito estabelecido internacionalmente como “*home learning enviowment*” e se funda a partir de modelo bioecológico de desenvolvimento (BRONFENBRENNER; MORRIS, 2006). Um estudo aprofundado sobre o Ambiente de Aprendizagem em Casa (ACC) na bibliografia nacional pode ser visto em Koslinski et al. (2021, PRELO).

Diferentes estratégias foram utilizadas para maximizar a taxa de resposta do questionário contextual da família. Foram desenvolvidas três abordagens distintas para captar essas informações, ficando a cargo das escolas quais seriam possíveis de serem implementadas: a) envio do questionário impresso para casa da criança e posterior devolução aos pesquisadores; b)

pesquisadores foram às escolas e aplicaram *in loco* os questionários juntamente dos responsáveis quando fossem buscar as crianças ao final de um dia letivo; c) envio de questionário de responsáveis digital, disponibilizado de forma *online* através de um *link* sendo possível de ser respondido a qualquer momento desde que tivesse conexão com a internet, tanto por aparelhos móveis, como *smartphones*, quanto computadores.

Algumas instituições optaram por utilizar todas as abordagens, outras apenas uma e, ainda, houve aquelas que indicaram para que os pais não fossem solicitados a responderem aos questionários. A coordenação da pesquisa negociou o que era possível e atendeu todos os pedidos das escolas participantes.

Diversos foram os desafios para coleta de dados contextuais. Exemplifica-se o envio de questionário impresso para casa, mas sem devolução; crianças que chegam desacompanhadas de seus responsáveis na escola, seja por irem de transporte escolar, transporte público, normalmente acompanhada por um irmão ou irmã mais velha, ou, ainda, por irem de motorista particular, casos que impossibilitam coleta de dados *in loco*. Para além disso, ainda que se fosse garantida o anonimato na pesquisa e o não compartilhamento de informações, foi observado que alguns pais ficavam receosos de responder questões relativas à renda domiciliar ou itens sobre o perfil socioeconômico, preocupados, por exemplo, que isso pudesse afetar na concessão ou permanência em programas de transferência de renda ou, ainda, bolsa de estudos em escolas. Por fim, mesmo com a disponibilidade do questionário digital e seu envio via agenda eletrônica da escola ou e-mail, diversos responsáveis não respondiam, seja por falta de interesse, fosse por desconfiança ou por ser mais uma obrigação no cotidiano, situações corriqueiras ao ato de pesquisar. Em todos os casos, os pesquisadores eram treinados para explicar a pesquisa, retirar dúvidas e, ainda, caso fosse necessário, os e-mails dos coordenadores da pesquisa eram disponibilizados para a necessidade de maiores esclarecimentos. Todos esses desafios impossibilitaram a coleta integral de dados contextuais para todas as famílias de crianças participantes da pesquisa.

A Tabela 4 apresenta a taxa de resposta do questionário de responsáveis considerando os diferentes estratos.

Tabela 4: Taxa de resposta dos questionários de responsáveis (apenas pré-I)

Rede	Estrato	N participante	N respondido	%Respondido
Pública	Pública	2716	1798	66,20%
Privada	Baixa mensalidade	140	63	45%
	Média mensalidade	138	68	49,30%
	Alta mensalidade	274	145	52,90%
TOTAL		3268	2074	63,46%

Fonte: Elaboração própria do autor.

As informações contextuais foram utilizadas como variáveis de controle nos modelos multivariados presentes nas análises da dissertação. Contudo, diante de dados faltantes para diversas crianças, optou-se por utilizar de técnica de múltipla imputação para que não houvesse perda amostral. Esse processo será descrito mais à frente no capítulo de dados faltantes/ *missing data* e imputação de dados, detalhando a proporção de dados faltantes para cada estrato de escolas, assim como, as estratégias adotadas no modelo de múltipla imputação. O Quadro 1, anteriormente exposto, apresenta uma descrição das principais variáveis da pesquisa.

3.2.2. Instrumento cognitivo: PIPS

Desenvolvido na Inglaterra em 1994 por pesquisadores do *Centre for Evaluation and Monitoring* (CEM) da Universidade de Durham, o *Performance Indicator in Primary Schools* (PIPS) foi pensado inicialmente para ser um sistema de monitoramento nacional sobre a educação de crianças no primeiro ano na escola, utilizando um teste adaptativo que gera dados padronizados e confiáveis sobre o desenvolvimento dos alunos no primeiro ano de escolarização obrigatória – na Inglaterra as crianças teriam entre 5 e 6 anos. O teste permite a descrição do desenvolvimento das crianças em linguagem e matemática e passou a ser usado por professores e

profissionais que atuam diretamente nas escolas como ferramenta de diagnóstico e apoio pedagógico, possibilitando devolutivas sobre desenvolvimento das crianças em sala de aula (TYMMS et al, 2014).

Desde então, o uso do PIPS vem sendo ampliado e utilizado para pesquisa em educação em diversos países do mundo – em geral com medidas sobre o desenvolvimento cognitivo no início e final do primeiro ano da escolarização obrigatória, o que pode variar bastante dependendo do país analisado. As demandas governamentais e de pesquisa por dados padronizados levaram no mundo todo ao desenvolvimento de avaliações com escalas comparáveis e confiáveis para diferentes contextos, tendo as avaliações como PISA, TIMMS e PIRS para tal, que focam, contudo, o Ensino Fundamental e Médio. Diante de uma lacuna de dados comparativos internacionais sobre o desenvolvimento das crianças quando iniciam o processo obrigatório de escolarização, o PIPS foi adaptado e traduzido para diferentes línguas com o intuito que fosse utilizado em diferentes países, recebendo tradução e sendo adaptado contextualmente. Desde o início da elaboração do instrumento, o PIPS já foi utilizado em países como Alemanha, Rússia, China, África do Sul, Austrália e Nova Zelândia. Testes de confiança de compatibilidade de escalas apresentam que as adequações mantêm as características originais do instrumento. O PIPS não busca criar ranqueamento entre escolas e crianças, mas ser suporte pedagógico para se observar de forma robusta e com dados de qualidade o desenvolvimento infantil, tanto no que tange aspectos pedagógicos em sala de aula quanto para o desenvolvimento de políticas públicas baseadas em evidências. Fica destacado tal premissa no projeto iPIPS que:

Tabelas de ranqueamento não serão produzidas. Em vez disso, cada país receberá um relatório que contextualiza os resultados com dados comparáveis. Medidas robustas serão tomadas, tanto na concepção das avaliações quanto a apresentação dos resultados, para evitar que o estudo se torne uma ferramenta de *'high stakes'* para escolas ou professores ou, ainda, para evitar que exerçam influência indevida sobre a natureza e o conteúdo de currículos pré-escolares. iPIPS, 2014. Tradução livre.

Para pesquisa, o PIPS é aplicado de forma clínica, com um pesquisador da UFRJ treinado realizando a avaliação com uma criança. Utilizando-se de um *tablet* e um livro ilustrado de suporte, o pesquisador faz perguntas à criança e solicita para que ela responda apontando para imagens e palavras no livro, anotando as respostas corretas e erradas no *software* do aparelho. Ainda que o PIPS completo levasse um longo tempo para ser respondido por uma criança em

idade pré-escolar, a aplicação dura, em média, entre 10 e 20 minutos, uma vez que se trata de um teste adaptativo, em que os itens considerados mais difíceis só são apresentados conforme a criança não erre de forma consecutiva os itens anteriores de cada tópico, considerados mais fáceis. Como foi criado pensado para ser aplicado no início e ao final do ano letivo, a criança, ao realizar o PIPS ao final do ano, inicia de uma etapa próxima da qual parou na primeira aplicação. Isso faz com que não seja exposta à maioria dos itens que já havia acertado, possibilitando um tempo de aplicação mais curto do que se partisse desde o início. Isso busca evitar um tempo longo de aplicação e, por conseguinte, um cansaço desnecessário à criança que poderia comprometer a qualidade da informação coletada.

O instrumento é dividido em itens de linguagem e matemática que são compostos de cinco grandes categorias cada.

A dimensão de linguagem é composta por:

- Vocabulário receptivo
- Consciência fonológica
- Ideias sobre leitura
- Identificação de letras
- Leitura (palavras e frases)

A dimensão de matemática é composta por:

- Ideias sobre matemática
- Contagem de objetos
- Identificação de formas
- Identificação de números, somas e subtração
- Solução de problemas de aritmética formal

O PIPS em versão brasileira foi construído através de um longo processo de adaptação. Foi traduzido para o português e, depois, realizada tradução reversa para o inglês no intuito de garantir que as perguntas estariam captando a mesma dimensão, independente da língua. Além disso, diversos itens sofreram adaptações culturais, como a inclusão de moedas nacionais em substituição a outras moedas em perguntas sobre matemática, por exemplo. Para verificar a escala, foram realizados pré-testes com crianças entre 4 e 7 anos de idade, indo do primeiro ano da pré-escola ao primeiro ano do ensino fundamental, buscando garantir que a escala não sofresse com problemas de medição como efeito-teto. (BARTHOLO et al., 2020). Os dados do instrumento cognitivo permitem, em tese, não apenas comparar as mesmas crianças ao longo do tempo, como também compará-las com outros contextos, sejam eles nacionais ou internacionais.

3.2.3. Indicadores em escala Rasch: matemática, linguagem, Nível Socioeconômico (NSE) e Ambiente de Aprendizagem em Casa (AAC)

Os resultados obtidos pelo PIPS são expressos em itens acertados, errados ou não respondidos. Utilizar a quantidade de acertos em uma avaliação é uma forma intuitivamente aplicada para analisar desempenho. Contudo, um problema importante nessa abordagem é que, indivíduos que obtiveram as mesmas notas podem ter acertado diferentes itens de diferentes complexidades. Ou seja, uma nota 5 de 10 acertos possíveis pode significar um desempenho completamente diferente de outra pessoa que acertou cinco questões, mas itens diferentes.

Isso significa que a escala da abordagem supracitada não detém um fator importante para garantir a confiabilidade de comparação dos resultados: linearidade da escala (BOONE, 2016). Ou seja, a distância entre a nota 4 e 5 (que é 1), não tem o mesmo significado na escala do que a distância entre 9 e 10, que também é 1. Numerais de valores iguais, mas significados diferentes. É como se “1cm” ao final de uma régua que mede a altura de uma pessoa não tivesse o mesmo espaçamento de “1cm” no início da mesma régua, o que geraria viés de medição.

Para contornar esse problema, é preciso lançar mão de uma escala que leve em consideração a dificuldade do item apresentado na avaliação, assim como, a distribuição de respostas. Para compor a escala de matemática e linguagem, que se expressam como as variáveis dependentes dos modelos da presente dissertação, utilizou-se de medida Rasch, que pode ser

compreendido como um subconjunto dos grupos de modelos conhecidos como Teoria de Resposta ao Item (TRI) (BOONE, 2016; KHINE, 2020). A escala foi gerada utilizando o pacote TAM (ROBITZSCH, KIEFER E WU, 2021) no software R através de técnica de modelos de resposta ao item unidimensional via *joint maximum likelihood estimation* (LINACRE, 1994), gerando resultados compatíveis com os softwares mais comumente utilizados, mas que não são gratuitos, como *Winsteps* e *Facets*. A mesma técnica foi utilizada na construção do indicador de Nível Socioeconômico (NSE) e Ambiente de Aprendizagem em Casa (AAC).

A construção de uma escala em TRI vai na direção do que já vem sendo apropriado pelas pesquisas educacionais de construir variáveis de traços latentes, isto é, que não medem diretamente o que se busca, mas utiliza-se de outros fatores que permitem medir indiretamente (SOARES, 2005). Assim é possível criar variáveis contínuas, de escala linear e comparativas, auxiliando contornar problemas de ordem de coleta de dados, como a dificuldade de perguntar diretamente a renda *per capita* familiar, medida que poderia ser utilizada como principal variável de nível socioeconômico, mas que raramente é preenchida pelos respondentes, como foi verificado em nosso campo de pesquisa.

Através dessa prática é possível captar diversas informações sobre uma dimensão e criar uma única variável de escala contínua. O Nível Socioeconômico (NSE), por exemplo, é uma dimensão de interesse e, para construir uma única variável contínua de NSE, foram perguntados no questionário de responsáveis diversos elementos que estão relacionados ao nível socioeconômico de uma família: se havia quartos exclusivos para os filhos em casa, conexão de internet na residência, posse de carro, TV à cabo, máquina de lavar, computador, *tablet*, impressora, responsável com formação superior e se a família recebia alguma auxílio do governo, como Bolsa Família ou Benefício de Prestação Continuada. Essas perguntas geram respostas dicotômicas de “sim” e “não” que dificultam aos pesquisadores utilizar nos modelos matemáticos. Para contornar isso, utiliza-se da técnica de TRI supracitada em que todos esses fatores são levados em consideração e geram uma variável única de escala contínua. Ou seja, todas as respostas são agrupadas e calculadas de forma unificada, e, ao invés de diversas variáveis com “sim” e “não”, teremos uma única variável com valores como “-3.54” e “2.21”. Quanto menor o valor, menor o NSE da família e, quanto maior o valor, maior o NSE da família. Na prática, uma resposta indicando que há responsável com ensino superior na casa aumenta o

NSE, enquanto uma resposta indicando que a família recebe auxílio governamental, diminui o NSE.

O mesmo vale para o resultado cognitivo de linguagem e matemática. Ao invés de termos respostas “sim” e “não”, teremos a opção “acertou” e “não acertou” – resposta em branco é considerada como um “não exposto”. Todos os itens de matemática, por exemplo, são agrupados e calculados em TRI. Acertar uma pergunta “mais difícil” agregará mais pontos em matemática do que acertar uma pergunta “mais fácil”. Ou seja, de forma simplificada, é uma melhor medida do que simplesmente somar os acertos, uma vez que o TRI permite distinguir quais perguntas tendem a ser difíceis de serem acertadas e quais tendem a ser mais fáceis, adicionando à nota o “grau de complexidade” do conhecimento. Isso permite, na prática, que mesmo que uma criança acerte mais perguntas do que outra, ela possa ter uma nota menor caso acerte poucas perguntas mais complexas, enquanto outra criança que errou uma ou outra pergunta menos complexa mas conseguiu acertar itens mais difíceis, tenha uma nota maior, ainda que no total tenha acertado menos.

Por fim, a variável latente de Ambiente de Aprendizagem em Casa (AAC) foi construída utilizando o mesmo procedimento. Foram perguntados no questionário de responsáveis questões de se havia livros infantis em casa, posse de jogos infantis e se realizavam diversas atividades com a criança como, ler, cantar músicas, desenhar, brincar com letras do alfabeto, números e formas geométricas. Através de técnica de Rasch, criou-se uma variável única e de escala contínua de AAC (KOSLINSKI et al, 2021 PRELO).

3.3. Imputação de dados faltantes

Os dados faltantes, correntemente chamados de *missings*, são comuns no contexto dos estudos sociais já que os participantes da pesquisa podem tanto optar por não responder aos questionários, como, também, podem se esquecer de responder alguns itens. Uma forma de diminuir o impacto causado por esse fenômeno é se utilizando de técnica de imputação de dados, preenchendo o *missing* com provável valor que ali estariam caso fosse respondido. Diante de diversas formas de preencher os dados faltantes, optou-se por utilizar nesta dissertação técnica de múltipla imputação, que se utiliza dos dados observados para prever os prováveis valores dos

dados não observados (BUUREN, 2018). De forma simplificada, serão utilizadas diversas variáveis que foram possíveis de serem captadas do indivíduo, como, por exemplo, resultado cognitivo em matemática e linguagem, idade, frequência à creche, para prever os valores mais prováveis que um valor *missing* teria com base nas tendências dos valores observados em toda a base de dados, como, por exemplo, para um dado faltante em nível socioeconômico.

Na tabela 4 abaixo é apresentada a proporção de dados faltantes para as variáveis explicativas que serão utilizados nos modelos matemáticos da dissertação.

Tabela 5: Dados faltantes e observados

Variável	Estrato	Respostas válidas (V)	Respostas faltantes (F)	Total crianças (T = V+F)	% Dados Faltantes (F/T)
Ambiente de Aprendizagem em Casa (AAC)	Público	1696	1014	2710	37,42%
	Baixa Mensalidade	63	77	140	55,00%
	Média mensalidade	67	71	138	51,45%
	Alta mensalidade	145	129	274	47,08%
	Total	1971	1291	3262	39,58%
Frequência à Creche	Público	2158	552	2710	20,37%
	Baixa Mensalidade	63	77	140	55,00%
	Média mensalidade	67	71	138	51,45%
	Alta mensalidade	145	129	274	47,08%
	Total	2433	829	3262	25,41%
Idade em meses	Público	2656	54	2710	1,99%
	Baixa Mensalidade	120	20	140	14,29%
	Média mensalidade	126	12	138	8,70%
	Alta mensalidade	255	19	274	6,93%
	Total	3157	105	3262	3,22%

Variável	Estrato	Respostas válidas (V)	Respostas faltantes (F)	Total crianças (T = V+F)	% Dados Faltantes (F/T)
Nível Socioeconômico (NSE)	Público	1786	924	2710	34,10%
	Baixa Mensalidade	63	77	140	55,00%
	Média mensalidade	68	70	138	50,72%
	Alta mensalidade	145	129	274	47,08%
	Total	2062	1200	3262	36,79%
Sexo	Público	2710	0	2710	0,00%
	Baixa Mensalidade	140	0	140	0,00%
	Média mensalidade	138	0	138	0,00%
	Alta mensalidade	274	0	274	0,00%
	Total	3262	0	3262	0,00%

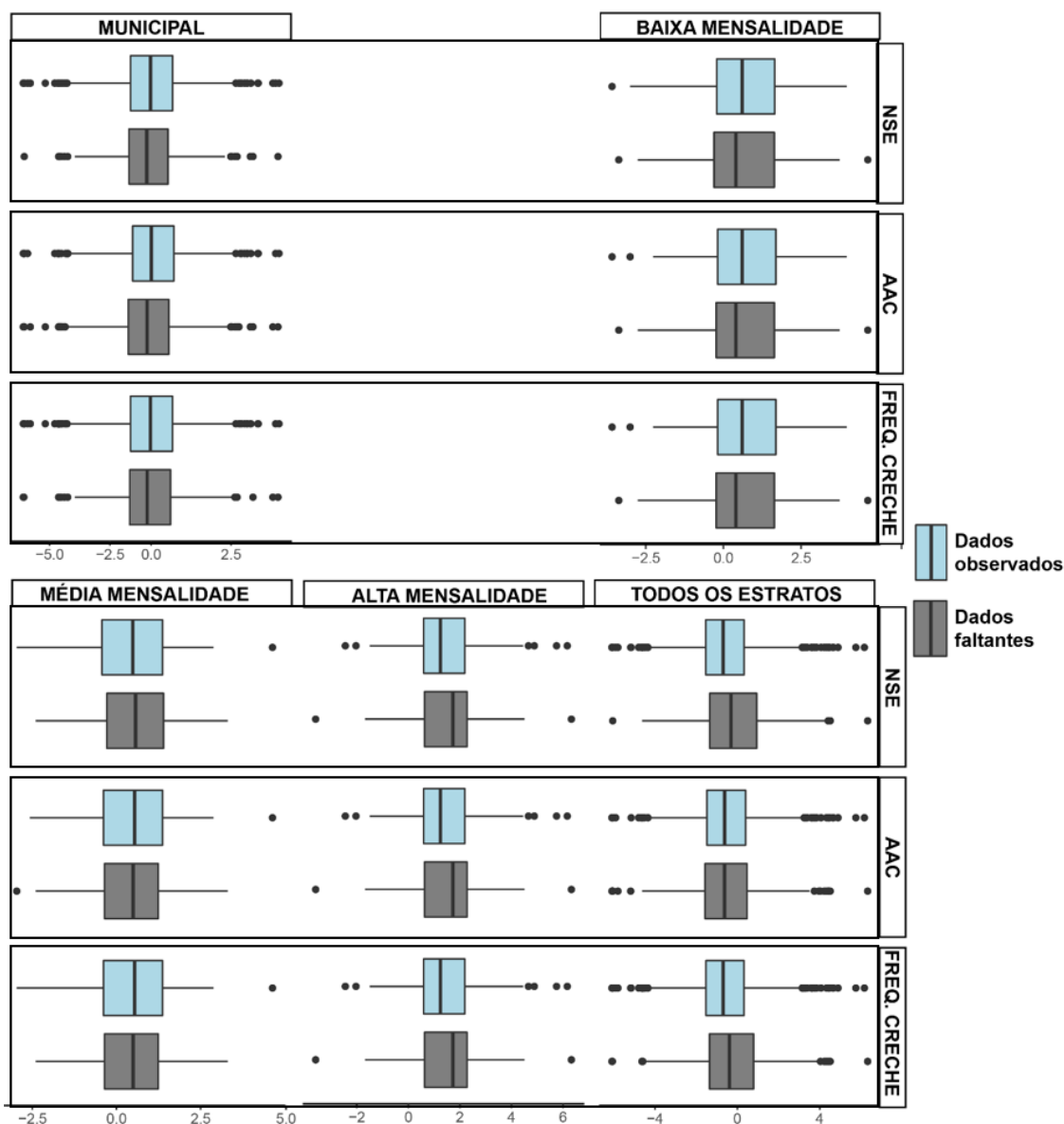
Fonte: Elaboração própria do autor.

Como pode ser visto, as variáveis de Nível Socioeconômico (NSE), Ambiente de Aprendizagem em Casa (AAC) e Frequência à Creche são as mais afetadas pela falta de dados, sendo, respectivamente, 36,79%, 39,58% e 25,41% no total. As taxas de dados faltantes para as escolas públicas tendem a ser menores pois, para além do uso de questionário de pais, foi possível incrementar os dados com informações provenientes do sistema centralizado de dados da Secretaria Municipal de Educação (SME). Nesse sentido, o processo de imputação teria como grande interesse evitar a perda de informações principalmente nessas três variáveis mais afetadas.

Para buscar compreender como os dados faltantes se comportam, as figuras 1 e 2 apresentam a distribuição dos dados observados e dos faltantes em relação ao desempenho de matemática e linguagem na 1ª Onda, respectivamente. Cada quadro na horizontal representa o comportamento de uma das variáveis explicativas que serão imputadas, sendo NSE, AAC e frequência à creche. Na vertical, é identificado o estrato da escola, podendo ser pública, baixa, média e alta mensalidade. No canto direito inferior de cada figura é apresentado o

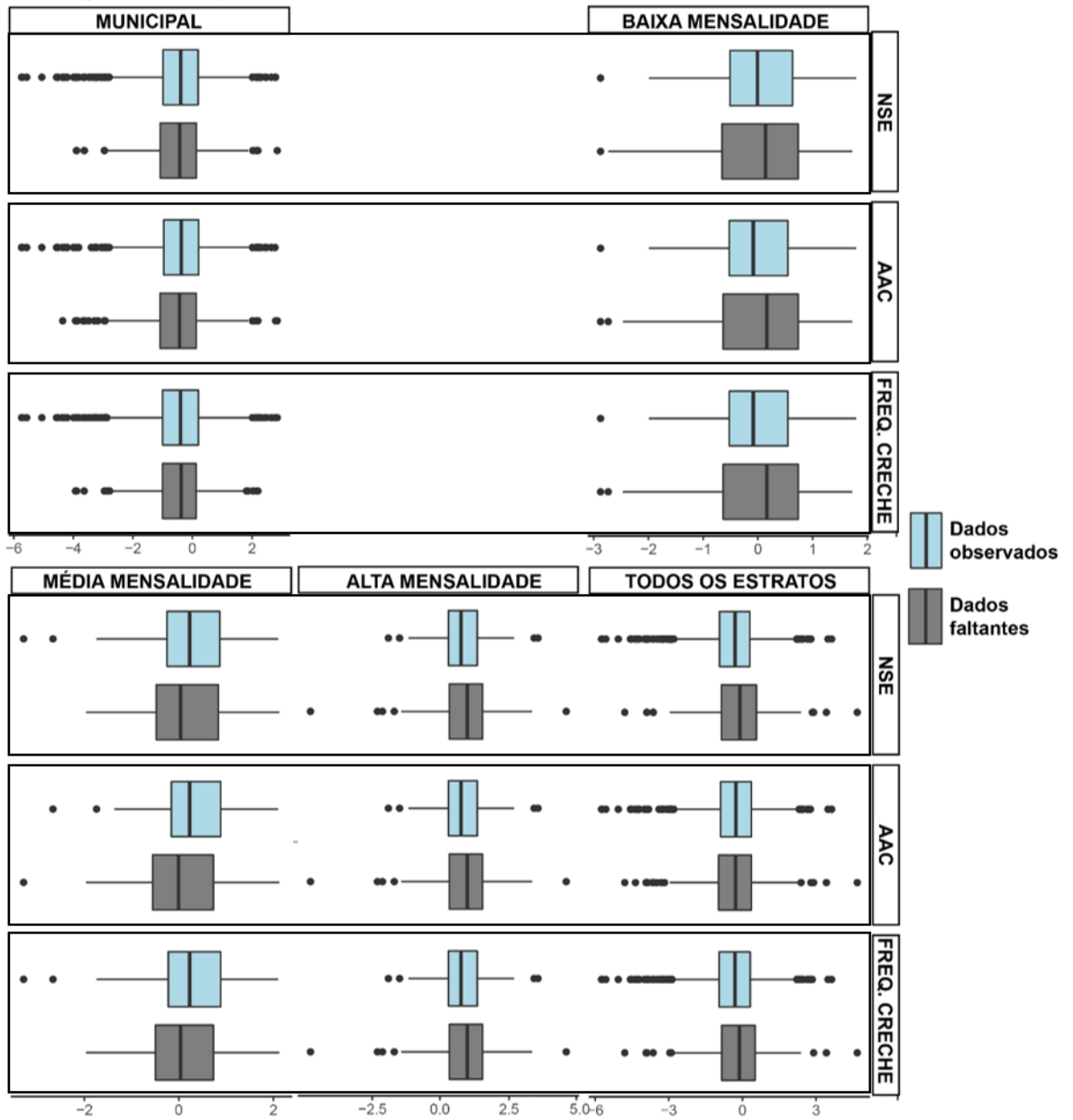
comportamento dos dados observados e faltantes levando em consideração o agregado de todos as crianças, sem diferenciação de estrato. É importante indicar que cada estrato tem em seu eixo X uma variação dos valores de desempenho cognitivo típico de seu agrupamento, uma vez que se está comparando as distribuições dos *boxplots* dentro do próprio estrato. Por isso, não é indicado a comparação entre os estratos, uma vez que os eixos X mudam conforme o estrato.

Figura 1: Distribuição de desempenho em Matemática na 1ª onda considerando dados observados e faltantes para as variáveis de NSE, AAC e frequência à creche.



Fonte: Elaboração própria do autor.

Figura 2: Distribuição de desempenho em Linguagem na 1ª onda considerando dados observados e faltantes para as variáveis de NSE, AAC e frequência à creche.



Fonte: Elaboração própria do autor.

O que se pode observar é que os dados faltantes para grande parte das variáveis de interesse analisadas se comportam de forma muito parecida com os dados observados, ainda que com leves deslizes para a direita, como em Matemática para alta mensalidade, ou mesmo para a esquerda, como em Linguagem para média mensalidade. Uma vez que o nível socioeconômico, o ambiente de aprendizagem em casa e a frequência em creche funcionam como preditores de desempenho, é presumível que quanto mais positivos são os desempenhos, por exemplo, maiores serão os valores das variáveis de interesse, em especial NSE e AAC.

Por essa característica provável de relação entre os dados observados e os dados faltantes, opta-se por realizar imputação múltipla como forma de contornar os malefícios de se perder casos por motivo de dados faltantes. Descrevo abaixo o processo.

Para realizar a imputação foi utilizado o pacote *mice* executado no software livre R, preenchendo os dados faltantes via imputações multivariadas por equações encadeadas (BUUREN e GROOTHUIS-OUDSHOORN, 2011). Compuseram o banco de dados a ser imputado as seguintes variáveis: desempenho em linguagem, desempenho em matemática, escala de nível socioeconômico (NSE), escala de ambiente de aprendizagem em casa (AAC), estrato da escola, rede da escola, indicador de sexo, indicador de ter frequentado creche e idade em meses.

As variáveis de Nível Socioeconômico, Ambiente de Aprendizagem em Casa e idade em meses foram imputadas utilizando método multinível através de modelos lineares mistos (comando *2l.pan*) indicando a escola enquanto segundo nível do modelo. Já a variável de frequência em creche foi imputada através de modelos generalizados mistos para variáveis binárias (comando *2l.bin*) com a escola enquanto segundo nível do modelo. Optou-se pela imputação multinível levando em consideração as orientações de Grund et al (2018), que apresenta que as imputações unidimensionais devem ser evitadas se a correlação de intraclasses (ICC) de um modelo de regressão multinível entre a variável independente e dependente principal for acima de 0.10. O ICC do modelo nulo entre os dados de matemática da 1ª Onda e escolas no nível 2 é de 0.33, enquanto para linguagem é 0.23, optando, portanto, por uma imputação multinível.

O resultado da imputação gerou cinco banco de dados empilhados, tendo em cada banco um valor distinto para cada dado faltante imputado. A prática mais usual de análise nesse cenário é através da chamada regra de Rubin (RUBIN, 1987), que leva em consideração o resultado obtido em cada banco de dados e depois os analisa de forma agregada, o que acaba por controlar de forma mais efetiva a incerteza dos resultados frente à imputação. Contudo, diante uma dificuldade técnica em gerar os *effect-sizes* após utilizar esse método, não se realizou as análises com os bancos de dados empilhados, isto é, através da regra de Rubin. Optou-se por escolher um banco de dados imputado em que a correlação entre as variáveis de AAC, NSE e frequência escolar com os resultados cognitivos de linguagem e matemática mais se aproximassem dessas correlações no banco de dados original (com dados faltantes). Após, extraiu-se as variáveis AAC, NSE e frequência escolar imputadas do banco escolhido para, então, compor as análises nos modelos. Antecipo que os resultados obtidos desse modo não modificam direção dos efeitos, mas foi utilizada como uma forma de contornar o problema técnico na geração de *effect-sizes*. As análises utilizando a regra de Rubin estão no anexo e podem ser comparadas com os as tabelas apresentadas nos resultados, se assim o leitor desejar. Debates mais aprofundadas entre as tabelas principais aqui apresentadas e a do anexo não serão apresentadas nesta dissertação por questão de tempo, mas serão foco de melhorias em trabalhos futuros, em especial após resolver o problema técnico que envolve o cálculo *effect-size* dos indicadores produzidos com a Regra de Rubin.

3.4. Resultado da imputação de dados faltantes

Nas análises abaixo, são apresentadas tabela e figuras que exploram os resultados da imputação em comparação aos dados observados.

Tabela 6: Descritiva do resultado de imputação para as variáveis de AAC, NSE e frequência à creche (observados e imputados).

Variável	Estrato	Observado		Imputado	
		Média	Desvio padrão	Média	Desvio padrão
Ambiente de Aprendizagem em Casa (AAC)	Público	-0,37	1,50	-0,41	1,48
	Baixa Mensalidade	0,93	1,01	0,68	1,24
	Média mensalidade	1,35	0,74	1,01	1,19
	Alta mensalidade	1,80	0,37	1,66	1,06
	Total	-0,11	1,56	-0,13	1,57

		Observado		Imputado	
Nível Socioeconômico (NSE)	Público	-0,46	1,42	-0,46	1,41
	Baixa Mensalidade	1,20	1,51	1,10	1,41
	Média mensalidade	1,84	1,23	1,46	1,30
	Alta mensalidade	2,45	1,05	2,08	1,30
	Total	-0,13	1,64	-0,10	1,62
Frequência à Creche (1 = Frequentou; 0 = Não frequentou)	Público	0,61	0,49	0,59	0,49
	Baixa Mensalidade	0,32	0,47	0,32	0,47
	Média mensalidade	0,42	0,50	0,23	0,42
	Alta mensalidade	0,54	0,50	0,45	0,50
	Total	0,60	0,49	0,55	0,50

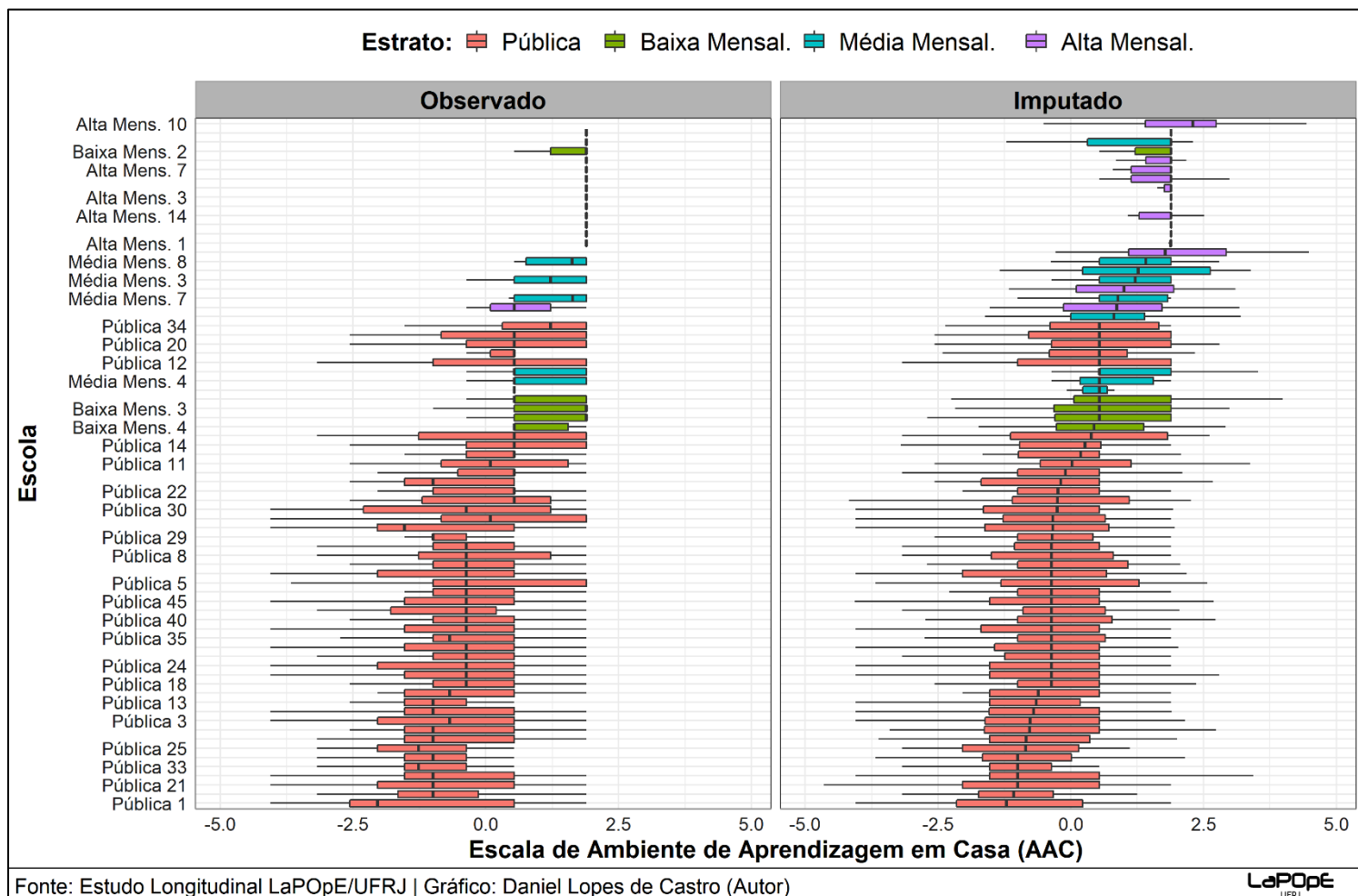
Fonte: Elaboração própria do autor.

Os dados da tabela 6 apresentam pequenas variações nos resultados observados e imputados. As mudanças ocorrem, principalmente nas casas decimais, mantendo, na grande maioria dos resultados, um desvio padrão próximo entre o observado e o imputado. As maiores diferenças observadas no desvio padrão são para o indicador de AAC para média e alta mensalidade, possivelmente porque aqueles que responderam nesses grupos deviam ter um perfil mais homogêneo, enquanto que, com a imputação, os perfis sem respostas passaram a receber informação, tornando o estrato mais heterogêneo e ampliando o desvio padrão. Outra explicação plausível para essa homogeneidade do valor de AAC é um possível efeito teto para certos grupos, em especial para as crianças matriculadas em escolas na média e alta mensalidade. Ou seja, os itens de práticas educativas em casa perguntados aos responsáveis foram basicamente todos respondidos positivamente, sendo incapaz de detectar variabilidades de práticas pedagógicas mais específicas no contexto do lar para esses grupos. Se esse for o caso, há um risco da imputação conter viés, uma vez que a imputação de dados para esses estratos ficaria de menor qualidade e, possivelmente, perderia capacidade de captar variabilidade para grupos de crianças matriculadas em escolas de média e alta mensalidade, superestimando os resultados desses estratos.

Esses possíveis comportamentos podem ser verificados na Figura 3 abaixo, em que se apresentam as distribuições de AAC observada e imputada para cada escola. As distribuições para NSE podem ser vistas na Figura 5. Ressalta-se que no eixo Y de cada figura foram omitidas alguns nomes apenas para priorizar a legibilidade. Os *boxplots* de cada figura estão ordenados

conforme a mediana dos dados imputados, fazendo com que quanto mais acima a escola, maior é a mediana de Ambiente de Aprendizagem em Casa (AAC) e Nível Socioeconômico (NSE) presente naquela determinada instituição. Além disso, foram incluídas cores para facilitar a compreensão da distribuição dos resultados conforme os estratos.

Figura 3: Distribuição de AAC para escolas da amostra (observado e imputado)



Fonte: Estudo Longitudinal LaPOPE/UFRJ | Gráfico: Daniel Lopes de Castro (Autor)

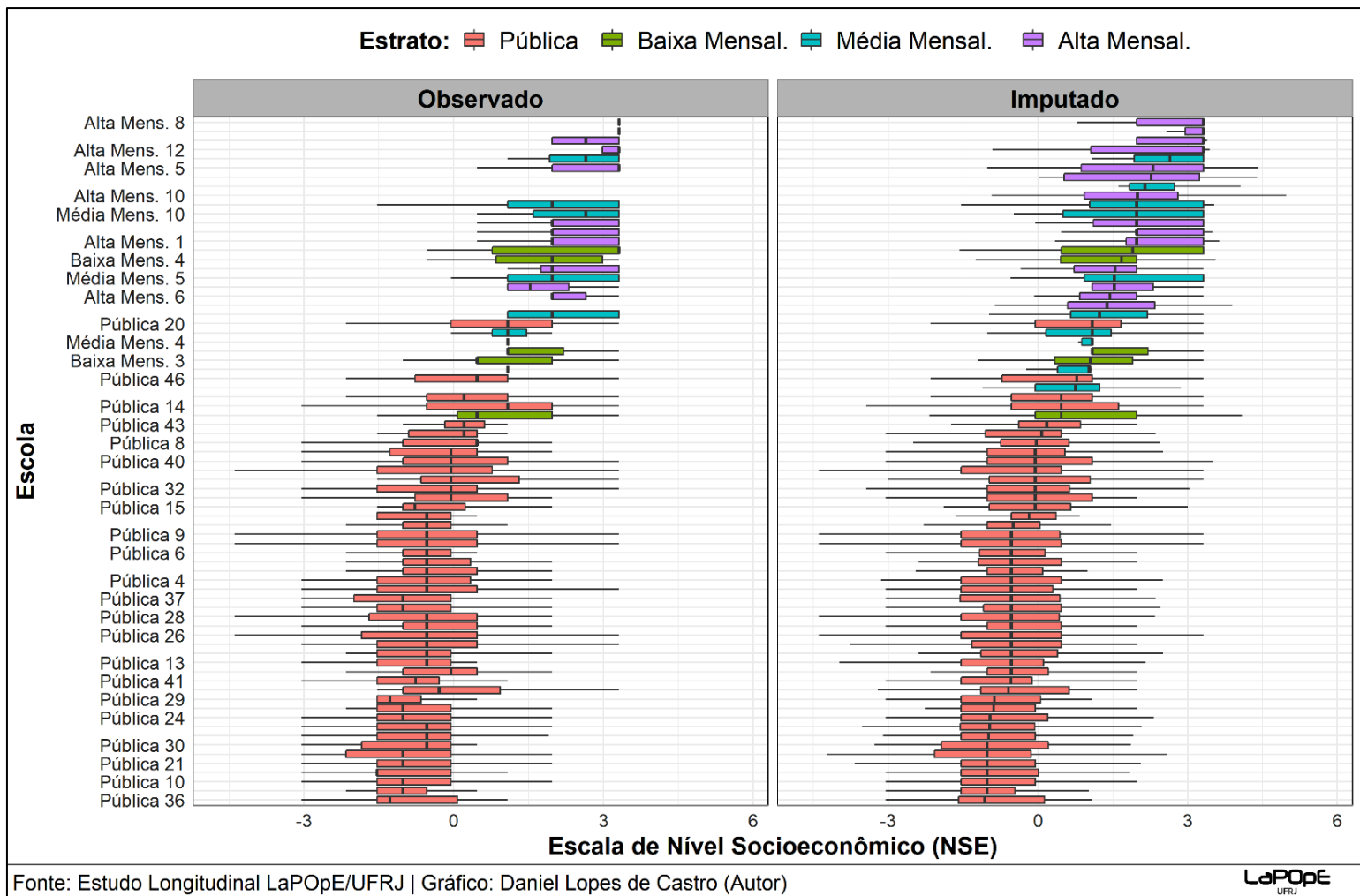
LaPOPE
UFRJ

Analisando a figura 3, o que podemos notar nos dados observados um possível efeito teto da variável de AAC. Podemos observar isso de forma mais clara nos grupos superiores do gráfico, principalmente relacionado às escolas privadas de alta mensalidade, que não há grande diferenciação entre os resultados obtidos por alunos. Isso pode ocorrer porque os responsáveis dessas crianças indicaram no questionário que realizam todas as atividades pedagógicas em casa

que foram perguntadas, como, por exemplo, ler para as crianças, brincar com números, contar e cantar. Em compensação, esse tipo de prática varia muito mais nos agrupamentos mais abaixo dos gráficos, completamente composta por famílias que têm seus filhos estudando em escolas públicas.

Ao imputar, ocorre a quebra do valor mínimo e máximo em ACC observado. Uma vez que a imputação se utiliza de regressão linear acaba por criar valores preditos conforme os controles estabelecidos e, possibilitando assim, imputar valores acima e abaixo dos observados. Mesmo diante dessa distorção, os resultados que serão apresentados ao final da dissertação seguem um padrão próximo aos dos dados observados, estando as tabelas de resultado com os dados observados localizados no anexo.

Figura 4: Distribuição de NSE para escolas da amostra (observado e imputado)



A figura 4 indica uma variabilidade de resultados percebida ao longo de todos os gráficos. Há um possível efeito teto também para a variável de NSE, mas que só é alcançada para os grupos localizados na parte superior. Nas duas figuras 4 e 5 o que chama a atenção é a distribuição das escolas. Há uma concentração grande de escolas públicas na base dos gráficos, enquanto as escolas privadas de alta mensalidade se concentram no topo do gráfico. Contudo, na figura 3, podemos ver no imputado um agrupamento de cinco escolas públicas que as famílias, na mediana, detêm valores de AAC acima de um agrupamento de escolas de baixa e média mensalidade, podendo indicar que práticas de aprendizagem em casa podem estar correlacionadas para além do nível socioeconômico, como elementos culturais, comunitários, compartilhamento de informações regionais entre pais e responsáveis ou, ainda, um possível efeito da própria escola sobre as informações que a comunidade detêm sobre práticas pedagógicas dentro de casa. Na

figura 5, o que chama atenção é que o padrão de distribuição do NSEs não é o mesmo de AAC. Nota-se a existência de famílias com alto nível socioeconômico, na mediana, mas que colocam seus filhos para estudar em escolas de baixa mensalidade. Isso pode estar ocorrendo porque essas são as escolas mais próximas de suas residências, por exemplo. Lembrando que os dados da dissertação foram obtidos em um grande município da região sudeste, fazendo com que a mobilidade possa ser um elemento importante na escolha das escolas por suas famílias. Apesar da presente dissertação não se debruçar em investigar tais elemento, defende-se que são questões que merecem futuras análises que podem auxiliar a compreender os resultados obtidos.

Nesse sentido, o que se nota é que o comportamento dos dados imputados seguem um padrão lógico do comportamento dos dados observados. A partir desse momento, os resultados que serão apresentados ao longo da dissertação estarão utilizando de dados imputados de AAC, NSE e indicador de frequência à creche antes de entrar na pré-escola

3.5. Análise dos dados

As análises passam por três etapas: I) análise descritiva, II) análise inferencial com uso de regressão hierárquica de dois níveis e III) apresentação da magnitude dos efeitos dos fatores associados à aprendizagem através de coeficientes de *effect-sizes*, possibilitando indicar a quantidade de meses de aprendizagem que cada fator, como nível socioeconômico e os estratos escolares, impactam no desenvolvimento cognitivo. Todos os resultados serão apresentados utilizando os dados imputados de Nível Socioeconômico (NSE), Ambiente de Aprendizagem em Casa (AAC) e indicador de frequência à creche. Os resultados dos modelos multiníveis com os dados observados (sem imputação) podem ser encontrados no anexo da dissertação.

Para a etapa de descrição dos dados serão apresentadas distribuições dos resultados de desenvolvimento cognitivo e aprendizagem conforme cada rede e estrato escolar, assim como para as variáveis explicativas de interesse. Para a terceira etapa serão apresentados modelos de análise de regressão hierárquica de dois níveis, sendo o primeiro nível com variáveis de aluno e, para o segundo nível, variável de nível socioeconômico médio da escola. A escolha pela modelagem multinível é feita para lidar com a natureza hierárquica natural dos dados educacionais (RAUDENBUSH, 1988). Essa natureza na educação significa que é preciso levar

em consideração os agrupamentos (*clusters*), isto é, o fato de que as crianças estão agrupadas em turmas, as turmas estão agrupadas em escolas, as escolas estão agrupadas em redes e as redes agrupadas em municípios, por exemplo. Ou seja, há uma hierarquia de relações que, a depender da análise, é preciso levar em conta uma vez que podem influenciar no resultado do desenvolvimento cognitivo da criança. Na prática o princípio da independência entre as medidas (no nosso caso crianças participantes do estudo) não estaria garantida e a incorporação dos *clusters* devem ser consideradas para permitir estimadores não enviesados. A modelagem hierárquica permite separar esses dois efeitos de forma mais adequada. Desse modo, sintetiza-se que o uso de modelos de regressão hierárquicos:

[...] permitem investigar a influência das características de cada nível da hierarquia no desempenho escolar dos alunos e na diferenciação entre as escolas e, ainda, separar a variabilidade nos resultados – associada às escolas – da variabilidade dentro de cada escola – associada aos alunos ou turmas de alunos. BONAMINO, 2013. P. 45.

A equação para elaboração do modelo hierárquico da presente dissertação pode ser verificada abaixo. Trata-se da equação apresentada por Bryk & Raundebush (1992), mas tendo apenas dois níveis, ao invés de três como originalmente expõem.

Nível 1:

$$Y_{ij} = \beta_{0j} + \sum_{f=1}^F \beta_{fij} X_{fij} + e_{ij}$$

Nível 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{0f} W_{sj} + \mu_{0j}$$

Y_{ij} representa o resultado da variável dependente da i -ésima criança matriculada na j -ésima escola. O termo X_{fij} indica a covariável (variável independente) f para o primeiro nível (criança) da i -ésima criança na j -ésima escola. W_{sj} indica s -ésima variável de nível dois da j -

ésima escola, podendo ser o nível socioeconômico da escola ou a variável indicadora de rede e estrato (pública e privada). Os coeficientes β_{fij} e γ_{0f} indicam os efeitos das covariáveis X_{fij} e NSE_j sobre a variável dependente. O termo μ_{0j} trata da incerteza relacionada ao coeficiente do nível socioeconômico médio (γ_{0f}) de cada escola (nível 2) e o termo e_{ij} indica a incerteza de todos os efeitos associados à variável dependente de cada criança, podendo ser, a depender da variável dependente, aprendizagem de matemática ou linguagem, ou, ainda, desempenho cognitivo em matemática ou linguagem para a 1ª e 2ª ondas.

Em um formato “menos” matemático, podemos apresentar como exemplo o modelo de regressão multinível da 1ª onda:

Modelo de nível 1:

$$Y = B0 + B1*(REDE) + B2*(IDADE) + B3*(SEXO) + B4*(CRECHE) + B5*(AAC_CRIANCA) + B6*(NSE_CRIANCA) + e$$

Sendo Y o resultado cognitivo de matemática ou linguagem para 1ª Onda.

Modelo de nível 2:

$$B0 = G00 + G01*(NSE_ESCOLA) + u0$$

A construção dos modelos será realizada através de técnica de *bottom-up*, partindo de um modelo nulo composto apenas pelo intercepto constante e adicionando cada covariável, gerando uma tabela aninhada que permite notar as mudanças nos coeficientes causadas pela adição de novas variáveis explicativas. Para evitar uma sobrecarga de modelos e facilitar a leitura, cada tabela de análise com as regressões multinível serão compostas de apenas cinco modelos, sendo o modelo nulo e os últimos quatro modelos mais completos. As tabelas de modelos aninhados contendo todas as covariáveis adicionadas por etapa podem ser vistas no anexo.

Por fim, os coeficientes dos modelos multiníveis controlados por todas as covariáveis serão apresentados em *effect size* (ES). Coe (2002) defende que o uso de medida em ES permite comparar de forma mais efetiva os resultados com outros contextos e com uma interpretação mais profunda ao estabelecer um coeficiente padronizado, indo além dos resultados obtidos pelo *p-value* ou dos coeficientes com escalas arbitrárias, como é o caso dos coeficientes em escala

Rasch utilizadas na presente dissertação. Enquanto o *p-value* aponta para uma diferença/efeito não aleatório – sem indicar o “tamanho” desse efeito –, o *effect size* (ES), por sua vez, permite apontar a magnitude desse efeito. A combinação do ES com o *p-value* auxiliam, então, a apontar tanto a existência do efeito quanto, também, a magnitude média do efeito.

Um outro elemento importante é que, uma vez que os coeficientes das regressões multiníveis são padronizados, o ES permite comparar diferentes estudos que se utilizam de instrumentos de medidas cognitivas distintas, tornando possível relacionar efeitos em diferentes contextos, tanto nacionais quanto internacionais. No cenário da pré-escola, medidas de ES produzidas através da padronização de coeficientes de regressões multiníveis foram utilizadas, por exemplo, na produção dos relatórios e estudos desenvolvidos pelo programa *Effective Pre-school and Primary Education* (EPPE), permitindo identificar, dentre diversos fatores, a proporção do impacto sobre o desenvolvimento infantil da ida à pré-escola e de estudar em período integral ou meio período (SAMMONS et al., 2008). Desse modo, reportar o *effect size* é de grande importância para contribuir com os estudos e esforços científicos na Educação, facilitando a interpretação dos resultados, contribuindo com as comparações e auxiliando em debates que balizam a produção e execução de políticas públicas e programas educacionais.

Para construir os coeficientes de *effect size* será utilizada a forma para estudos experimentais na educação proposto por Tymms (2004), realizada através de desvios padrão. São duas fórmulas que se diferenciam quanto à característica das variáveis a serem analisadas, separadas em variáveis categóricas ou contínuas.

Para as variáveis categorias em formato *dummy* (0 ou 1), a fórmula utilizada é:

$$\Delta = \frac{\beta_1}{\sigma_e}$$

Em que σ_e é o desvio padrão referente ao nível do aluno e β_1 o coeficiente de inclinação da variável categórica resultante da regressão multinível, como, por exemplo, o coeficiente obtido de ser escola de alta mensalidade, ser menino ou ter frequentado a creche.

E, para variáveis contínuas:

$$\Delta = \frac{2\beta_1}{\sigma_e}$$

Sendo que, neste caso em específico para as variáveis contínuas, β_1 deve ser calculado através de variável padronizada em escore-Z.

A interpretação dos coeficientes de ES será feita em meses de aprendizagem. Para isso será utilizada a classificação apresentada na Tabela 7 proposta por Higgins, Kokotsaki e Coe (2012) e Higgins et al (2014), que foi construída através de revisões sistemáticas e meta-análises de estudos experimentais em Educação realizadas na Inglaterra e internacionalmente. Essa prática de reportar resultados não em Effect Sizes, mas sim em meses de aprendizagem é comumente utilizada pela *Education Endowment Foundation*¹ (EEF), organização inglesa sem fins lucrativos que financia diversas avaliações experimentais e outras pesquisas sobre eficácia escolar que tem como principal objetivo identificar programas e práticas eficazes para impulsionar o aprendizado dos alunos. Apresentar os resultados em “meses de aprendizagem” pode ser vista como uma estratégia adicional para comunicar de forma mais eficaz com profissionais da educação e membros da sociedade civil resultados de estudos científicos que podem ajudar na formulação de política pública no campo da educação.

Tabela 7: Interpretação de Effect Size em meses de aprendizagem

Meses de aprendizagem	Effect Size entre:	Impacto do efeito
0	-0.01 e 0.01	Muito baixo ou inexistente
1	0.02 e 0.09	Baixo
2	0.10 e 0.18	Baixo
3	0.19 e 0.26	Moderado
4	0.27 e 0.35	Moderado
5	0.36 e 0.44	Moderado

¹ Para saber mais sobre a *Education Endowment Foundation* (EEF), acesse: <https://educationendowmentfoundation.org.uk/>

6	0.45 e 0.52	Alto
7	0.53 e 0.61	Alto
8	0.62 e 0.69	Alto
9	0.70 e 0.78	Muito alto
10	0.79 e 0.87	Muito alto
11	0.88 e 0.95	Muito alto
12	0.96 e >1.0	Muito alto

Fonte: Adaptado de Higgins et al (2014). Tradução livre.

Sendo assim, se um coeficiente de um modelo multivariado e padronizado em *effect size* apresentar um valor, por exemplo, de 0.33, é possível interpretar o dado como evidencia de impacto moderado e equivalente a 3 meses de aprendizagem escolar. É importante ressaltar que as variáveis categóricas detêm sempre um valor base de comparação. No caso da rede e dos estratos, a rede pública é a base. Para sexo, ser do sexo feminino e, para o indicador de ter frequentado à creche, a base de comparação é não ter frequentado a creche. Ou seja, se *hipoteticamente* a variável de sexo apresentasse um coeficiente ES de 0.10, indica-se que ser do sexo masculino tem o equivalente, na média e controlado por todas outras covariáveis, dois meses de aprendizagem a mais na variável dependente (matemática ou linguagem) em comparação a ser do sexo feminino, sendo um impacto considerado baixo.

As variáveis contínuas, por outro lado, são interpretadas levando em consideração a adição de desvios padrão (score-Z). De forma exemplificada, se *hipoteticamente* o coeficiente de nível socioeconômico (NSE) em ES for de 0.62, significa que, crianças que estão um desvio padrão (um score-Z) acima da média de NSE, detêm, na média e controlado por todas as covariáveis, o equivalente a oito meses de aprendizagem escolar a mais no desenvolvimento de matemática ou linguagem do que aquelas que estão na média de NSE, sendo considerado um impacto alto.

Serão apresentados dois agrupamentos de variáveis dependentes, tanto para regressão multinível quanto para *effect size*. O primeiro agrupamento de modelos utiliza como variável dependente os resultados cognitivos na 1ª onda em linguagem e matemática, possibilitando compreender o tamanho das desigualdades cognitivas para essas duas dimensões do

conhecimento na entrada da Educação Básica obrigatória, isto é, no início do ano letivo do 1º ano da pré-escola. Ressalta-se que, uma vez que as crianças estão iniciando a vida escolar, todos os resultados apenas indicam o tamanho da desigualdade e estratificação entre as redes pública e privada e não um impacto do tipo da instituição sobre o desenvolvimento das crianças.

O segundo agrupamento de modelos utiliza como variável dependente os resultados cognitivos de linguagem e matemática na 2ª onda, que é a medida de proficiência ao final do primeiro ano da pré-escola. Os modelos completos, com as covariáveis e a medida cognitiva do início do ano (1ª onda) como variável explicativa, são considerados modelos de valor agregado (*value-added*). Com isso, busca identificar os fatores que impactam a aprendizagem das crianças, possibilitando verificar possíveis efeito-escola e efeitos dos perfis escolares e familiares. Ressalta-se que, como as tabelas de regressão multinível serão apresentadas através de técnica *bottom-up*, será considerado valor agregado (*value-added*) apenas o último modelo de cada tabela com variável dependente da 2ª onda, que detém a medida cognitiva do início do ano -se a dependente for linguagem 2ª onda, então medida de linguagem 1ª onda como variável explicativa. Com isso, todos os outros modelos antes do último dessas tabelas são reconhecidos como modelos de *status*, que permitem identificar a estratificação e desigualdade no desempenho cognitivo ao final do ano, mas não aprendizagem, pois não detém uma medida de desempenho anterior, com exceção do último modelo da tabela.

No quadro 2 abaixo é apresentada um resumo da composição dos modelos completos de regressão multinível e *effect size*.

Quadro 2: Resumo das variáveis que compõem as análises multinível e *effect size*

Tipo de análise	Variável dependente	Variáveis explicativas	Interpretação
Transversal	Linguagem 1ª onda	<ul style="list-style-type: none"> ➤ Rede ou estrato* ➤ Idade em meses ➤ Sexo ➤ Indicador de frequência à creche ➤ Ambiente de Aprendizagem em Casa (AAC) ➤ Nível Socioeconômico (NSE) 	Indica a estratificação educacional medida através da desigualdade de desenvolvimento cognitivo no início da Educação Básica obrigatória (1º ano da pré-escola). Não apresenta qualquer impacto de efeito-escola.
	Matemática 1ª onda		

Valor agregado	Linguagem 2ª onda	<ul style="list-style-type: none"> ➤ Rede ou estrato* ➤ Idade em meses ➤ Sexo ➤ Indicador de frequência à creche ➤ Ambiente de Aprendizagem em Casa (AAC) 	Análise de efeito-escola, indicando quais fatores mais impactam a aprendizagem, possibilitando analisar o possível efeito agregado por estudar em diferentes redes escolares.
	Matemática 2ª onda	<ul style="list-style-type: none"> ➤ Nível Socioeconômico (NSE) ➤ NSE médio da escola ➤ Linguagem 1ª onda ➤ Matemática 1ª onda 	

* Serão gerados modelos que contém como variável explicativa a rede (pública e privada) e outros modelos que utilizam estrato (pública e privada de baixa, média e alta mensalidade).

Fonte: elaboração própria do autor.

4. RESULTADOS

As análises apresentadas nesse capítulo buscam evidenciar padrões de desigualdade no desenvolvimento cognitivo em matemática e linguagem no primeiro ano da pré-escola. As análises descritivas buscam comparar o ponto de partida e aprendizado ao longo do primeiro ano da escolarização obrigatória entre os alunos matriculados na rede pública e privada. É importante reforçar que os dados apresentados neste primeiro subcapítulo são meramente descritivos e não permitem estimar o impacto das diferentes redes de ensino no aprendizado. Os dados devem ser interpretados como uma evidência preliminar das diferenças observadas no desenvolvimento cognitivo entre alunos matriculados nas redes pública e privada e não uma evidência de eficácia das escolas ou redes.

Os dados coletados no início do ano letivo permitem estabelecer uma medida de base e descrever padrões de desigualdade educacional no momento inicial da escolarização obrigatória. Essa medida é chave para análises que pretendem investigar o efeito das escolas e/ou redes no aprendizado e na diminuição ou ampliação das desigualdades educacionais. A coleta de dados ao final do ano letivo permite, ao comparar com essa linha de base, realizar análises de valor agregado para identificar o impacto escolar na aprendizagem das crianças em idade pré-escolar.

Inicialmente serão apresentadas análises descritivas, buscando identificar padrões relacionados às variáveis de perfil discente e a rede escolar nos resultados cognitivos. Por fim, nos capítulos 4.2, 4.4 e 4.4, serão apresentadas as análises multivariadas e inferenciais no intuito de evidenciar impactos de perfis familiares na estratificação no início do ano escolar e na aprendizagem, assim como o impacto das diferentes redes escolares na aprendizagem cognitiva das crianças, explorando o como se comporta a produção de desigualdade escolar ao longo do início da Educação Básica obrigatória.

4.1. Descrição dos resultados para desenvolvimento em linguagem e matemática no início e final do ano letivo do pré-I (1ª e 2ª onda)

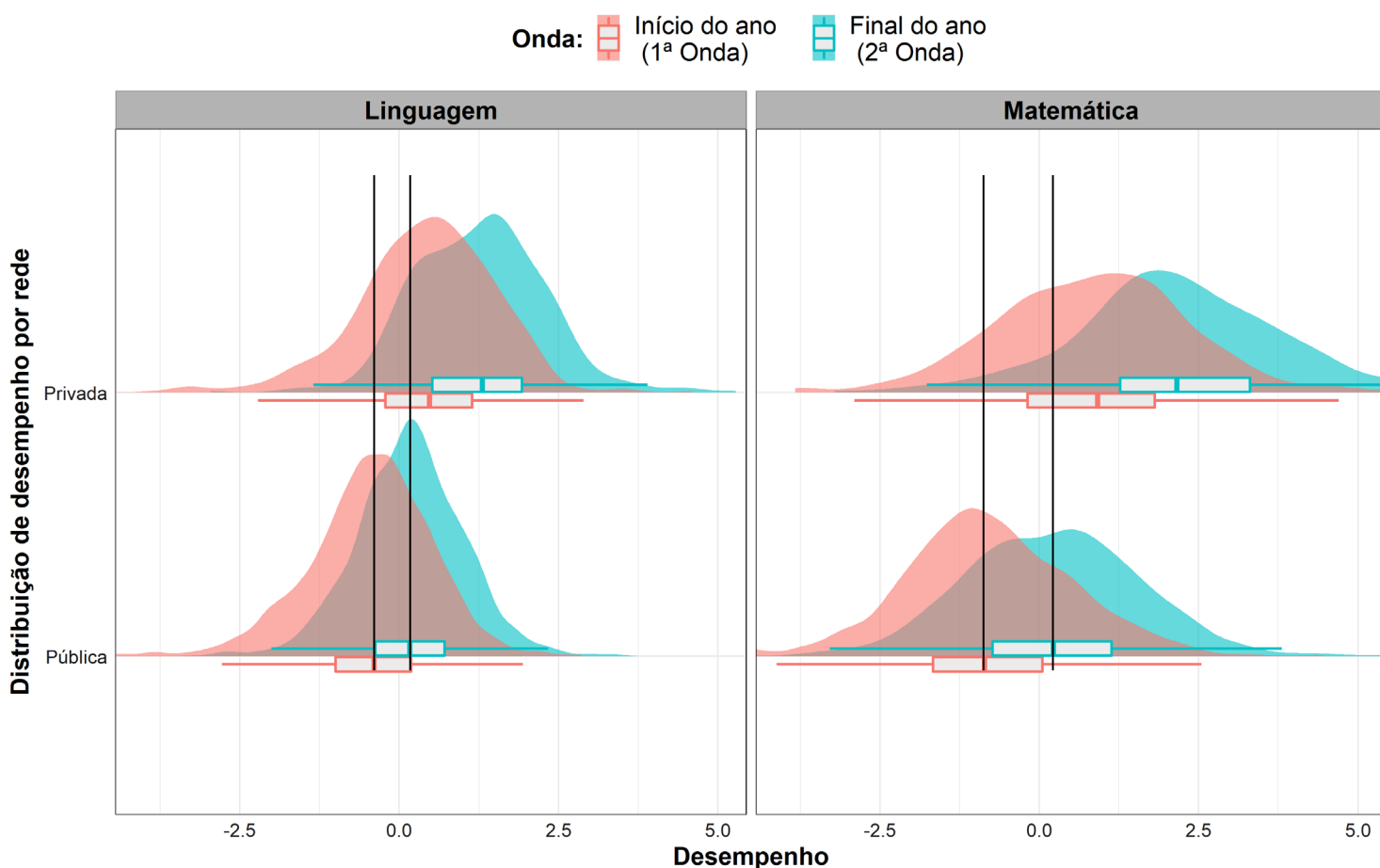
O estudo teve participação de 3262 crianças na 1ª onda. Dentre essas, 2710 estavam matriculadas em 46 escolas públicas e 552 crianças em 29 escolas privadas. As escolas privadas, aqui agrupadas sob a denominação de “rede privada”, subdividem-se em três estratos conforme a mensalidade, sendo baixa (até R\$400), média (R\$401 – R\$1000) e alta mensalidade (acima de R\$1000), em um total de, respectivamente, 5, 10 e 14 escolas. Para o final do ano (2ª onda) são consideradas nas análises apenas crianças que estavam presentes no início do ano, possibilitando, assim, análises longitudinais. Ao total, para a 2ª onda, participam do estudo 2712 crianças, indicando uma mortalidade amostral de 550 casos².

Os gráficos das figuras 5 e 6 auxiliam compreender a distribuição dos resultados observados no estudo. No quadro esquerdo de cada figura é apresentado a distribuição de desenvolvimento em linguagem e, no direito, para matemática. São utilizadas de gráficos de densidade e, logo abaixo de cada densidade, um *boxplot* que auxilia compreender a distribuição desses dados. As cores indicam a onda, sendo a cor roseada a distribuição de 1ª onda e, em azul, a distribuição do desenvolvimento cognitivo para a 2ª onda. O distanciamento entre as duas cores é indicativo da aprendizagem cognitiva entre o começo (1ª onda) e final do ano letivo (2ª onda). Quanto mais para a direita se apresenta a distribuição, maior tende a ser o desenvolvimento cognitivo do referido grupo. Duas linhas verticais em cor preta destacadas foram inseridas em cada quadrante, representando as medianas de resultados das crianças matriculadas em escolas

² Para mais informações sobre composição de amostra, mortalidade amostral e tratamento nos dados, ver capítulo de metodologia.

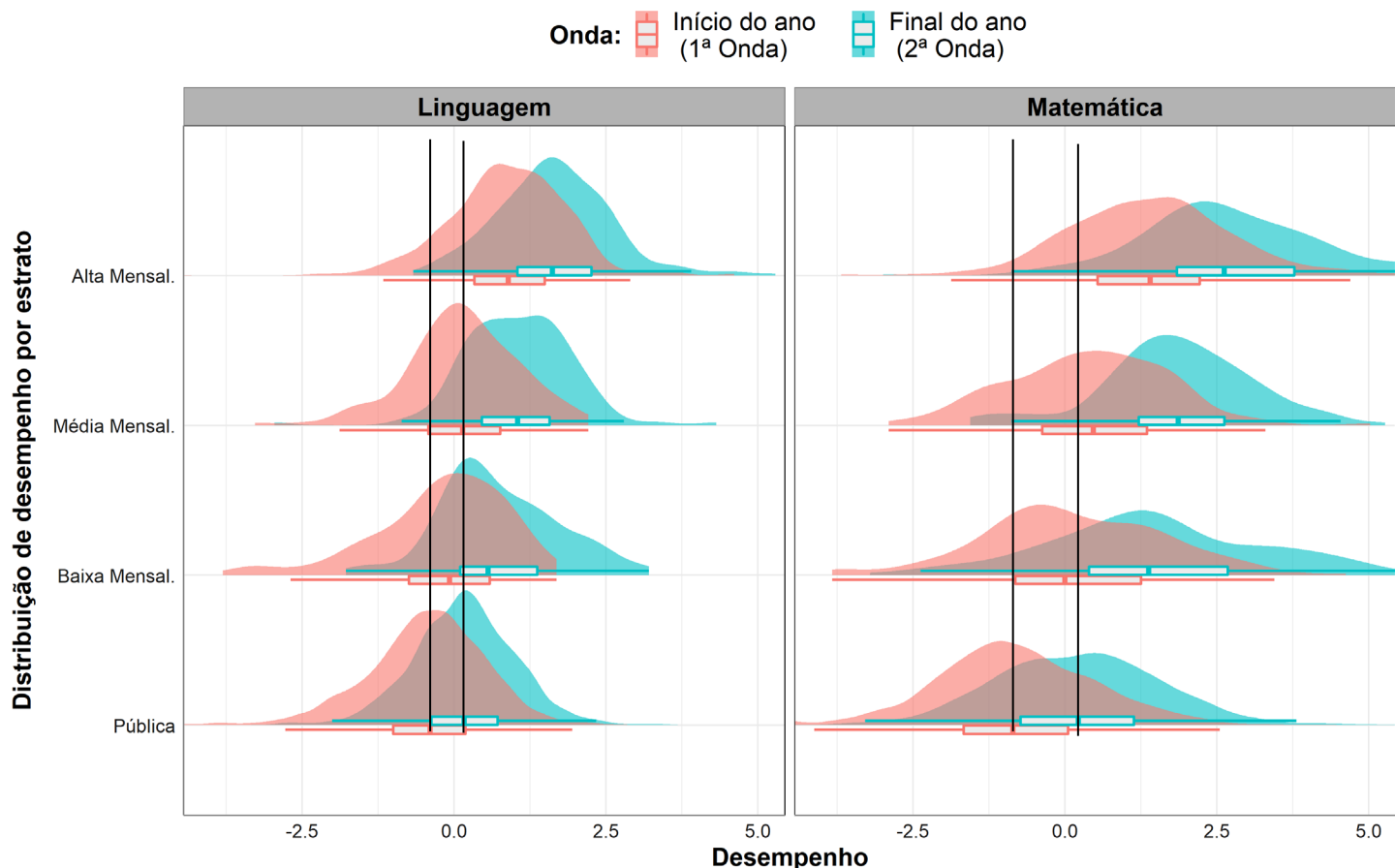
públicas, sendo a da esquerda a mediana de 1ª onda e, da direita, a mediana de 2ª onda. Essas linhas permitem uma melhor visualização das desigualdades observadas entre as redes. A escala para linguagem e matemática são diferentes e, por isso, a comparação entre elas deve ser apenas para se referir ao padrão de distribuição e não ao desenvolvimento cognitivo, uma vez que o valor 5 em matemática não tem o mesmo significado do que o valor 5 em linguagem. Ressalta-se que os valores no eixo X são em escala Rasch (logit) e não representam a quantidade de acertos que as crianças tiveram no instrumento PIPS.

Figura 5: Distribuição do desenvolvimento em Linguagem e Matemática por rede



Fonte: Estudo Longitudinal LaPOpE/UFRJ | Gráfico: Daniel Lopes de Castro (Autor)

Figura 6: Distribuição do desenvolvimento em Linguagem e Matemática por estrato



Fonte: Estudo Longitudinal LaPOPE/UFRJ | Gráfico: Daniel Lopes de Castro (Autor)

LaPOPE
UFRJ

Ambas as figuras apresentam resultados com um comportamento aproximado de distribuição normal em todos os grupos, tanto para redes quanto para estratos em linguagem e matemática. Os gráficos permitem visualizar possíveis estratificações entre as redes escolares ao observar o distanciamento da mediana da rede pública em comparação à distribuição de desenvolvimento da rede privada no início e no ao final do ano letivo.

Nos gráficos que focam a divisão entre as redes (figura 5), observando a linha vertical à esquerda em cada quadrante, que se refere à mediana das escolas públicas no início do ano (1ª onda), indica-se que metade das crianças das escolas públicas do estudo iniciam a Educação Básica obrigatória no pré-I com um desenvolvimento inferior ao quartil inferior das crianças matriculadas nas escolas privadas, tanto para linguagem quanto matemática. Uma visualização

ainda mais preocupante é que, ao finalizar o primeiro ano da pré-escola, metade das crianças matriculadas em escola pública do estudo detém desempenho inferior do que o foi observado no início do ano para mais da metade das crianças matriculadas em escolas privadas. De forma resumida, o comportamento nas amostras analisadas indica que ao menos metade das crianças das escolas públicas finalizam o primeiro ano da pré-escola sem saber os conhecimentos em linguagem e matemática que a maior parte das crianças de escolas privadas sabem logo que iniciam a vida escolar no primeiro ano da pré-escola.

Padrão semelhante é observado na figura 6, contudo as desigualdades aumentam à medida que avançamos nos três estratos da rede privada. Esse resultado indica estratificação que existe na rede privada, muito marcada pelo valor das mensalidades cobradas e perfil socioeconômico das famílias dos estudantes.

É importante ressaltar que as análises desses gráficos são meramente descritivas e não consideram as características dos alunos e suas famílias. Não há possibilidade, portanto, de falarmos sobre efeito e eficácia das escolas e redes no aprendizado. Os dados descritivos são importantes para ressaltar o tamanho das desigualdades observadas no início da pré-escola e como há uma clara segregação do sistema escolar que separa, desde o momento inicial da escolarização obrigatória, crianças com perfil socioeconômico diferentes e com estágios de desenvolvimento também diferentes. Considerando o ponto de partida apenas, os resultados coletados sugerem um Effect Size de 0,906 para linguagem e 1,259 para matemática comparando o ponto de partida das crianças matriculadas na rede pública e privada. Isso equivale a uma diferença inicial de 11 meses no desenvolvimento em linguagem e de 15 meses em matemática entre estudantes matriculados na rede pública e privada

. Observe que não estamos falando do impacto da pré-escola ou buscando explicar quais fatores estão associados a essa diferença observada. A ideia aqui é simplesmente destacar a diferença inicial mensurada. Ou seja, o que a evidência descritiva aponta é que, no contexto dessa metrópole do sudeste, crianças matriculadas em escolas privadas iniciam a vida escolar com conhecimento médio em linguagem equivalente a 11 meses de aprendizagem escolar a mais do que a média das crianças que iniciam sua vida em escola pública, além de 15 meses, em média, para matemática. Essa expressiva diferença é um provável fruto da origem social dos estudantes,

seus contextos familiares e vivências culturais, não do impacto da rede sobre a aprendizagem. Afinal, é uma diferença logo no início da vida escolar obrigatória e não haveria tempo para impacto proveniente da escola, ao menos não da pré-escola. Análises de fatores associados a esse cenário serão realizados no próximo capítulo.

Tabela 8: Effect Size (EF) em Linguagem e Matemática no início do ano (1ª onda) por rede e estrato

1ª Onda (Referência: rede pública)		
Variável	Linguagem	Matemática
Rede:Privada (Agregado)	0.906* (11 meses)	1.259* (15 meses)
Estrato:Baixa Mensalidade	0.284* (4 meses)	0.742* (9 meses)
Estrato:Média Mensalidade	0.586* (7 meses)	0.893* (11 meses)
Estrato:Alta Mensalidade	1.351* (17 meses)	1.691* (20 meses)

Fonte: Elaboração própria do autor – * $p < 0.05$

Compreender de forma mais profunda esse fenômeno no contexto brasileiro exige a elaboração de novas pesquisas, em especial aquelas que observam a primeiríssima infância. Ou seja, aqui não devemos desenhar estudos pensando o impacto da escola no desenvolvimento, mas sim das oportunidades no desenvolvimento desde o período gestacional, e durante a chamada os primeiros 3 anos de vida. Estamos falando de políticas intersetoriais que garantam o acesso da população ao sistema de saúde, saneamento básico, um ambiente de aprendizagem em casa rico, creche de qualidade, ambiente seguro e acolhedor nos espaços públicos das cidades, entre outros fatores que podem impactar de forma importante o desenvolvimento infantil. Os resultados descritivos da dissertação sugerem que políticas intersetoriais focadas na primeira infância devem ser pensadas como políticas públicas prioritárias para a diminuição das desigualdades educacionais.

Mesmo não havendo políticas que intencionalmente segreguem alunos considerando sua habilidade/conhecimento prévio ou perfil socioeconômico, na prática, os resultados reforçam o enorme impacto das desigualdades sociais no modelo de sistema educacional vigente no país.

Cabe reforçar que os dados apresentados foram coletados em uma grande cidade no Sudeste. Apesar das limitações das comparações entre as amostras da rede pública e privada, os dados descritivos apresentam evidências inéditas no Brasil sobre o tamanho das desigualdades educacionais, aqui medidas a partir do desenvolvimento cognitivo, no início da escolarização obrigatória.

Outro ponto que chama a atenção nas figuras 5 e 6 é que as diferenças para matemática parecem maiores do que as observadas para linguagem. Isso vale para a medida inicial no ano letivo e ao final. Estudos futuros devem investigar as diferenças descritas nessa dissertação buscando observar se há elementos relacionados a rotina das famílias (ambiente de aprendizagem em casa) (KOSLINSKI et al., 2021, PRELO), qualidade das creches ou currículo na pré-escola que ajudem a explicar os resultados preliminares descritos.

Abaixo, as tabelas 9 até 12 apresentam estatística descritiva (média, desvio padrão, mínimo e máximo) do desenvolvimento cognitivo considerando, os estratos, as características das crianças e das famílias. Essas tabelas apresentam valores brutos em escala *Rasch* e não devem ser confundidos com resultados reportados em *Effect Size*. As tabelas 9 e 10 apontam para os dados de início do ano (1ª onda), auxiliando a caracterizar a desigualdade cognitiva na entrada da pré-escola. As tabelas 11 e 12 apontam para a estratificação observada ao final do primeiro ano da pré-escola. Já a tabela 13³ descreve a distribuição dos valores de nível socioeconômico (NSE), ambiente de aprendizagem em casa (AAC) e indicador de frequência por cada estrato de escola.

Os dados descritivos não possibilitam indicar qualquer impacto das covariáveis sobre o desenvolvimento cognitivo das crianças. Contudo, essas estatísticas são importantes para compreendermos a distribuição dos dados e descrever padrões observados entre o desenvolvimento cognitivo e características das crianças, famílias e escolas.

³ Trata-se de um recorte das estatísticas descritivas com dados imputados da tabela 5 do capítulo de imputação de dados.

Tabela 9: Estatísticas descritivas para Matemática 1ª onda

Variável	Grupo	Média	DP	Mín	Max	N
Unificado	Total	-0,53	1,52	-6,10	6,08	3251
Rede	Pública	-0,81	1,37	-6,10	4,47	2700
	Privada	0,83	1,51	-3,83	6,08	551
Estrato	Pública	-0,81	1,37	-6,10	4,47	2700
	Baixa mensalidade	0,13	1,52	-3,83	4,62	140
	Média mensalidade	0,39	1,31	-2,90	5,02	138
	Alta mensalidade	1,42	1,37	-3,68	6,08	273
Idade em meses (Quartis)	1º quartil	-1,18	1,44	-6,10	4,06	788
	2º quartil	-0,66	1,41	-6,10	4,43	788
	3º quartil	-0,35	1,41	-5,92	5,02	787
	4º quartil	-0,04	1,56	-5,19	6,08	783
Sexo	Feminino	-0,52	1,52	-6,10	5,02	1550
	Masculino	-0,54	1,52	-6,07	6,08	1701
NSE (Quartis)	1º quartil	-1,10	1,32	-6,10	4,43	814
	2º quartil	-0,84	1,38	-6,07	4,33	811
	3º quartil	-0,43	1,44	-6,10	5,02	1014
	4º quartil	0,46	1,57	-4,26	6,08	612
AAC (Quartis)	1º quartil	-0,99	1,32	-6,10	4,33	813
	2º quartil	-0,76	1,41	-6,07	3,61	815
	3º quartil	-0,61	1,39	-6,10	4,56	811
	4º quartil	0,23	1,66	-5,92	6,08	812
Frequentou creche	Não frequentou	-0,50	1,59	-6,10	6,08	1463
	Frequentou	-0,55	1,47	-6,10	5,62	1788

Fonte: Elaboração própria do autor.

Tabela 10: Estatísticas descritivas para Linguagem 1ª onda

Variável	Grupo	Média	DP	Mín	Max	N
Unificado	Total	-0,31	1,07	-5,74	4,61	3261
Rede	Pública	-0,46	1,00	-5,74	2,85	2709
	Privada	0,41	1,11	-4,93	4,61	552
Estrato	Pública	-0,46	1,00	-5,74	2,85	2709
	Baixa mensalidade	-0,22	1,09	-3,81	1,69	140
	Média mensalidade	0,13	0,93	-3,28	2,21	138
	Alta mensalidade	0,87	1,00	-4,93	4,61	274
Idade em meses (Quartis)	1º quartil	-0,74	1,09	-5,74	2,11	792
	2º quartil	-0,39	1,00	-4,93	2,90	788
	3º quartil	-0,19	0,99	-5,71	2,79	790
	4º quartil	0,01	1,05	-4,36	4,61	786
Sexo	Feminino	-0,22	1,02	-4,93	2,90	1556
	Masculino	-0,40	1,11	-5,74	4,61	1705
NSE (Quartis)	1º quartil	-0,71	1,01	-5,71	2,79	815
	2º quartil	-0,49	0,99	-5,74	2,31	815
	3º quartil	-0,22	1,01	-5,56	2,85	1018
	4º quartil	0,29	1,06	-4,52	4,61	613
AAC (Quartis)	1º quartil	-0,60	0,99	-5,74	2,85	815
	2º quartil	-0,50	1,03	-5,71	2,48	815
	3º quartil	-0,35	0,98	-5,56	2,78	815
	4º quartil	0,19	1,11	-3,92	4,61	816
Frequentou creche	Não frequentou	-0,32	1,13	-5,74	4,61	1469
	Frequentou	-0,31	1,02	-5,56	3,62	1792

Fonte: Elaboração própria do autor.

Tabela 11: Estatísticas descritivas para Matemática 2ª onda

Variável	Grupo	Média	DP	Mín	Max	N
Unificado	Total	0,60	1,71	-6,14	15,88	2712
Rede	Pública	0,22	1,44	-6,14	15,88	2212
	Privada	2,28	1,78	-3,21	11,08	500
Estrato	Pública	0,22	1,44	-6,14	15,88	2212
	Baixa mensalidade	1,51	1,82	-3,21	7,00	129
	Média mensalidade	1,85	1,27	-1,56	5,27	122
	Alta mensalidade	2,89	1,76	-2,99	11,08	249
Idade em meses (Quartis)	1º quartil	-0,07	1,50	-6,14	4,90	671
	2º quartil	0,42	1,50	-4,57	6,09	666
	3º quartil	0,78	1,68	-3,75	15,88	662
	4º quartil	1,21	1,83	-3,21	11,08	642
Sexo	Feminino	0,60	1,61	-4,63	15,88	1276
	Masculino	0,61	1,79	-6,14	11,08	1436
NSE (Quartis)	1º quartil	-0,13	1,32	-4,57	4,29	650
	2º quartil	0,16	1,42	-6,14	6,30	668
	3º quartil	0,76	1,66	-3,67	15,88	857
	4º quartil	1,79	1,84	-3,21	11,08	537
AAC (Quartis)	1º quartil	-0,01	1,38	-4,57	7,00	655
	2º quartil	0,31	1,55	-4,63	7,84	688
	3º quartil	0,54	1,53	-3,67	6,31	651
	4º quartil	1,50	1,91	-6,14	15,88	718
Frequentou creche	Não frequentou	0,76	1,79	-4,63	11,08	1140
	Frequentou	0,49	1,63	-6,14	15,88	1572

Fonte: Elaboração própria do autor.

Tabela 12: Estatísticas descritivas para linguagem 2ª onda

Variável	Grupo	Média	DP	Mín	Max	N
Unificado	Total	0,37	1,00	-6,15	6,20	2712
Rede	Pública	0,17	0,88	-6,15	6,20	2212
	Privada	1,27	1,02	-2,95	5,29	500
Estrato	Pública	0,17	0,88	-6,15	6,20	2212
	Baixa mensalidade	0,74	0,99	-1,78	3,21	129
	Média mensalidade	1,05	0,90	-2,95	4,32	122
	Alta mensalidade	1,64	0,94	-0,67	5,29	249
Idade em meses (Quartis)	1º quartil	0,00	0,97	-6,15	4,32	671
	2º quartil	0,23	0,92	-2,95	3,90	666
	3º quartil	0,46	0,94	-2,82	6,20	662
	4º quartil	0,74	1,00	-2,78	5,29	642
Sexo	Feminino	0,44	0,96	-2,96	6,20	1276
	Masculino	0,31	1,03	-6,15	5,29	1436
NSE (Quartis)	1º quartil	-0,03	0,80	-2,81	3,03	650
	2º quartil	0,12	0,88	-6,15	3,39	668
	3º quartil	0,44	0,97	-3,23	6,20	857
	4º quartil	1,05	1,04	-2,81	5,29	537
AAC (Quartis)	1º quartil	0,02	0,81	-2,96	3,39	655
	2º quartil	0,19	0,90	-2,81	3,82	688
	3º quartil	0,38	0,90	-3,23	4,32	651
	4º quartil	0,86	1,13	-6,15	6,20	718
Frequentou creche	Não frequentou	0,46	1,05	-2,95	5,29	1140
	Frequentou	0,31	0,96	-6,15	6,20	1572

Fonte: Elaboração própria do autor.

Tabela 13: Estatísticas descritivas dos valores de AAC, NSE e frequência à creche por estrato escolar

Variável	Estrato	Média	Desvio padrão
Ambiente de Aprendizagem em Casa (AAC)	Público	-0,41	1,48
	Baixa Mensalidade	0,68	1,24
	Média mensalidade	1,01	1,19
	Alta mensalidade	1,66	1,06
	Total	-0,13	1,57
Nível Socioeconômico (NSE)	Público	-0,46	1,41
	Baixa Mensalidade	1,10	1,41
	Média mensalidade	1,46	1,30
	Alta mensalidade	2,08	1,30
	Total	-0,10	1,62
Frequência à Creche (1 = Frequentou; 0 = Não frequentou)	Público	0,59	0,49
	Baixa Mensalidade	0,32	0,47
	Média mensalidade	0,23	0,42
	Alta mensalidade	0,45	0,50
	Total	0,55	0,50

Fonte: Elaboração própria do autor.

As descrições dos dados apontam para uma consistente desigualdade observada no primeiro ano da pré-escola, assim como já observado pelos gráficos das figuras 5 e 6. Crianças de escolas públicas iniciam sua vida escolar com desenvolvimento em matemática e linguagem em torno de um desvio padrão abaixo da média das crianças matriculadas em escolas privadas que, como já vimos em resultados reportados em *Effect Size* (ver tabela 8), é uma diferença equivalente a 11 meses de aprendizagem para linguagem 15 meses de aprendizagem para matemática no agregado da rede privada.

Para a variável estrato, nota-se um progressivo distanciamento das médias em comparação à rede pública. Conforme a mensalidade aumenta, maior é a desigualdade observada entre o estrato e a rede pública. A tabela 9, que apresenta os resultados para desenvolvimento de matemática na 1ª onda, indica que o resultado médio para as escolas públicas é de -0.81, mas de 0.13, 0.39 e 1.42 para estrato de escolas privadas de baixa, média e alta mensalidade,

respectivamente. Se interpretarmos essa distância através do coeficiente em *Effect Size* (ver tabela 8), expressa em 0.74 para baixa, 0.89 para média e 1.69 para alta mensalidade, o conhecimento em matemática das crianças provenientes de famílias que matriculam seus filhos nesses estratos se distanciam das que matriculam na rede pública, em média, o equivalente a 9 meses (baixa mensalidade), 11 meses (média mensalidade) e 20 meses de aprendizagem (alta mensalidade). Para linguagem 1ª onda, na tabela 10, esse padrão crescente também é observado. Assim como interpretado em matemática, para linguagem as distâncias são equivalentes, em média, a 4 meses, 7 meses e 16 meses de aprendizagem para baixa, média e alta mensalidade, respectivamente, em comparação às crianças matriculadas na rede pública.

As mensalidades das escolas se comportam enquanto uma *proxy* de nível socioeconômico e os dados descritivos permitem observar uma indicação de estratificação educacional não apenas entre a rede pública e privada, mas, também, internamente à própria rede privada, reforçando uma direção que correlaciona os resultados educacionais com características sociais.

Essa direção de comportamento dos dados é vista também para nível socioeconômico (NSE) e ambiente de aprendizagem em casa (AAC). A cada crescimento de quartil, observa-se um aumento médio nos resultados de desenvolvimento cognitivo. Para os resultados da 1ª onda em matemática (ver tabela 9), os quartis de NSE têm médias progressivamente crescentes, sendo, -1.1, -0.84, -0.43 e 0.46, detendo comportamento de mesmo modo para AAC conforme aumenta-se o quartil, sendo -0.99, -0.76, -0.61 e 0.23. Os resultados para linguagem (tabela 10) indicam a mesma direção, e esse comportamento é visto, também, para os resultados de 2ª onda, isto é, ao final do ano letivo no pré-I (tabelas 11 e 12) tanto para matemática quanto linguagem. Desse modo, indica-se para uma possível persistência de estratificação educacional com origens provenientes dos fatores sociofamiliares.

Essa constante na direção que indica que, na média, quanto maior os fatores socioeconômicos (NSE) e práticas de aprendizagem no seio familiar (AAC) maior tenderá o desenvolvimento cognitivo em linguagem e matemática, acende um alerta. Os dados apresentam, como podem ser vistos na tabela 13, que escolas públicas tendem a concentrar grupos de crianças com menor nível socioeconômico (NSE) e ambiente de aprendizagem em casa (AAC) em comparação aos estratos de escolas privadas. Nesse sentido, se esses fatores estiverem

associados à aprendizagem nos modelos multivariados, ampliará ainda mais a importância das escolas públicas enquanto parte de uma rede de proteção contra efeitos perversos de vulnerabilidades sociais. A rede privada se beneficiaria, nessa lógica, por um efeito de ter matriculado em suas instituições crianças que, na média, são provenientes de famílias com NSE e AAC acima da média, enquanto a rede pública precisaria estar incluída em estratégias de políticas públicas ou, mesmo, lançar-se em práticas pedagógicas para contrabalancear a composição discente repleta de vulnerabilidades sociais.

Ao menos no que tange o desenvolvimento cognitivo em matemática e linguagem, não buscar contrabalancear esses efeitos através de práticas pedagógicas ou políticas públicas podem dificultar no estabelecimento do que Nelson do Valle Silva (2003) chama de Sistemas Educacionais abertos ou democráticos, isto é, sistemas educacionais efetivos em diminuir a influência de fatores de origem familiar nos resultados educacionais, produzindo maior equidade nos resultados educacionais. Em outras palavras, as análises que serão apresentadas nos capítulos subsequentes poderão ajudar a compreender melhor os padrões de desigualdades educacional e sua relação com a oferta educacional. Se nas análises multivariadas as variáveis de controle de origem social AAC e NSE estiverem associadas com a aprendizagem (modelo de valor agregado), será ainda mais relevante refletirmos sobre as implicações desses resultados para a qualidade e equidade dos sistemas educacionais no Brasil, em especial nos anos iniciais da escolarização obrigatória.

As tabelas apontam também para outros padrões de comportamento das variáveis para o desenvolvimento cognitivo em linguagem e matemática. Para idade, a cada aumento de quartil há um crescimento médio do desenvolvimento cognitivo, tanto para 1ª quanto 2ª onda em linguagem e matemática, indicando importância de característica biológica e maturação envolvida no desenvolvimento cognitivo. Esse fator pode tanto apontar que crianças alguns meses mais velhas aprendem mais do que crianças alguns meses mais novas, como também, na média, a vivência de alguns meses a mais de vida do que outras crianças antes de iniciar a pré-escola é o suficiente para que haja diferenças cognitivas no início da vida escolar. Nesse sentido, podemos interpretar esse padrão, caso haja diferenças estatisticamente significante entre essas médias, como uma evidência do chamado Período Sensível do desenvolvimento humano que ocorre entre os 0 e 5

anos de idade, que são momentos da vida com maior capacidade de modificação e maleabilidade do cérebro em resposta às experiências ambientais (SINGER, 1995; KOLB et al., 2013).

Isso torna o debate sobre qualidade da oferta Educação Infantil ainda mais urgente, uma vez que cultivar um ambiente com estímulos de qualidade para as crianças nos primeiros anos de vida pode potencializar o aprendizado futuro e garantir que as crianças desenvolvam todo o seu potencial (PEISNER-FEINBERG et al., 2001; SAMMONS et al., 2008; SYLVA, MELHUIS, & SAMMONS, 2010; SYLVA et al., 2006; TYMMS, JONES, ALBONE, & HENDERSON, 2009). Ainda que a dissertação se debruce sobre resultados de desenvolvimento em linguagem e matemática, a idade pré-escolar deve ser entendida como momento de grande importância para o desenvolvimento humano e o estabelecimento de vínculos com outros adultos e crianças fora do círculo familiar. Além disso, as experiências vividas durante a Educação Infantil devem trabalhar elementos de habilidades socioemocionais, físico-motor e de socialização.

No que tange o sexo, se feminino ou masculino, os dados apresentam uma diferença mais destacável quando se observa para o desenvolvimento de linguagem. Esse comportamento vai ao encontro dos achados no estudo Geres, que indicam que meninas tendem a ter um desempenho mais acentuado para Língua Portuguesa do que os meninos (BROOKE et al., 2014).

Por fim, as descritivas observadas para o indicador de frequência à creche apresentam uma direção constante, mas um tanto contraintuitiva. A princípio, sem considerar qualquer significância estatística que garantam a diferença das médias, crianças que frequentaram a creche tendem ter, em média, resultados de desenvolvimento menores do que as crianças que não foram à creche. Os resultados das análises multivariadas, após os diversos controles, podem apontar para outra direção de efeito da frequência à creche, porém.

A próxima etapa da análise de dados se apresenta nos modelos multivariados, buscando identificar fatores que explicam a desigualdade educacional no início da Educação básica obrigatória (1º ano da pré-escola), assim como quais fatores impactam na aprendizagem das crianças. Desse modo será possível investigar mais a fundo o comportamento da estratificação educacional, assim como sua mudança ao longo do início da Educação Básica obrigatória.

4.2. Fatores associados ao desenvolvimento cognitivo e desigualdade educacional no ingresso da pré-escola

Abaixo são apresentados os resultados das análises multivariadas. As tabelas são colocadas em pares para cada resultado de desenvolvimento cognitivo (matemática ou linguagem), sendo primeiramente observadas as análises dos modelos multinível e a tabela subsequente reportando os coeficientes em Effect Size. Sendo assim, as tabelas 14 e 15 apresentam análises para matemática e as tabelas 16 e 17 para linguagem.

As tabelas 14 e 16 apresentam os coeficientes do modelo multinível que estima o desenvolvimento das crianças no início do processo de escolarização obrigatória – início do ano letivo no 1º ano da pré-escola. O modelo hierárquico apresenta dois níveis: aluno e escola. As variáveis no nível 1 do modelo permitem identificar fatores extraescolares associados ao desenvolvimento das crianças no início do 1º ano da pré-escola. Além das informações sobre a idade e as características da família, incluímos a informação se a criança teve a oportunidade frequentar a creche entre zero e três anos. Não discriminamos a informação sobre o tempo que a criança frequentou a creche 1, 2 ou 3 anos.

Os coeficientes do nível dois indicam a rede a que as escolas pertencem, a saber, pública e privada. Nesse primeiro modelo não estratificamos a rede privada considerando o valor das mensalidades. O coeficiente estimado para as redes não deve ser interpretado como uma evidência de efeito escola no desenvolvimento cognitivo. Não há nenhuma garantia que os alunos da amostra tenham frequentado a creche nas mesmas escolas que estão matriculados na pré-escola. Nesse sentido, os coeficientes estimados no nível 2 do modelo devem ser interpretados como evidência de viés de seleção ou estratificação no início da escolarização obrigatória. Sabemos que a alocação dos alunos dentro de uma rede (por exemplo, dentro da rede pública) ou analisando a composição social das escolas entre diferentes redes de ensino não ocorre de forma aleatória (Bartholo; Costa; 2018). A alocação dos alunos é fortemente impactada pela segregação residencial, além de outros fatores, como a cobrança de mensalidade ou estratégias da burocracia educacional e das famílias para garantir o acesso às escolas desejadas. Todo esse processo de escolha e seleção do alunado pode gerar um sistema educacional altamente estratificado quando

pensamos o perfil dos alunos dentro da rede pública ou entre as redes de ensino. Os coeficientes das redes sugerem que no início da escolarização obrigatória já há grande seleção dos alunos, em especial quando comparamos a alocação entre a rede pública e privada de ensino.

Tabela 14: análise de estratificação no início do 1º ano da pré-escola por rede - Regressão multinível para desenvolvimento de matemática 1ª onda:

Variável dependente: Desenvolvimento em Matemática no início do 1º ano da pré-escola.					
	Modelo nulo	Modelo 1	Modelo 2	Modelo 3	Modelo 4
(Intercepto)	-0.21 (-0.43 0.00)	-0.81 *** (-0.94 -0.68)	-6.75 *** (-7.43 -6.07)	-6.76 *** (-7.43 -6.08)	-6.73 *** (-7.39 -6.06)
Escola:Privada	---	1.69 *** (1.45 1.92)	1.72 *** (1.50 1.94)	1.53 *** (1.32 1.74)	1.24 *** (1.03 1.45)
Idade (meses)	---	---	0.11 *** (0.10 0.12)	0.11 *** (0.10 0.12)	0.11 *** (0.10 0.12)
Sexo:Masculino	---	---	-0.01 (-0.10 0.07)	0.01 (-0.08 0.10)	0.00 (-0.09 0.09)
Creche:Frequentou	---	---	0.12 ** (0.03 0.22)	0.11 * (0.02 0.20)	0.10 * (0.01 0.20)
Indicador de Ambiente de Aprendizagem em Casa	---	---	---	0.12 *** (0.09 0.15)	0.06 *** (0.03 0.09)
Indicador de Nível socioeconômico (Indivíduo)	---	---	---	---	0.18 *** (0.14 0.21)
Random Effects					
σ^2	1.80	1.80	1.63	1.61	1.57
τ_{00}	0.82 _{ESC}	0.17 _{ESC}	0.13 _{ESC}	0.11 _{ESC}	0.08 _{ESC}
ICC	0.31	0.08	0.07	0.06	0.05
N	⁷⁵ ESC	⁷⁵ ESC	⁷⁵ ESC	⁷⁵ ESC	⁷⁵ ESC
Observations	3251	3251	3146	3146	3146
Marginal R ² / Conditional R ²	0.000 / 0.314	0.170 / 0.240	0.239 / 0.295	0.254 / 0.301	0.280 / 0.316

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Fonte: Elaboração própria do autor.

Tabela 15: análise de estratificação no início do 1º ano da pré-escola por rede – Effect Size para desenvolvimento de matemática 1ª onda

Modelo 4 da Tabela 13	Effect-Size	Equivalente em meses de aprendizagem
Escola:Privada	0.991*	12 meses
Idade meses (Z-escore)	0.644*	8 meses
Sexo:Masculino	0	-
Creche:Frequentou	0.083*	1 mês
AAC (Z-Escore)	0.147*	2 meses
NSE Indivíduo (Z-escore)	0.451*	6 meses

Fonte: Elaboração própria do autor. - * $p < 0.05$.

Tabela 16: análise de estratificação no início do 1º ano da pré-escola por rede - Regressão multinível para desenvolvimento de linguagem 1ª onda

Variável dependente: Desenvolvimento em Linguagem no início do 1º ano da pré-escola.					
	Modelo nulo	Modelo 1	Modelo 2	Modelo 3	Modelo 4
(Intercepto)	-0.14 * (-0.26 -0.01)	-0.45 *** (-0.54 -0.35)	-4.26 *** (-4.77 -3.76)	-4.27 *** (-4.77 -3.77)	-4.25 *** (-4.74 -3.76)
Escola: Privada	---	0.89 *** (0.72 1.06)	0.92 *** (0.75 1.08)	0.75 *** (0.60 0.91)	0.55 *** (0.39 0.70)
Idade (meses)	---	---	0.07 *** (0.06 0.08)	0.07 *** (0.06 0.08)	0.07 *** (0.06 0.08)
Sexo: Masculino	---	---	-0.16 *** (-0.23 -0.10)	-0.15 *** (-0.21 -0.08)	-0.15 *** (-0.21 -0.08)
Creche: Frequentou	---	---	0.09 ** (0.02 0.16)	0.08 * (0.01 0.15)	0.08 * (0.01 0.14)
Indicador de Ambiente de Aprendizagem em Casa	---	---	---	0.10 *** (0.07 0.12)	0.05 *** (0.03 0.08)
Indicador de Nível socioeconômico (Indivíduo)	---	---	---	---	0.13 *** (0.10 0.15)

Random Effects					
σ^2	0.97	0.97	0.89	0.88	0.86
τ_{00}	0.27 _{ESC}	0.09 _{ESC}	0.08 _{ESC}	0.06 _{ESC}	0.05 _{ESC}
ICC	0.22	0.08	0.08	0.07	0.05
N	75 _{ESC}	75 _{ESC}	75 _{ESC}	75 _{ESC}	75 _{ESC}
Observations	3261	3261	3156	3156	3156
Marginal R ² / Conditional R ²	0.000 / 0.216	0.096 / 0.169	0.161 / 0.227	0.181 / 0.235	0.208 / 0.251

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Fonte: Elaboração própria do autor.

Tabela 17: análise de estratificação no início do 1º ano da pré-escola por rede – Effect Size para desenvolvimento de linguagem 1ª onda

Modelo 4 da Tabela 15	Effect-Size	Equivalente em meses de aprendizagem
Escola: Privada	0.59*	7 meses
Idade meses (Z-escore)	0.572*	7 meses
Sexo: Masculino	-0.161*	-2 meses (em comparação ao feminino)
Creche: Frequentou	0.081*	1 mês
AAC (Z-Escore)	0.184*	2 meses
NSE Indivíduo (Z-escore)	0.45*	6 meses

Fonte: Elaboração própria do autor. - * $p < 0.05$

Os resultados dos coeficientes nas tabelas de 14 a 17 permitem uma primeira análise sobre fatores associados ao desenvolvimento das crianças no início da pré-escola em uma amostra diversificada de crianças matriculadas na rede pública e privada. Os coeficientes estimados no nível 1 do modelo trazem contribuições importantes para o debate sobre as desigualdades educacionais no início da escolarização obrigatória. O indicador de nível socioeconômico sugere um *effect size* de 0.45, tanto para matemática quanto para linguagem. A interpretação sugere que crianças com um desvio padrão abaixo da média para o indicador de NSE apresentam um ponto de partida mais baixo com uma diferença média de 5 a 6 meses de

aprendizado. O indicador de ambiente de aprendizagem em casa (AAC) apresenta coeficientes estatisticamente significativos com um *effect size* de 0.18 em linguagem e 0.15 em matemática. Comparando crianças com um desvio padrão abaixo da média, os dados sugerem uma diferença média de 2 a 3 meses de aprendizagem. Esses resultados são importantes e sugerem que, controlado pelo nível socioeconômico das famílias e outras características da criança e escolas, as rotinas e brincadeiras realizadas pelos adultos dentro de casa estão associados com o desenvolvimento cognitivo em linguagem e matemática. Os resultados corroboram achados preliminares publicados no Brasil e em outros países sobre a importância das rotinas estabelecidas dentro de casa e reforçam a necessidade de introduzir essa dimensão em novos estudos educacionais (BARTHOLO et al., 2020; SYLVA et al. 2010).

O efeito da maturação biológica é mensurado com a variável idade medida em meses. Essa é uma variável por vezes ausente dos estudos no campo da educação na educação infantil e nos anos iniciais do ensino fundamental. Os resultados do modelo multivariado sugerem que a idade medida em meses de vida tem correlação positiva forte com o desenvolvimento cognitivo, mesmo após o controle do nível socioeconômico e outras características da família e criança.

Meninos e meninas apresentam desenvolvimento muito semelhante para matemática no início da escolarização obrigatória. No entanto, para a medida sobre o desenvolvimento da linguagem, o modelo multivariado sugere que na média as meninas apresentam um resultado superior – equivalente a dois meses de aprendizado – com um *effect size* de 0.16.

A última variável no nível 1 do modelo indica se a criança teve a oportunidade de frequentar a creche. Idealmente, os pesquisadores gostariam de obter uma medida sobre a qualidade da oferta na creche e, também, o tempo de exposição da criança em meses de frequência. No entanto, os dados não existem e, portanto, a variável indica apenas se a criança frequentou ou não a creche.

Os resultados apontam para uma associação positiva e estatisticamente significativa entre frequência a creche e o desenvolvimento cognitivo no início da escolarização obrigatória. A associação parece um pouco mais forte (mas ainda fraca) para a medida de matemática – EF de 0.10 para matemática e 0.08 para linguagem. Isso corresponde a um ganho entre 1 e 2 meses de aprendizado para as crianças que frequentaram a creche, controlado pelo nível socioeconômico,

idade e todas as demais covariáveis do modelo. Os resultados apresentados são de um modelo transversal e devem ser interpretados com cautela. É desejável que outros estudos com desenho robusto (por exemplo, com dados longitudinais na creche) sejam realizados para analisar os efeitos no bem-estar e desenvolvimento das crianças. Independente das limitações do modelo, os dados apresentados nessa dissertação são superiores a outros estudos que buscaram estimar o impacto da frequência a creche no desenvolvimento cognitivo das crianças, uma vez que associação a frequência a creche com uma medida no momento de ingresso na pré-escola (4 anos) e não medidas de avaliações externas realizadas anos depois no ensino fundamental (BARROS, 2011; NCPI, 2014)

O nível 2 do modelo apresenta apenas as variáveis que indicam a rede a que a escola pertence: pública ou privada. Os resultados sugerem que há forte estratificação no sistema escolar quando consideramos o estágio inicial do desenvolvimento cognitivo no início da escolarização obrigatória. Os resultados corroboram estudos anteriores sobre segregação escolar realizados em capitais no Brasil com estudantes matriculadas na rede pública e privada (Bartholo; Costa, 2018). Não obstante, os dados da dissertação são pioneiros porque analisam a etapa da pré-escola e reforçam a estratificação no sistema não apenas considerando características da composição social do alunado, mas também o seu desenvolvimento cognitivo.

O coeficiente que compara a alocação dos alunos nas escolas da rede pública e privada sugere que, no início da escolarização obrigatória, há uma grande diferença no desenvolvimento cognitivo das crianças – effect size de 0.59 ou 7 meses a mais de aprendizado em linguagem e 0.99 effect size ou 12 meses a mais de aprendizado em matemática para as crianças da rede privada. Essa diferença pode ser explicada pelas oportunidades que as crianças tiveram na chamada primeiríssima infância (0 a 3 anos). Estudos na área da saúde pública e educação sugerem que programas de acesso a pré-natal e sistema de saúde, saneamento básico, os estímulos recebidos pelo bebê recém-nascido em casa, acesso a uma creche de qualidade, entre outros fatores, têm impacto positivo no desenvolvimento infantil (ABUCHAIM, 2016).

Os resultados apresentados devem ser interpretados como uma evidência de estratificação no sistema escolar e não como impacto das escolas ou redes de ensino no desenvolvimento cognitivo. Esse é o primeiro estudo em larga escala no Brasil que mediu as desigualdades educacionais no aprendizado no início da pré-escola – crianças com aproximadamente 4 anos –

em uma cidade de grande porte. Os dados reforçam a grande estratificação do sistema educacional observada em estudos semelhantes realizados com amostras no ensino fundamental (DALBEN, A., 2014; BROOKE et al, 2014).

Tabela 18: análise de estratificação no início do 1º ano da pré-escola por estrato - Regressão multinível para desenvolvimento de matemática 1ª onda

Variável dependente: Desenvolvimento em Matemática no início do 1º ano da pré-escola.					
	Modelo nulo	Modelo 1	Modelo 2	Modelo 3	Modelo 4
(Intercepto)	-0.21 (-0.43 0.00)	-0.81 *** (-0.91 -0.71)	-6.68 *** (-7.36 -6.01)	-6.69 *** (-7.36 -6.02)	-6.66 *** (-7.32 -6.00)
Escola:Baixa mens.	---	0.99 *** (0.62 1.36)	0.99 *** (0.64 1.34)	0.87 *** (0.53 1.20)	0.65 *** (0.33 0.97)
Escola:Média mens.	---	1.20 *** (0.88 1.51)	1.33 *** (1.03 1.63)	1.17 *** (0.87 1.46)	0.92 *** (0.64 1.21)
Escola:Alta mens.	---	2.27 *** (2.01 2.52)	2.23 *** (1.99 2.47)	2.00 *** (1.76 2.23)	1.67 *** (1.44 1.90)
Idade (meses)	---	---	0.11 *** (0.09 0.12)	0.11 *** (0.10 0.12)	0.11 *** (0.10 0.12)
Sexo:Masculino	---	---	-0.02 (-0.11 0.07)	0.00 (-0.09 0.09)	-0.00 (-0.09 0.09)
Creche:Frequentou	---	---	0.11 * (0.02 0.21)	0.10 * (0.01 0.20)	0.09 * (0.00 0.19)
Indicador de Ambiente de Aprendizagem em Casa	---	---	---	0.11 *** (0.08 0.15)	0.06 *** (0.02 0.09)
Indicador de Nível socioeconômico (Indivíduo)	---	---	---	---	0.17 *** (0.14 0.21)
Random Effects					
σ^2	1.80	1.79	1.63	1.61	1.56
τ^2_{00}	0.82 _{ESC}	0.08 _{ESC}	0.06 _{ESC}	0.05 _{ESC}	0.04 _{ESC}
ICC	0.31	0.04	0.04	0.03	0.03
N	⁷⁵ ESC	⁷⁵ ESC	⁷⁵ ESC	⁷⁵ ESC	⁷⁵ ESC
Observations	3251	3251	3146	3146	3146
Marginal R ² / Conditional R ²	0.000 / 0.314	0.195 / 0.230	0.262 / 0.291	0.276 / 0.299	0.298 / 0.318

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Fonte: Elaboração própria do autor.

Tabela 19: análise de estratificação no início do 1º ano da pré-escola por estrato – Effect Size para desenvolvimento de matemática 1ª onda

Modelo 4 da Tabela 17	Effect-Size	Equivalente em meses de aprendizagem
Escola:Baixa mensalidade	0.523*	6 meses
Escola:Média mensalidade	0.739*	9 meses
Escola:Alta mensalidade	1.336*	16 meses
Idade meses (Z-escore)	0.638*	8 meses
Sexo:Masculino	-0.001	0 meses
Creche:Frequentou	0.076*	1 mês
AAC (Z-Escore)	0.14*	2 meses
NSE Indivíduo (Z-escore)	0.442*	5 meses

Fonte: Elaboração própria do autor. - * $p < 0.05$

Tabela 20: análise de estratificação no início do 1º ano da pré-escola por estrato - Regressão multinível para desenvolvimento de linguagem 1ª onda

Variável dependente: Desenvolvimento em Linguagem no início do 1º ano da pré-escola.					
	Modelo nulo	Modelo 1	Modelo 2	Modelo 3	Modelo 4
(Intercepto)	-0.14 * (-0.26 -0.01)	-0.45 *** (-0.51 -0.38)	-4.21 *** (-4.71 -3.72)	-4.22 *** (-4.71 -3.72)	-4.20 *** (-4.69 -3.72)
Escola:Baixa mens.	---	0.28 * (0.03 0.53)	0.24 * (0.00 0.49)	0.14 (-0.09 0.38)	-0.01 (-0.23 0.21)
Escola:Média mens.	---	0.58 *** (0.36 0.79)	0.66 *** (0.44 0.87)	0.53 *** (0.32 0.73)	0.35 *** (0.15 0.55)
Escola:Alta mens.	---	1.33 *** (1.16 1.50)	1.33 *** (1.16 1.49)	1.13 *** (0.97 1.30)	0.89 *** (0.73 1.06)
Idade (meses)	---	---	0.07 *** (0.06 0.08)	0.07 *** (0.06 0.08)	0.07 *** (0.06 0.08)
Sexo:Masculino	---	---	-0.16 *** (-0.23 -0.10)	-0.15 *** (-0.21 -0.08)	-0.15 *** (-0.21 -0.08)
Creche:Frequentou	---	---	0.08 * (0.00 0.16)	0.07 * (0.00 0.14)	0.07 (0.00 0.14)

			(0.02 0.15)	(0.01 0.14)	(-0.00 0.13)
Indicador de Ambiente de Aprendizagem em Casa	---	---	---	0.09 *** (0.07 0.12)	0.05 *** (0.03 0.08)
Indicador de Nível socioeconômico (Indivíduo)	---	---	---	---	0.13 *** (0.10 0.15)
Random Effects					
σ^2	0.97	0.97	0.89	0.88	0.85
τ^2_{00}	0.27 _{ESC}	0.03 _{ESC}	0.03 _{ESC}	0.02 _{ESC}	0.02 _{ESC}
ICC	0.22	0.03	0.03	0.03	0.02
N	⁷⁵ ESC	⁷⁵ ESC	⁷⁵ ESC	⁷⁵ ESC	⁷⁵ ESC
Observations	3261	3261	3156	3156	3156
Marginal R ² / Conditional R ²	0.000 / 0.216	0.126 / 0.154	0.193 / 0.218	0.211 / 0.231	0.236 / 0.251

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Fonte: Elaboração própria do autor.

Tabela 21: análise de estratificação no início do 1º ano da pré-escola por estrato – Effect Size para desenvolvimento de linguagem 1ª onda

	Effect-Size	Equivalente em meses de aprendizagem
Escola:Baixa mensalidade	-0.014	-
Escola:Média mensalidade	0.376*	5 meses
Escola:Alta mensalidade	0.967*	12 meses
Idade meses (Z-escore)	0.566*	7 meses
Sexo:Masculino	-0.161*	-2 meses (em comparação ao feminino)
Creche:Frequentou	0.073	-
AAC (Z-Escore)	0.176*	2 meses
NSE Indivíduo (Z-escore)	0.439*	5 meses

Fonte: Elaboração própria do autor. - * $p < 0.05$

As Tabelas 18 a 21 apresentam modelos semelhantes aos descritos nas Tabelas 14 a 17, porém com a diferença de que a variável de rede (pública e privada) foi substituída por variável de estrado, isto é, considerando o valor da mensalidade cobrada para as escolas privadas— baixa, média e alta mensalidade. A categoria de base mais uma vez são as escolas da rede pública de ensino. Os coeficientes estimados para os estratos reforçam os resultados descritos nas Tabelas 14 a 17, no entanto, evidenciam a grande estratificação dentro da rede privada e trazem novas evidências sobre o processo de estratificação no início da escolarização obrigatória.

Os coeficientes do nível 1 do modelo são muito semelhantes aos descritos nas tabelas que utilizam variável de rede e não serão discutidos aqui. Iremos concentrar as análises nos coeficientes do nível 2 do modelo, que destacam os estratos das escolas privadas tendo a rede pública de ensino como categoria de base. Os resultados que estimam o desenvolvimento em linguagem no início da escolarização obrigatória (ver Tabela 20) sugerem que não há diferenças entre o estrato das escolas privadas de baixa mensalidade e a rede pública, após a introdução das covariáveis de controle. O Modelo 1 apresenta apenas as variáveis do nível 2, sem os controles do nível 1, e sugere uma diferença entre o estrato das escolas privadas de baixa mensalidade e a rede pública – effect size de 0.28 para o desenvolvimento em linguagem (ver Tabela 8). No entanto, no modelo 4, com todas as variáveis de controle, o coeficiente que estima o viés de seleção no início da escolarização obrigatória entre a rede pública e as escolas privadas de baixa mensalidade sugere que não há diferenças entre os dois grupos.

Os coeficientes que estimam o viés de seleção entre as escolas privadas de média mensalidade e a rede pública sugerem que há diferenças estatisticamente significativas entre os dois grupos no momento inicial da escolarização obrigatória. Controlado por todas as covariáveis do modelo, os resultados apresentados na Tabela 21 sugerem um effect size de 0.38 para as escolas privadas de média mensalidade, o que pode ser traduzido como uma diferença média de 5 meses de aprendizado. Os resultados apresentados corroboram a hipótese de que há um viés de seleção baseado não apenas no perfil socioeconômico das famílias, mas também no estágio inicial de desenvolvimento das crianças quando comparamos a rede pública de ensino com as escolas privadas classificadas como média mensalidade.

Por fim, o último coeficiente no nível 2 do modelo identifica as escolas privadas de alta mensalidade e compara esse grupo com a rede pública de ensino. Os resultados evidenciam uma estratificação não apenas entre a rede pública e privada de ensino, mas também dentro da rede privada corroborando uma das hipóteses do estudo que ressaltam a grande estratificação dentro da própria rede privada. Isso cria um novo objeto de estudo dentro dos estudos sobre estratificação, segregação e desigualdades educacionais que devem considerar não apenas a distinção rede pública e privada, mas também diferenças existentes dentro da rede privada. As desigualdades dentro da rede pública de ensino tem sido objeto de diferentes grupos de pesquisa no Brasil e os resultados preliminares da dissertação evidenciam a necessidade de mais pesquisas com escolas privadas que incorporem a grande diversidade dentro dessa rede. O uso do valor das mensalidades como uma proxy da composição social dos alunos parece fazer sentido e são corroboradas pelos resultados preliminares.

Controlado pelas demais covariáveis do modelo, os resultados do Modelo 4 para linguagem sugerem que há um claro viés de seleção dos alunos matriculados nas escolas privadas de alta mensalidade comparado com a rede pública. Os coeficientes sugerem um *effect size* de 0.97 ou aproximadamente 12 meses de diferença no aprendizado em linguagem no momento inicial do processo obrigatória de escolarização. Essa é uma diferença muito grande e reforça as imensas desigualdades sociais e educacionais do país. É importante destacar mais uma vez que os modelos apresentados até o momento não analisam o aprendizado dos alunos, mas sim buscam identificar fatores associados ao desenvolvimento das crianças no início da pré-escola. Os coeficientes do nível 2 do modelo devem ser interpretados como uma evidência de estratificação ou viés de seleção no início da escolarização obrigatória no Brasil, não como um impacto do tipo de rede no desenvolvimento cognitivo infantil. Esses impactos serão vistos a partir do próximo capítulo através de análises multivariadas de modelagem de valor agregado.

Os resultados estimando a medida de 1ª onda de matemática (ver Tabelas 18 e 19), no início da escolarização obrigatória, apresentam um padrão semelhante ao descrito anteriormente para os resultados de linguagem, no entanto, os dados sugerem uma estratificação ainda maior para essa dimensão. O coeficiente que compara os alunos matriculados nas escolas privadas de baixa mensalidade com os pares na rede pública sugere que há viés de seleção dos alunos. Esse resultado é diferente do que foi observado para a medida no desenvolvimento de linguagem.

Estudos futuros devem buscar explicar esse resultado. Uma hipótese é que famílias que nível socioeconômico mais alto apresentam na média um indicador de ambiente de aprendizagem em casa maior. Alguns itens do indicador das rotinas dentro de casa observam jogos que envolvam a contagem objetos, números ou figuras geométricas. Outra hipótese pode ser uma exposição à creche de maior qualidade desse grupo, o que poderia proporcionar vivências com conteúdos relacionados à matemática. Por exemplo, ideias sobre matemática (conceitos como maior e menor, mais e menos), brincadeiras com números, blocos de montar e figuras geométricas.

O coeficiente estimado para as escolas de baixa mensalidade indica um effect size de 0.52 ou entre 6 e 7 meses de diferença no aprendizado entre a rede pública e privada de baixa mensalidade. O coeficiente que estimou as escolas de mensalidade média sugere um effect size de 0.74 ou até 9 meses de diferença na média entre as crianças das escolas da rede privada de média mensalidade e os alunos matriculados na rede pública de ensino. Os números são ainda maiores para as escolas privadas de alta mensalidade. Após o controle de todas as covariáveis, o Modelo 4 sugere um effect size de 1.34 ou 16 meses de diferença no aprendizado entre a rede pública e as escolas privadas de alta mensalidade. Isso é mais do que um ano de aprendizado e uma diferença muito grande, ainda mais considerando que estamos analisando o início do processo de escolarização obrigatório. Trata-se, portanto, de uma evidência clara da forte estratificação do sistema de educação básica e das enormes desigualdades sociais e educacionais observadas no país.

Os resultados apresentados devem ser interpretados com cautela devido as limitações da amostra realizada nas escolas privadas de ensino. Conforme discutido no capítulo de métodos, a amostra das escolas privadas foi intencional e há uma ameaça a validade externa dos resultados apresentados para a rede privada de ensino na cidade objeto do estudo. Diante dessa limitação, não é possível afirmar que as diferenças descritas nesse estudo correspondem as diferenças observadas na população de alunos matriculadas nas redes pública e privada. Não obstante, esses são os primeiros resultados reportados sobre estratificação escolar no início da pré-escola que apresenta o desempenho cognitivo como uma das variáveis chave da análise. Os dados reforçam a compreensão de que temos um sistema educacional com altos níveis de segregação escolar considerando não apenas a composição social das escolas, mas também o estágio de desenvolvimento cognitivo dos alunos no momento inicial da escolarização obrigatória.

Os modelos multivariados reforçam a ideia de que há um grande viés de seleção e estratificação no momento inicial do processo obrigatório de escolarização. Alunos com nível socioeconômico e ponto de partida no seu desenvolvimento também distintos tendem a frequentar escolas diferentes quando analisamos os estratos. Os coeficientes estimados para as escolas de alta mensalidade reforçam a tendência de um padrão que segrega em redes diferentes (pública e privada) crianças que na primeiríssima infância (período de zero a três anos) tiveram mais chance de ter uma infância saudável e com estímulos adequados. Isso pode ser exemplificado com mais acesso a saneamento básico, acesso ao sistema de saúde, creche de alta qualidade, ambientes de aprendizagem em casa mais ricos, entre outros fatores. A grande desigualdade social observada e documentada em diferentes estudos no Brasil parece marcar de forma clara a trajetória escolar e as oportunidades de aprendizagem desde o momento inicial de entrada no Ensino Básico obrigatório.

4.3. Fatores associados ao aprendizado em linguagem no primeiro ano da pré-escola

Este capítulo e o próximo (capítulo 4.4), apresentam os dados dos modelos hierárquicos para identificar fatores associados ao aprendizado em linguagem e matemática no primeiro ano da escolarização obrigatória. A principal questão é: há diferenças no aprendizado quando comparamos os alunos da rede pública e privada de ensino? No capítulo anterior, foi possível identificar fatores associados ao desenvolvimento no início da pré-escola e analisar se havia evidência de viés de seleção ou estratificação no início da escolarização obrigatória. Nestes dois capítulos serão apresentados modelos hierárquicos de dois níveis (indivíduo e escola) com dados longitudinais sobre o aprendizado no primeiro ano na pré-escola, tendo a medida da 2ª onda (final do ano) como variável dependente e a medida de 1ª onda (início do ano) sobre o desenvolvimento como principal preditor do modelo.

Tabela 22: análise de aprendizagem no 1º ano da pré-escola por rede - Regressão multinível para desenvolvimento de linguagem 2ª onda

	Modelo Nulo	Modelo 1	Modelo 2	Modelo 3	Modelo 4
(Intercepto)	0.55 *** (0.40 0.70)	0.15 ** (0.05 0.25)	-3.36 *** (-3.83 -2.89)	-0.97 *** (-1.39 -0.56)	-0.20 (-0.56 0.16)
Escola:Privada	---	1.14 *** (0.97 1.31)	0.25 (-0.02 0.53)	0.08 (-0.11 0.27)	0.18 * (0.01 0.36)
Idade (meses)	---	---	0.07 *** (0.06 0.08)	0.03 *** (0.02 0.04)	0.01 *** (0.01 0.02)
Sexo:Masculino	---	---	-0.10 ** (-0.16 -0.04)	-0.08 ** (-0.14 -0.03)	-0.02 (-0.06 0.03)
Creche:Frequentou	---	---	-0.00 (-0.07 0.06)	-0.03 (-0.08 0.03)	-0.03 (-0.08 0.01)
Indicador de Ambiente de Aprendizagem em Casa	---	---	0.06 *** (0.04 0.08)	0.04 *** (0.02 0.06)	0.02 * (0.00 0.04)
Indicador de Nível socioeconômico (Indivíduo)	---	---	0.10 *** (0.07 0.12)	0.04 *** (0.02 0.06)	0.01 (-0.01 0.03)
Indicador de Nível socioeconômico (Escola)	---	---	0.27 *** (0.15 0.39)	0.14 *** (0.06 0.22)	0.09 * (0.02 0.17)
Desempenho em matemática início do ano	---	---	---	0.36 *** (0.34 0.38)	0.18 *** (0.16 0.20)
Desempenho em linguagem início do ano	---	---	---	---	0.45 *** (0.42 0.48)
Random Effects					
σ^2	0.74	0.74	0.65	0.45	0.33
τ_{00}	0.40 _{ESC}	0.09 _{ESC}	0.03 _{ESC}	0.01 _{ESC}	0.01 _{ESC}
ICC	0.35	0.11	0.05	0.02	0.03
N	⁷⁵ ESC	⁷⁵ ESC	⁷⁵ ESC	⁷⁵ ESC	⁷⁵ ESC
Observations	2712	2712	2641	2632	2631

	Modelo Nulo	Modelo 1	Modelo 2	Modelo 3	Modelo 4
Marginal R ² / Conditional R ²	0.000 / 0.348	0.189 / 0.280	0.322 / 0.355	0.533 / 0.544	0.655 / 0.666

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Fonte: Elaboração própria do autor.

Tabela 23: análise de aprendizagem no 1º ano da pré-escola por rede – Effect Size para desenvolvimento de linguagem 2ª onda

	Effect-Size	Equivalente em meses de aprendizagem
Escola:Privada	0.314*	4 meses
Idade meses (Z-escore)	0.184*	3 meses
Sexo:Masculino	-0.033	-
Creche:Frequentou	-0.057	-
AAC (Z-Escore)	0.102*	2 meses
NSE Indivíduo (Z-escore)	0.059	1 mês
Indicador de Nível socioeconômico (Escola)	0.293*	4 meses
Desempenho em matemática início do ano	0.942*	11 meses
Desempenho em linguagem início do ano	1.681*	18 meses

Fonte: Elaboração própria do autor. - * $p < 0.05$

A Tabela 22 apresenta os resultados do modelo multinível que estima o desenvolvimento da criança em linguagem no final do primeiro ano da escolarização obrigatória. Um primeiro ponto que merece atenção é o ICC (Intraclass Correlation) observado no modelo nulo. A partição da variância mostra que aproximadamente 40% da variância em linguagem e matemática observada é explicada pelo nível 2 do modelo – escola. Esses são valores bem maiores do que o reportado anteriormente em outros estudos utilizando o mesmo instrumento para outras amostras

utilizando apenas dados da rede pública (KOSLINSKI; BARTHOLO, 2019; BARTHOLO et al., 2020).

O Modelo 2 introduz apenas os coeficientes referentes ao tipo de rede das escolas e os modelos subsequentes introduzem outras covariáveis de controle. É possível observar que os coeficientes da rede perdem força à medida que mais covariáveis de controle são introduzidas, em especial a medida inicial sobre o desenvolvimento das crianças. Esse é um resultado esperado uma vez que as análises do capítulo anteriores e os dados descritivos evidenciaram forte estratificação quando comparamos a rede privada e pública, não apenas em no perfil socioeconômico dos alunos, mas também no ponto inicial do desenvolvimento dos alunos no início da escolarização obrigatória.

Os dados sobre o aprendizado no primeiro ano da escola em linguagem sugerem que meninos e meninas aprendem em ritmo semelhante. O coeficiente negativo sugere que as meninas apresentam leve vantagem, mas o valor é muito pequeno e não apresenta significância estatística. O indicador de nível socioeconômico no nível 1 (aluno) também apresenta um coeficiente pequeno e sem significância estatística. Esse resultado pode parecer contraintuitivo em um primeiro momento, mas devemos lembrar que esse é um modelo longitudinal que incorpora ambas as medidas de desenvolvimento cognitivo (língua e matemática) como preditores no modelo. A proporção da variância explicada é alta comparada com outros estudos no campo educacional no Brasil, tendo aqui um R^2 de 65%. Não obstante, chama a atenção o coeficiente para o nível socioeconômico agregado no nível da escola (nível 2 do modelo). Os dados sugerem um effect size de 0.29, o que pode ser interpretado como uma evidência de efeito par ou efeito da composição social do alunado no aprendizado (BARTHOLO et al., 2020; BARTHOLO; COSTA, 2016). Esse é um resultado já observado em outros estudos no Brasil com dados para o ensino fundamental e pré-escola, no entanto, é importante destacar que os efeitos anteriores utilizando apenas dados da rede pública de ensino eram menores. Os valores reforçam a forte segregação no sistema educacional brasileiro no início da escolarização obrigatória e seus possíveis efeitos no aumento das desigualdades educacionais observadas ao longo da escolarização obrigatória.

Existe uma vasta literatura sobre segregação escolar e seus efeitos no aprendizado. O chamado efeito par, “efeito mistura” ou efeito da composição social do alunado é tema de

interesse em diversos países do mundo e gera debates importantes sobre a necessidade de criar legislação e programas que previnam que crianças pobres ou que apresentem alguma outra característica de desvantagem (por exemplo, sejam estrangeiros e não falem a língua local) sejam alocadas juntas em escolas específicas. A evidência internacional mostra que separar os alunos em desvantagem potencial pode influenciar a forma como são tratados na escola, a qualidade do ensino e suas aspirações para etapas subsequentes do processo de escolarização. Esse é um primeiro resultado importante da dissertação e que merece um debate amplo entre pesquisadores, gestores e membros da sociedade civil sobre que sistema educacional queremos. Os dados apresentados reforçam que a forte estratificação observada na alocação dos alunos no início da escolarização obrigatória está associada com um aumento das desigualdades educacionais observada no primeiro ano na escola. Em outras palavras, um sistema educacional com grande proporção de matrículas em escolas privadas que cobram mensalidade, separa, desde o momento inicial, crianças com perfil socioeconômico diferentes em diferentes escolas. O fato de crianças com perfil socioeconômico diferentes estudarem separadas pode ser visto como um fator que aumenta as desigualdades educacionais medidas aqui a partir do aprendizado em linguagem.

Outro resultado interessante que aparece na Tabela 20 é o coeficiente estimado para o indicador de ambiente de aprendizagem em casa. Diferente do que foi observado para o indicador de nível socioeconômico no nível 1 do modelo, o ambiente de aprendizagem em casa explica parte da variação no aprendizado ao longo do primeiro ano na escola. O coeficiente é estatisticamente significativo e apresenta um effect size de 0.1, ou um ganho adicional de 1 a 2 meses de aprendizado. Efeitos dessa magnitude por vezes são interpretados como efeitos pequenos em estudos que avaliam o impacto de programas educacionais. No entanto, é importante considerar que os efeitos das rotinas da família podem ser cumulativos – não necessariamente irão parar no segundo ano da pré-escola ou no ensino fundamental. Isso reforça a importância de políticas educacionais que orientem os pais na construção de uma rotina em casa que envolvam brincadeiras e estímulos adequados para as crianças. A escola pode ter um papel importante na socialização dessas informações e apoio às famílias. É importante reforçar que os dados analisados nessa dissertação evidenciam uma correlação positiva entre o indicador de nível socioeconômico e o indicador de ambiente de aprendizagem em casa. Desse modo, a construção

de programas de apoio e orientação das famílias pode ter impacto positivo na diminuição das desigualdades educacionais.

A frequência à creche não está associada com o aprendizado no primeiro ano da pré-escola. O coeficiente é muito pequeno e não apresenta significância estatística. Cabe destacar mais uma vez, a importância de realizar novos estudos que possam associar não apenas a frequência à creche, mas também incorporar medidas sobre a qualidade da oferta, e associar com o desenvolvimento cognitivo no início da escolarização obrigatória, assim como, o aprendizado ao longo da pré-escola.

Os últimos coeficientes no nível 1 do modelo são as medidas sobre o desenvolvimento cognitivo no início do ano letivo. Além de incluir a medida inicial sobre o desenvolvimento em linguagem, optou-se também por incluir a medida inicial sobre o desenvolvimento em matemática seguindo modelo já realizado em outros estudos internacionais que utilizam o PIPS. É possível observar que a medida inicial sobre matemática explica parte da variação e aumenta a variância total explicada no modelo.

O nível 2 do modelo apresenta a medida agregada de nível socioeconômico e a variável que indica se a escola é da rede pública ou privada. É possível observar que o coeficiente que identifica a rede da escola perde força à medida que mais covariáveis são incluídas no modelo. O modelo final sugere que a variável que identifica as escolas públicas e privadas está associada com o aprendizado. O coeficiente positivo sugere que as crianças matriculadas na rede privada, controlado pelo ponto de partida e demais covariáveis de controle, aprenderam de forma mais acelerada – effect size de 0.31 ou aproximadamente 4 meses a mais de aprendizado do que as crianças matriculadas na rede pública após um do início da Educação básica obrigatória (primeiro ano da pré-escola). Esse é um resultado que deve ser interpretado com cautela, em especial considerando as limitações da amostra na rede privada. Nossos dados sugerem que as escolas de mensalidade estão sobrerrepresentadas na amostra, o que poderia ser uma ameaça para os resultados apresentados, em especial se existir diferenças nos resultados que serão analisados a seguir discriminando as escolas privadas em 3 grupos – baixa, média e alta mensalidade. Nesse caso, a diferença observada estaria sendo superestimada no modelo.

Uma outra ameaça que pode auxiliar a superestimar a aprendizagem das crianças matriculadas nas escolas privadas frente às públicas é o ano de coleta. Ao final de 2017 foi instituída nacionalmente a Base Nacional Comum Curricular (BNCC) que estabeleceu diretrizes curriculares para Educação Infantil e Fundamental para todo o país, obrigando a adaptação curricular de instituições públicas e privadas até 2020. A amostra de escolas públicas é coletada em 2017, enquanto a de escolas privadas ocorre em 2019. É possível que, já em 2019, instituições privadas da amostra tenham reestruturados seus currículos para se adaptar à base, causando, assim, um risco de viés histórico aos resultados. Ou seja, os coeficientes para as escolas privadas podem estar captando, em parte, o impacto de implementação da BNCC, superestimando a diferença para as escolas públicas.

A Tabela de 24 abaixo apresenta os resultados para modelos hierárquicos que estima a medida de linguagem no final do ano letivo (mesma variável dependente da Tabela 20), porém a principal diferença nos modelos é que a comparação agora é com os três estratos das escolas privadas, a saber, escolas privadas de baixa, média e alta mensalidade. Cabe ressaltar que os coeficientes estimados no nível 1 do modelo são muito parecidos com o descrito anteriormente na Tabela 20 e não serão discutidos aqui. Iremos discutir os coeficientes no nível 2 do modelo – nível da escola. O primeiro ponto que chama a atenção é o indicador de nível socioeconômico médio da escola que deixa de ser estatisticamente significativo. Uma possível explicação para esse resultado é que o número de escolas em cada estrato diminuiu uma vez que as escolas privadas foram divididas de acordo com o valor da mensalidade. No entanto, a direção e força da associação permanecem muito semelhantes reforçando o resultado anterior apresentado.

Tabela 24: análise de aprendizagem no 1º ano da pré-escola por estrato - Regressão multinível para desenvolvimento de linguagem 2ª onda

Regressão multinível - Fatores associados à aprendizagem e linguagem do 1º ano da pré-escola.

	Modelo Nulo	Modelo 1	Modelo 2	Modelo 3	Modelo 4
(Intercepto)	0.55 *** (0.40 0.70)	0.15 *** (0.08 0.23)	-3.38 *** (-3.85 -2.91)	-0.98 *** (-1.40 -0.57)	-0.21 (-0.57 0.15)
Escola:Baixa mens.		0.62 *** (0.35 0.90)	0.10 (-0.18 0.38)	-0.01 (-0.22 0.20)	0.14 (-0.06 0.34)

Escola:Média mens.		0.88 *** (0.65 1.11)	0.39 ** (0.10 0.67)	0.16 (-0.06 0.38)	0.22 * (0.02 0.42)
Escola:Alta mens.		1.51 *** (1.32 1.69)	0.69 *** (0.37 1.01)	0.28 * (0.04 0.52)	0.24 * (0.01 0.47)
Idade (meses)			0.07 *** (0.06 0.08)	0.03 *** (0.02 0.04)	0.01 *** (0.01 0.02)
Sexo:Masculino			-0.10 ** (-0.16 -0.04)	-0.09 ** (-0.14 -0.03)	-0.02 (-0.06 0.03)
Creche:Frequentou			-0.01 (-0.07 0.06)	-0.03 (-0.09 0.02)	-0.03 (-0.08 0.01)
Indicador de Ambiente de Aprendizagem em Casa			0.06 *** (0.03 0.08)	0.04 *** (0.02 0.05)	0.02 * (0.00 0.04)
Indicador de Nível socioeconômico (Indivíduo)			0.10 *** (0.07 0.12)	0.04 *** (0.02 0.06)	0.01 (-0.01 0.03)
Indicador de Nível socioeconômico (Escola)			0.16 ** (0.05 0.28)	0.09 * (0.01 0.18)	0.08 (-0.01 0.16)
Desempenho em matemática início do ano				0.36 *** (0.34 0.38)	0.18 *** (0.16 0.20)
Desempenho em linguagem início do ano					0.45 *** (0.42 0.48)
Random Effects					
σ^2	0.74	0.74	0.65	0.45	0.33
τ_{00}	0.40 _{ESC}	0.05 _{ESC}	0.02 _{ESC}	0.01 _{ESC}	0.01 _{ESC}
ICC	0.35	0.07	0.03	0.02	0.03
N	⁷⁵ ESC	⁷⁵ ESC	⁷⁵ ESC	⁷⁵ ESC	⁷⁵ ESC
Observations	2712	2712	2641	2632	2631
Marginal R ² / Conditional R ²	0.000 / 0.348	0.216 / 0.267	0.329 / 0.350	0.535 / 0.544	0.656 / 0.666

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Fonte: Elaboração própria do autor.

Tabela 25: análise de aprendizagem no 1º ano da pré-escola por estrato – Effect Size para desenvolvimento de linguagem 2ª onda

	Effect-Size	Equivalente em meses de aprendizagem
Baixa mensalidade	0.244	3 meses
Média mensalidade	0.384*	5 meses
Alta mensalidade	0.419*	5 meses
Idade meses (Z-escore)	0.184*	3 meses
Sexo:Masculino	-0.034	-
Creche:Frequentou	-0.057	-
AAC (Z-Escore)	0.1*	2 meses
NSE Indivíduo (Z-escore)	0.06	-
Indicador de Nível socioeconômico (Escola)	0.245	3 meses
Desempenho em matemática início do ano	0.941*	11 meses
Desempenho em linguagem início do ano	1.678*	18 meses

Fonte: Elaboração própria do autor. - * $p < 0.05$

Os coeficientes que identificam os estratos das escolas privadas permitem comparar a categoria de base, aqui definida como as escolas da rede pública, com os três estratos das escolas privadas. Inicialmente é possível observar que a associação entre o estrato e o aprendizado aumenta à medida que o valor da mensalidade aumenta. Em outras palavras, há uma clara relação entre o valor médio das mensalidades e o aprendizado das crianças no primeiro ano de escolarização obrigatório. O coeficiente estimado para as escolas privadas de baixa mensalidade sugere uma correlação positiva sem significância estatística. Esse resultado dialoga em grande medida com o resultado descrito na tabela 18 que descreveu fatores associados ao desenvolvimento em e linguagem no início da pré-escola. Os resultados descritos no capítulo anterior sugeriam que não havia claro viés de seleção entre o público da escola pública e as escolas da rede privada de baixa mensalidade. Os modelos da tabela 24 sugerem que não há evidência de diferenças estatisticamente significativas no aprendizado em linguagem dos alunos matriculadas na rede pública e privada de baixa mensalidade. É importante destacar duas

limitações na interpretação dos dados. A primeira já amplamente discutida na dissertação é referente a amostra na rede privada ser uma amostra por conveniência. O segundo é o número baixo de unidades/escolas no modelo que compara a rede pública (46 escolas) com a rede privada de mensalidade baixa (5 escolas).

O coeficiente para as escolas privadas de média mensalidade indica uma associação positiva e estatisticamente significativa entre o estrato e o aprendizado, sempre tendo a rede pública de ensino como a categoria de base. Controlado pelo ponto de partida em linguagem e matemática e demais covariáveis de controle do modelo, os resultados sugerem um effect size de 0.38 ou 5 meses a mais de aprendizado na média para os alunos matriculados nas escolas privadas de média mensalidade. Os dados evidenciam que as diferenças observadas entre a rede pública e privada parecem estar menos associadas às escolas de baixa mensalidade e mais às escolas de média e alta mensalidade. Cabe ressaltar mais uma vez que a amostra intencional das escolas privadas aliadas a um número baixo de unidades/escolas de média mensalidade sugerem cautela na interpretação dos resultados.

Por fim, o coeficiente que identifica as escolas privadas de alta mensalidade sugere novamente uma associação entre o aprendizado e a frequência nessas escolas. No total temos 14 escolas identificadas no grupo de alta mensalidade – o maior número entre os estratos das escolas privadas. É possível observar que se trata do maior coeficiente entre as escolas privadas, reforçando o argumento que há uma relação entre escolas que cobram maiores mensalidades e o aprendizado no primeiro ano na pré-escola. O modelo final com todas as covariáveis de controle sugere um effect size de 0.42 ou aproximadamente 5 meses a mais de aprendizado na média para os alunos matriculados nas escolas privadas de alta mensalidade comparado com os alunos matriculados na rede pública de ensino.

Os dados apresentados nas Tabelas de 24 e 25 sugerem que as desigualdades mensuradas no início da escolarização obrigatória – ver resultados do capítulo 4.2. – crescem ao longo do primeiro ano da escolarização obrigatória quando comparamos a rede pública com a rede privada – ver coeficientes da Tabela 24 e 25. A separação das escolas da rede privada em três estratos apresentam evidências preliminares que talvez as diferenças observadas estejam concentradas em escolas privadas que cobram mensalidades de valor médio ou alto (considerando os pontos de corte definidos nesse estudo). Estudos complementares devem buscar replicar essas análises

com uma amostra representativa de escolas da rede privada. Não obstante, os resultados apresentados não possuem paralelo em outros estudos realizados no Brasil para a faixa etária da pré-escola e, mesmo considerando as limitações da amostra na rede privada, devem ser consideradas como evidência preliminar que as desigualdades educacionais aumentam após o primeiro ano da pré-escola considerando a oferta estratificada pela rede pública e privada. Implicações para o sistema educacional no Brasil e para política educacional serão debatidos no capítulo final da dissertação.

4.4. Fatores associados ao aprendizado em matemática no primeiro ano da pré-escola

Nesse capítulo serão apresentados os resultados dos modelos hierárquicos que estimam o aprendizado em matemática no primeiro ano da escolarização obrigatória. A Tabela 26 apresenta o modelo que tem como variável dependente a medida de matemática ao final do primeiro ano da escola. Essa primeira análise tem como principal interesse analisar diferenças no aprendizado considerando os alunos matriculados na rede pública e privado de ensino. Em seguida, apresentaremos as análises que separam as escolas privadas em três estratos a partir dos critérios já explicitados.

O primeiro ponto relevante é destacar o ICC mais alto do que o relatado em estudos anteriores realizados no Brasil com amostras de crianças na pré-escola matriculadas na rede pública de ensino (KOSLINSKI; BARTHOLO, 2019; BARTHOLO et al., 2020). Esse dado reforça mais uma vez a importância de incluirmos dados da rede privada de ensino nos estudos sobre desigualdades educacionais. As análises que trabalham apenas com as matrículas na rede pública correm um grande risco de subestimar o fenômeno da segregação escolar e das desigualdades educacionais.

Os modelos apresentados para estimar o aprendizado em matemática seguem a mesma lógica dos modelos descritos para linguagem. É possível observar que os padrões observados para o aprendizado em matemática não têm necessariamente o mesmo do que foi observado em linguagem e, por isso, faremos uma descrição inicial detalhada de todas as variáveis do nível 1 do

modelo e em seguida iremos interpretar as variáveis que identificam as escolas da rede pública e privada – nível 2 dos modelos.

Tabela 26: análise de aprendizagem no 1º ano da pré-escola por rede - Regressão multinível para desenvolvimento de matemática 2ª onda

Regressão multinível - Fatores associados à aprendizagem em matemática no 1º ano da pré-escola.

	Modelo Nulo	Modelo 1	Modelo 2	Modelo 3	Modelo 4
(Intercepto)	0.96 *** (0.69 1.22)	0.20 * (0.04 0.37)	-6.05 *** (-6.83 -5.26)	-2.94 *** (-3.66 -2.22)	-0.57 (-1.17 0.02)
Escola:Privada		2.11 *** (1.81 2.40)	0.87 *** (0.38 1.35)	0.88 *** (0.46 1.29)	0.56 *** (0.25 0.87)
Idade (meses)			0.12 *** (0.11 0.13)	0.06 *** (0.05 0.08)	0.03 *** (0.02 0.04)
Sexo:Masculino			0.05 (-0.05 0.15)	0.17 *** (0.08 0.26)	0.11 ** (0.03 0.18)
Creche:Frequentou			0.00 (-0.11 0.11)	-0.03 (-0.12 0.07)	-0.05 (-0.13 0.03)
Indicador de Ambiente de Aprendizagem em Casa			0.08 *** (0.04 0.12)	0.04 * (0.00 0.07)	0.02 (-0.01 0.05)
Indicador de Nível socioeconômico (Indivíduo)			0.18 *** (0.14 0.22)	0.09 *** (0.06 0.13)	0.04 ** (0.01 0.07)
Indicador de Nível socioeconômico (Escola)			0.33 ** (0.12 0.54)	0.16 (-0.02 0.33)	0.04 (-0.09 0.18)
Desempenho em linguagem início do ano				0.71 *** (0.66 0.76)	0.23 *** (0.18 0.28)
Desempenho em matemática início do ano					0.67 *** (0.63 0.70)
Random Effects					
σ^2	2.07	2.07	1.80	1.38	0.90
τ^{00}	1.31 _{ESC}	0.28 _{ESC}	0.11 _{ESC}	0.08 _{ESC}	0.04 _{ESC}

ICC	0.39	0.12	0.06	0.05	0.04
N	⁷⁵ ESC	⁷⁵ ESC	⁷⁵ ESC	⁷⁵ ESC	⁷⁵ ESC
Observations	2712	2712	2641	2640	2631
Marginal R ² / Conditional R ²	0.000 / 0.387	0.221 / 0.313	0.344 / 0.382	0.495 / 0.522	0.672 / 0.686

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Fonte: Elaboração própria do autor.

Tabela 27: análise de aprendizagem no 1º ano da pré-escola por rede – Effect Size para desenvolvimento de matemática 2ª onda

	Effect-Size	Equivalente em meses de aprendizagem
Escola:Privada	0.588*	7 meses
Idade meses (Z-escore)	0.205*	3 meses
Sexo:Masculino	0.111*	2 meses
Creche:Frequentou	-0.05	-
AAC (Z-Escore)	0.071	1 mês
NSE Indivíduo (Z-escore)	0.139*	2 meses
Indicador de Nível socioeconômico (Escola)	0.08	1 mês
Desempenho em linguagem início do ano	0.512*	6 meses
Desempenho em matemática início do ano	2.128*	26 meses

Fonte: Elaboração própria do autor. - * $p < 0.05$

Os coeficientes estimados para o primeiro ano na escola sugerem que meninos aprendem mais ao longo do ano letivo – effect size de 0.11 ou aproximadamente 2 meses de vantagem para os meninos. Não é simples interpretar esse resultado e estudos futuros devem buscar identificar os mecanismos dentro das escolas e, eventualmente na rotina das próprias famílias, que explicam o resultado. Esse resultado reforça o que foi observado no estudo Geres, assim como a diferença verificada em avaliações padronizadas, como por exemplo a Prova Brasil ou o PISA, que destaca que os meninos ou adolescentes homens, apresentam proficiência média mais alta do que as

meninas em matemática, mesmo após controle das variáveis de nível socioeconômico (BROOKE et al, 2014). Os dados sugerem que parte desse efeito diferencial no aprendizado pode estar associado com a escola.

O indicador de nível socioeconômico está associado com o aprendizado em matemática, mesmo após o controle das medidas iniciais sobre o desenvolvimento cognitivo (matemática e linguagem) no início do ano letivo. Os coeficientes estimados indicam um effect size de 0.14 ou dois meses a mais de aprendizado comparando crianças cujo as famílias têm um desvio padrão abaixo ou acima da média do indicador. Esse resultado não foi encontrado nos modelos estimando o aprendizado em linguagem e devem ser aprofundados em estudos futuros. Cabe lembrar que as diferenças observadas no desenvolvimento em matemática no início da escolarização obrigatória foram consideravelmente maiores, assim como a estratificação entre a rede pública e privada. Não é claro porque o aprendizado em matemática está mais associado com o nível socioeconômico das famílias, mas deve ser objeto de análises futuras, inclusive com coleta de dados qualitativos para compreender os mecanismos que podem ajudar na formulação de hipóteses plausíveis para explicar os resultados descritos.

O indicador de ambiente de aprendizagem em casa apresenta correlação positiva com o aprendizado em matemática, com magnitude parecida ao descrito para o aprendizado em linguagem (effect size de 0.07), porém o coeficiente estimado não apresenta significância estatística.

A frequência à creche, assim como descrito para o aprendizado em linguagem, não apresenta coeficiente estatisticamente significativo após o controle das demais covariáveis. O modelo sugere uma associação negativa, porém com magnitude muito pequena. Novamente reforçamos a importância de estudos que incorporem medidas sobre a qualidade da oferta na creche para compreender melhor a relação entre a frequência à creche e o aprendizado.

Os principais preditores dos modelos apresentados na Tabela 26 são as medidas iniciais sobre o desenvolvimento cognitivo no início do ano letivo. Além de incluir a medida inicial sobre o desenvolvimento em matemática, optamos também por incluir a medida inicial sobre o desenvolvimento em linguagem seguindo modelo já realizado em outros estudos internacionais que utilizam o PIPS, assim como nos modelos já explicitados nas Tabelas 20 e 21.

As variáveis no nível 2 do modelo apresentam o indicador de nível socioeconômico agregado para a escola e a variável de rede, que identifica se a escola é pública ou privada. O indicador de nível socioeconômico da escola apresenta associação positiva com o aprendizado em matemática, porém o coeficiente estimado não apresenta significância estatística.

Os resultados estimados permitem analisar se há uma relação entre o aprendizado em matemática e a rede de ensino (escolas públicas ou privadas) que os alunos frequentam. Os resultados do modelo final sugerem que há uma relação entre frequentar as escolas da rede privada e um aprendizado mais acelerado. O coeficiente estimado sugere um effect size de 0.59 ou aproximadamente 7 meses de aprendizado a mais para os alunos matriculados na rede privada.

O resultado é maior do que o descrito para linguagem e reforça em parte os resultados descritos nos modelos que têm como variável dependente a medida inicial de matemática. É oportuno reforçar mais uma vez as limitações da amostra das escolas privadas e a necessidade de interpretar os resultados com cautela. Há um risco claro dos resultados apresentados estarem superestimando as diferenças entre as redes pública e privada uma vez que escolas de alta mensalidade apresentam uma maior representatividade na amostra intencional realizada.

A Tabela 28 apresenta os resultados dos modelos hierárquicos que estimam o aprendizado em matemática com a principal diferença que a variável de rede é substituída pela variável de estrato. As análises comparam os três estratos das escolas privadas (baixa, média e alta mensalidade) com a categoria de base – as escolas da rede pública. Cabe ressaltar que os coeficientes estimados no nível 1 do modelo (nível aluno) são muito parecidos com o descrito anteriormente na Tabela 26 e 27 e não serão discutidos aqui. As análises irão se concentrar nos coeficientes que identificam os diferentes estratos das escolas privadas. O principal objetivo é identificar se há uma relação entre o aprendizado das crianças em matemática e os diferentes estratos das escolas privadas, sempre tendo a rede pública como a categoria de base.

Tabela 28: análise de aprendizagem no 1º ano da pré-escola por estrato - Regressão multinível para desenvolvimento de matemática 2ª onda

Regressão multinível - Fatores associados à aprendizagem e matemática do 1º ano da pré-escola.

	Modelo Nulo	Modelo 1	Modelo 2	Modelo 3	Modelo 4
(Intercepto)	0.96 *** (0.69 1.22)	0.21 ** (0.07 0.34)	-6.09 *** (-6.87 -5.30)	-2.96 *** (-3.68 -2.24)	-0.59 (-1.18 0.01)
Escola:Baixa mens.		1.37 *** (0.90 1.85)	0.65 ** (0.17 1.13)	0.80 *** (0.34 1.25)	0.52 ** (0.17 0.87)
Escola:Média mens.		1.65 *** (1.24 2.05)	1.01 *** (0.51 1.50)	0.93 *** (0.47 1.39)	0.58 ** (0.23 0.94)
Escola:Alta mens.		2.69 *** (2.36 3.01)	1.60 *** (1.05 2.15)	1.21 *** (0.69 1.73)	0.75 *** (0.35 1.15)
Idade (meses)			0.12 *** (0.10 0.13)	0.06 *** (0.05 0.08)	0.03 *** (0.02 0.04)
Sexo:Masculino			0.05 (-0.06 0.15)	0.17 *** (0.07 0.26)	0.10 ** (0.03 0.18)
Creche:Frequentou			-0.00 (-0.11 0.11)	-0.03 (-0.13 0.07)	-0.05 (-0.13 0.03)
Indicador de Ambiente de Aprendizagem em Casa			0.08 *** (0.04 0.12)	0.04 * (0.00 0.07)	0.02 (-0.01 0.05)
Indicador de Nível socioeconômico (Indivíduo)			0.18 *** (0.14 0.22)	0.09 *** (0.06 0.13)	0.04 ** (0.01 0.07)
Indicador de Nível socioeconômico (Escola)			0.16 (-0.04 0.36)	0.08 (-0.11 0.27)	-0.00 (-0.15 0.14)
Desempenho em linguagem início do ano				0.70 *** (0.65 0.75)	0.23 *** (0.18 0.27)
Desempenho em matemática início do ano					0.67 *** (0.63 0.70)
Random Effects					
σ^2	2.07	2.08	1.80	1.38	0.90

r^2_{00}	1.31 _{ESC}	0.16 _{ESC}	0.07 _{ESC}	0.07 _{ESC}	0.04 _{ESC}
ICC	0.39	0.07	0.04	0.05	0.04
N	⁷⁵ ESC	⁷⁵ ESC	⁷⁵ ESC	⁷⁵ ESC	⁷⁵ ESC
Observations	2712	2712	2641	2640	2631
Marginal R ² / Conditional R ²	0.000 / 0.387	0.245 / 0.300	0.353 / 0.377	0.498 / 0.522	0.673 / 0.686

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Fonte: Elaboração própria do autor.

Tabela 29: análise de aprendizagem no 1º ano da pré-escola por estrato – Effect Size para desenvolvimento de matemática 2ª onda

	Effect-Size	Equivalente em meses de aprendizagem
Baixa mensalidade	0.546*	7 meses
Média mensalidade	0.612*	7 meses
Alta mensalidade	0.79*	10 meses
Idade meses (Z-escore)	0.205*	3 meses
Sexo:Masculino	0.11*	2 meses
Creche:Frequentou	-0.052	-
AAC (Z-Escore)	0.069	1 mês
NSE Indivíduo (Z-escore)	0.14*	2 meses
Indicador de Nível socioeconômico (Escola)	-0.004	-
Desempenho em linguagem início do ano	0.507*	6 meses
Desempenho em matemática início do ano	2.126*	26 meses

Fonte: Elaboração própria do autor. - * $p < 0.05$

Inicialmente é possível observar que a associação entre o estrato e o aprendizado aumenta à medida que o valor da mensalidade aumenta – mesmo padrão observado para o aprendizado em linguagem. Isso reforça o argumento que há uma relação entre o aprendizado e os diferentes estratos de instituições privadas. O coeficiente estimado para as escolas privadas de baixa mensalidade sugere uma correlação positiva e estatisticamente significativa com o aprendizado em matemática. Esse resultado é diferente do observado para linguagem. O coeficiente estimado apresenta um effect size de 0.55 ou aproximadamente 7 meses a mais de aprendizado. Os coeficientes estimados para as escolas de média e alta mensalidade apresentam respectivamente um effect size de 0.61 e 0.79. Esses são efeitos grandes de aproximadamente 7 meses para as escolas de mensalidade média e 9/10 meses para as escolas de alta mensalidade.

Os resultados apresentados nas Tabelas 28 e 29 são consistentes e sugerem que o aprendizado em matemática está associado com as escolas privadas e, de forma ainda mais específica, com os estratos das escolas privadas. Quanto maior o estrato, maior o aprendizado associado. Os resultados preliminares descritos devem ser interpretados considerando as limitações da amostra. Não obstante, os dados reforçam que as desigualdades descritas no início da escolarização obrigatória, mensuradas a partir do viés de seleção/estratificação, entre as redes pública e privada, se intensificam ao longo do primeiro ano na escola. O cenário sugere uma distribuição dos alunos entre as escolas da cidade com padrão desigual quando pensamos o perfil socioeconômico das famílias e o seu desenvolvimento cognitivo no início da escolarização obrigatória. Essa desigualdade na linha base, isto é, no início da Educação Básica obrigatória, amplia-se ao longo do primeiro ano pré-escolar, em especial para a medida sobre o desenvolvimento em matemática. Fatores associados as famílias como o indicador de nível socioeconômico (em especial para o aprendizado em linguagem) e o indicador de ambiente de aprendizagem em casa (para o aprendizado em linguagem) parecem explicar parte das variações observadas.

Os resultados descritos devem ser seguidos de outros estudos com amostras representativas para a faixa etária estudada. Os dados apresentam evidências preliminares de um sistema altamente segregado no ponto de partida e que potencialmente amplia as desigualdades pré-existentes ao separar em escolas diferentes, alunos com perfil socioeconômico e estágio inicial de desenvolvimento distintos.

CONSIDERAÇÕES FINAIS

O presente capítulo de considerações finais foca em três principais pontos. O primeiro é realizar um apanhado dos principais resultados evidenciados na dissertação. O segundo busca atrelar os resultados com sugestões de políticas públicas que possam auxiliar no combate das desigualdades educacionais, tanto nas de início da vida escolar quanto na produção da desigualdade ao longo dos anos escolares. Por fim, aponta para os próximos passos a serem desenvolvidos em pesquisas futuras.

Investigar o efeito escolar na aprendizagem e nas desigualdades educacionais, em especial olhando para a estratificação entre as redes pública e privada, é primordial para produzir evidências que orientem políticas públicas com o objetivo de melhorar a qualidade e equidade do sistema educacional no Brasil. Contudo, são diminutas as evidências sobre o tamanho da estratificação escolar na literatura nacional que agregam dados de ambas as redes, em especial no que tange à aprendizagem cognitiva, elemento fundamental enquanto medida de qualidade educacional. Sendo assim, a presente dissertação buscou investigar a) o tamanho da desigualdade observada no primeiro ano da escolarização obrigatória no desenvolvimento cognitivo de crianças matriculadas em instituições públicas e privadas e b) o impacto de ofertas pública e privada de ensino no desenvolvimento cognitivo ao longo do primeiro ano de escolarização obrigatória.

Compreender profundamente as estruturas de desigualdades educacionais passa por dois principais pontos. O primeiro é descrever as desigualdades educacionais anteriores ao início da vida escolar, mensurando os padrões de estratificação que não têm relação com a escola, mas, sim, com as famílias, as vivências socioculturais e políticas de primeira infância às quais as crianças são expostas (FRANCO; BROOKE; ALVES, 2008; GOLDSTEIN, 1997). Com isso, é possível mensurar o tamanho da desigualdade educacional assim que a vida escolar se inicia e estabelecer uma linha de base de comparação para os anos subsequentes. O segundo ponto é medir a aprendizagem ao longo da vida escolar, verificando o quanto que essa desigualdade de início se modifica ao longo do tempo e identificar os fatores escolares que impactam na

ampliação ou diminuição dessa desigualdade. Para tanto, é necessário estudos longitudinais e de modelagem de valor agregado para controle de características tanto familiares e socioculturais quanto fatores escolares, possibilitando identificar o quanto que as redes escolares impactam no desenvolvimento cognitivo e sustentam um sistema que produz desigualdades ou equidades educacionais (BONAMINO; OLIVEIRA, 2013; LEE, 2010; REARDON; RAUDENBUSH, 2009; SOARES et al., 2017).

A revisão bibliográfica apontou, porém, que nos últimos 20 anos apenas um estudo longitudinal de larga escala foi realizado no Brasil para identificar fatores escolares e extraescolares associados à aprendizagem e as desigualdades educacionais no início da escolarização obrigatória: o projeto Geres (FRANCO; BROOKE; ALVES, 2008). A partir de 2009, contudo, a Educação Básica obrigatória foi ampliada, tendo como seu marco inicial não mais o Ensino Fundamental – foco do Geres -, mas, sim, a pré-escola. A presente dissertação busca contribuir com essa lacuna na literatura nacional sobre desigualdade escolar e com os estudos de primeira infância no país, lançando mão de dados primário de uma pesquisa de larga escala realizada em uma grande cidade no sudeste do país.

O estudo de desenho longitudinal, analisou resultados cognitivos em linguagem em matemática de 2712 crianças de 4 a 5 anos de escolas públicas e privadas, que realizaram a avaliação *Performance Indicator in Primary Schools (PIPS)* no início do primeiro ano pré-escolar (1ª onda) e ao final do mesmo ano letivo (2ª onda). O *PIPS* é uma ferramenta cognitiva de linguagem e matemática desenvolvida e utilizada para elaborar devolutivas pedagógicas para professores da Educação Infantil, sendo amplamente utilizada e validada por docentes de diversos países. Para o Brasil, a ferramenta foi traduzida, adaptada conforme características culturais regionais e validada com uso exclusivo para pesquisa. Os dados foram coletados pelo Laboratório de Pesquisa em Oportunidades Educacionais (LaPOpE/UFRJ). As coletas de dados da presente dissertação ocorreram em dois anos, sendo em 2017 com as escolas públicas e em 2019 com as privadas, contendo medidas cognitivas e contextuais de famílias e das instituições escolares participantes. As análises realizadas nesta dissertação foram um recorte desse grande estudo que continha mais de cinco mil crianças participantes.

Uma vez que a pré-escola é o começo da Educação Básica obrigatória, os resultados da 1ª onda (início do ano letivo do primeiro ano da pré-escola) do estudo apontam para as desigualdades do início da vida escolar, não relacionadas com a escola em que estavam matriculadas, mas, principalmente, com as vivências sociais, culturais e ambientais anteriores à escola. Através da 2ª onda de coleta de dados, ao final do ano letivo, e comparando com os resultados da 1ª onda, foi possível mensurar a aprendizagem ao longo do primeiro ano da pré-escola, calculando o impacto do perfil de rede escolar na aprendizagem e desigualdade escolar, além de identificar fatores socioculturais na estratificação educacional e aprendizagem. Para além das análises descritivas dos dados, utilizou-se de modelagem hierárquica/multinível de dois níveis (aluno e escola) para cálculo de impacto da rede escolar e medidas de desigualdade educacional. Os resultados cognitivos de linguagem e matemática compunham as variáveis dependentes dos modelos, tendo sido calculadas em escala de Rasch (LINACRE, 1994), técnica que pode ser compreendida como um subconjunto da Teoria de Resposta ao Item (TRI). As variáveis independentes que compuseram os modelos podem ser resumidas em indicador de frequência à creche, idade da criança em meses, Ambiente de Aprendizagem em Casa (AAC), nível socioeconômico (NSE) da família da criança, nível socioeconômico médio da escola (em análises de valor agregado), rede (pública ou privada) e estrato (pública ou privada de baixa, média ou alta mensalidade), assim como os resultados cognitivos da 1ª onda quando analisados os modelos de valor agregado (aprendizagem).

Para facilitar a interpretação dos resultados, os coeficientes dos modelos hierárquicos foram convertidos em Effect Size (ES) e interpretados em meses escolares de aprendizagem, utilizando como referência Higgins et al (2014). Os resultados se dividiram em duas etapas, conforme os objetivos da dissertação. A primeira analisa as desigualdades no início do ano letivo no primeiro ano da pré-escola, ou seja, as desigualdades de início da vida escolar e que estão majoritariamente relacionadas aos perfis familiares e oportunidades anteriores à escolarização. A segunda etapa de análise buscou medir o impacto da rede escolar (público e privada) na aprendizagem pré-escolar e, por conseguinte, nas desigualdades educacionais logo que se inicia a vida escolar obrigatória.

Para além da desigualdade entre rede pública e privada, foi analisado o panorama de estratificação interna à rede privada. Para tanto, optou-se por realizar análises que separavam a

rede privada em três estratos: baixa mensalidade (até R\$400), média mensalidade (R\$401 a R\$1000) e alta mensalidade (acima de R\$1000). Isso possibilita comparar os estratos com os resultados da rede pública e explorar estratificações internas à própria rede privada.

É importante ressaltar que há diferentes elementos que possam contribuir para uma possível superestimação do impacto da rede privada frente à rede pública e na desigualdade no início da educação básica obrigatória, como foi evidenciado nos resultados da dissertação. O primeiro deles já é amplamente debatido ao longo da dissertação que se expressa na amostra privada ser intencional e agregar mais escolas privadas de alta mensalidade do que de outros estratos (baixa ou média mensalidade). Uma vez que há forte correlação entre o nível socioeconômico, mensalidade e desenvolvimento cognitivo no início da vida escolar, o agrupamento sobrerrepresentado de escolas de alta mensalidade pode aumentar a média das escolas privadas para a desigualdade educacional quando comparado com a rede pública.

Outro fator que pode possibilitar a superestimação do impacto da rede privada é o efeito teto do indicador de Ambiente de Aprendizagem em Casa (AAC). As análises observadas no capítulo de imputação de dados, indica que essa variável tem pouca variabilidade para escolas de alta mensalidade e, por isso, pode acabar perdendo capacidade de explicar os efeitos na aprendizagem, inflando o resultado das escolas de alta mensalidade. É preciso de maiores investigações para compreender se isso ocorre.

Um último fator é um risco de viés histórico. Os dados das escolas públicas foram coletados no ano de 2017, enquanto para as escolas privadas em 2019. A Base Nacional Comum Curricular (BNCC) foi aprovada ao final de 2017 e, por isso, é possível que instituições privadas da amostra tenham implementado modificações curriculares através das orientações da BNCC. Nesse sentido, é possível que parte do impacto em aprendizagem da rede privada seja efeito de uma mudança curricular fruto da implementação da BNCC na rotina escolar. Seriam necessárias novas investigações para compreender se isso ocorre. Apesar dessas ameaças, os dados utilizados continuam a prover as melhores evidências no cenário nacional até o momento para a etapa da pré-escola, no que tange desigualdade cognitiva e aprendizado ao longo do primeiro ano da Educação Básica obrigatória.

Os resultados da dissertação apontam para uma intensa desigualdade educacional logo no começo da vida escolar quando comparamos o desenvolvimento cognitivo em linguagem e matemática entre crianças matriculadas na rede pública e privada. As análises descritivas apresentam um preocupante cenário. Crianças matriculadas na rede privada iniciam a vida escolar com o equivalente, em média, a 11 meses de aprendizagem escolar em linguagem (ES 0.91) e 15 meses em matemática (ES 1.26) a mais do que crianças que iniciam a vida escolar matriculadas na rede pública. É preciso ressaltar que essa desigualdade não é fruto da rede escolar, mas, sim, do contexto familiar da criança, suas vivências socioculturais no cotidiano fora da escola e políticas de primeira infância a que foram expostas em períodos anteriores ao início da Educação Básica obrigatória (antes dos 4 anos de idade).

Os modelos hierárquicos auxiliam a explicar os fatores que se associam a essa desigualdade do início da vida escolar. Para matemática, não se observa diferenças estatisticamente significativas entre crianças no sexo feminino e masculino. Por outro lado, crianças que frequentaram a creche iniciam a pré-escola, em média e controlado todas outras covariáveis, com um 1 mês (ES 0.081) a mais de desenvolvimento em matemática. Crianças com indicador de Ambiente de Aprendizagem em Casa (AAC) um desvio padrão acima da média desse indicador, apresentam 2 meses (ES 0.184) de desenvolvimento em matemática a mais. Crianças um desvio padrão acima da média de nível socioeconômico, assim como para idade em meses, apresentam alto impacto no desenvolvimento em matemática logo no início da escolarização, apontando, respectivamente, para 6 meses (ES 0.45) e 8 meses (ES 0.644) de aprendizado a mais em matemática.

Para linguagem, as tendências se mantêm conforme visto em matemática, com exceção de sexo. Meninas iniciam a pré-escola, em média e controlado por todas outras covariáveis, com 2 meses a mais de desenvolvimento linguagem do que os meninos. Outro ponto a se destacar é que o tamanho das desigualdades de linguagem e matemática são discrepantes. A distância média observada entre crianças matriculadas na rede pública e privada para linguagem é, nos modelos hierárquicos controlados, de 7 meses (ES 0.59) enquanto, para matemática, é de 12 meses (ES 0.991) de diferença de aprendizagem escolar. Uma vez que a rede escolar nessa primeira análise deve ser vista como um indicador de viés de seleção de perfil discente, o que pode ser interpretado entre a discrepância entre linguagem e matemática é que o contato com matemática

no período anterior à Educação Básica obrigatória é mais intenso com as famílias mais ricas (crianças matriculadas em escolas privadas) do que com famílias mais vulneráveis (crianças matriculadas em escolas públicas). É necessário que essa hipótese seja mais profundamente investigada em outros estudos, contudo, aponta que as desigualdades no desenvolvimento cognitivo em matemática para o início da vida escolar são mais intensas do que as desigualdades observadas em linguagem.

Os resultados dos modelos hierárquicos que subdividem a rede privada em estratos de baixa (até R\$400), média (R\$401 – R\$1000) e alta (acima de R\$1000) mensalidade apontam que, conforme a instituição escolar se torna financeiramente mais seletiva (aumento de mensalidade), maior tende a ser a discrepância média no desenvolvimento cognitivo em matemática e linguagem no início da vida escolar quando comparado com o desenvolvimento médio das crianças matriculadas na rede pública. Para matemática, crianças matriculadas no início da pré-escola em instituições de baixa mensalidade apresentam aprendizagem em matemática equivalente a 6 meses (ES 0.523) escolares a mais do que a média da rede pública. Esses valores se ampliam conforme o aumento no estrato, indicando 9 meses (ES 0.739) para média mensalidade e 16 meses (ES 1.336) para alta mensalidade. Já, para linguagem, não há resultado estatisticamente significativo de diferença no desenvolvimento cognitivo para o início da vida escolar entre rede pública e baixa mensalidade. Contudo, crianças matriculadas em escolas de média mensalidade apresentam 5 meses (ES 0.376) de aprendizado escolar a mais em linguagem e 12 meses para alta mensalidade, quando comparado com o resultado médio de linguagem das crianças matriculadas na rede pública. Nesse sentido, as análises que utilizam os estratos de mensalidade apontam não apenas que a desigualdade de início da vida escolar entre crianças matriculadas na rede pública e privada tende a ampliar conforme a escola é mais seletiva (aumento de mensalidade), mas, também, que há uma estratificação educacional intensa dentro da própria rede privada.

Desse modo, os resultados que focam o primeiro objetivo da dissertação, o de analisar a desigualdade de início da vida escolar obrigatória, apontam resumidamente para três principais fatores explicativos: A) as oportunidades as quais as crianças foram expostas na primeiríssima infância (0 a 3 anos), B) maturação biológica e C) a seletividade das instituições escolares.

As oportunidades extraescolares são captadas, principalmente, pelo indicador de Ambiente de Aprendizagem em Casa (ACC), o Nível Socioeconômico das famílias (NSE) e ter realizado a creche. Crianças expostas a mais práticas pedagógicas em casa, como leitura de livros e brincar com números, apresentam maior desenvolvimento em linguagem e matemática logo no início da vida escolar com 4 anos de idade. O NSE mais alto das famílias é fortemente associado aos resultados mais positivos e devem ser interpretados como contextos mais protetivos e oferecedor de oportunidades, como exposição à melhores condições de nutrição, saúde e acesso a bens e serviços, impactando direta e indiretamente a qualidade de estímulos no início da vida. A participação na creche, porém, tende a apresentar resultados não muito sólidos, apontando para um pequeno impacto no desenvolvimento, ainda que estatisticamente significativo. Estudos mais aprofundados sobre os impactos desses fatores no desenvolvimento cognitivo na primeira infância devem ser estimulados e conduzidos para um melhor entendimento deles nos mais diversos contextos brasileiros. Mais do que pesquisas que foquem a pré-escola, são necessários estudos que busquem compreender os mecanismos de desigualdade e desenvolvimento cognitivo que se estruturam ao longo da primeiríssima infância, isto é, dos 0 aos 3 anos.

No que tange ao tempo de vida, crianças mais velhas tendem a iniciar a escolarização básica obrigatório com estágio de desenvolvimento mais positivo para linguagem e matemática do que as crianças mais novas. É importante ressaltar que a medida utilizada é calculada em meses de idade. Ter vivido apenas alguns meses a mais do que outras crianças é indicativo de substancial desenvolvimento cognitivo em matemática e linguagem. Isso é visto como uma evidência do Período Sensível do desenvolvimento humano que ocorre na primeira infância e torna esse momento da vida especialmente potencial para aprendizagem (SINGER, 1995; KOLB et al., 2013).

Por fim, quanto ao primeiro objetivo de mensurar desigualdade educacional no início da Educação Básica obrigatória, fica evidenciado que a divisão pública e privada estabelece uma seletividade de perfis disparees entre as redes, estabelecendo um sistema educacional altamente segregado e desigual. As análises dos estratos da rede privada apontam, ainda, que a estratificação não é apenas entre a rede pública e privada, mas, também, é expressiva e interna à própria rede privada, demarcada, principalmente, pela mensalidade das instituições privadas.

A desigualdade educacional no início da Educação básica obrigatória é multifatorial e sua busca por redução dessa desigualdade também exigirá intervenções de diversas ordens. Uma vez que famílias mais vulneráveis carecem de maiores condições protetivas e acesso a bens e serviços de qualidade, é importante reafirmar a relevância do poder público na garantia de direitos e proteções sociais a esse perfil, observado aqui, em especial, as famílias das crianças matriculadas em escola pública.

Os resultados apontam que as oportunidades logo nos três primeiros anos de vida são definidoras de uma desigualdade assoladora no início da Educação Básica obrigatória (aos 4 anos de idade/pré-escola). Há fortes evidências que programas de acesso a pré-natal e sistema de saúde, assim como saneamento básico impactam o desenvolvimento infantil (ABUCHAIM, 2016). Políticas intersetoriais devem ser incentivadas, em especial as que envolvam garantias de saúde e pleno desenvolvimento infantil, atrelados a práticas de estímulo cognitivo em ambientes extraescolares, como ações pedagógicas em ambientes socializadores e do lar. Isto é, políticas, empreendimentos e incentivos que ensinem e estimulem pais e familiares a terem práticas pedagógicas no lar e contextos extraescolares com seus filhos, envolvendo leitura, contagem e interações de estímulo cognitivo. Essas práticas, infelizmente, são mais produzidas por famílias de maiores condições socioeconômicas, como vista nos resultados da dissertação. Esse conjunto de intervenções deve mirar a redução das desigualdades educacionais produzidas ao longo dos primeiros três anos de vida, tornando mais comum estratégias de estímulo cognitivo na primeiríssima infância e diminuindo a correlação dessas práticas e acesso a bens, serviços e políticas com níveis socioeconômicos das famílias.

Lidar com a desigualdade cognitiva no início da vida pode vir a ser um dos mais eficazes mecanismos de redução da desigualdade educacional. Como já visto, a primeira infância é um Período Sensível e, tanto os estímulos negativos quanto os positivos são geradores de impactos profundos e com alta probabilidade de permanência ao longo da vida. É preciso que as políticas públicas para essa faixa etária sejam compreendidas não apenas como urgentes, mas, também, pela alta eficácia potencial e baixo custo, diminuindo a necessidade de remediar problemas de desigualdade futuro, que são muito mais dispendiosos e complexos de se combater (HECKMAN, 2006).

No que se refere aos resultados do segundo objetivo da dissertação, isto é, o impacto de ofertas pública e privada de ensino no desenvolvimento cognitivo ao longo do primeiro ano de escolarização obrigatória, as análises apontaram para um aprendizado mais acelerado das crianças matriculadas na rede privada no desenvolvimento cognitivo em linguagem e matemática, indicando um aumento nas desigualdades educacionais na amostra estudada ao final do primeiro ano da pré-escola.

As análises descritivas retratam que, ao final do primeiro ano da pré-escola, mais da metade das crianças matriculadas em rede pública não havia aprendido os conhecimentos de linguagem e matemática que a maior parte (acima da mediana) das crianças matriculadas na rede privada já sabiam logo que iniciaram o ano letivo. Ou seja: ainda que as crianças matriculadas em escola pública tenham aprendido ao longo de um ano letivo da pré-escola, esse aprendizado (em linguagem e matemática) não foi suficiente para suprimir o tamanho da desigualdade educacional estabelecida no início do ano letivo, isto é, a desigualdade de início da vida escolar. Esse cenário reforça o quão profundo é a desigualdade de início da vida escolar, podendo explicar grande parte da desigualdade escolar ao longo de toda a Educação Básica.

Ademais, as crianças matriculadas em instituições privadas tiveram, em média, um aprendizado mais acelerado do que as crianças de escola pública. Ao comparar a distância entre o desenvolvimento médio em linguagem e matemática no início e final do ano letivo, os dados descritivos apontam que, após um ano de pré-escola, houve um aumento na distância entre as médias da rede pública e privada na ordem dos 25% tanto para matemática quanto para linguagem. Ou seja, as desigualdades de início da vida escolar se tornaram ainda maiores após o primeiro ano da escolarização obrigatória, devido aos diferentes ritmos médios de aprendizagem das crianças matriculadas em ambas as redes.

Esse aprendizado mais acelerado das crianças matriculadas em escola privada e, portanto, um aumento da desigualdade educacional de início, é evidenciado através das análises de regressão hierárquica de valor agregado. Os modelos seguiam o mesmo padrão das análises anteriores, contudo os resultados cognitivos em linguagem e matemática do final do ano letivo (2ª onda) compunham as variáveis dependentes, enquanto os resultados de 1ª onda eram adicionados às variáveis independentes, estabelecendo um modelo de análise de valor agregado.

Os dados sobre o aprendizado no primeiro ano da escola em linguagem sugerem que as crianças matriculadas em escola privada aprendem o equivalente a 4 meses (ES 0.314) a mais do que crianças matriculadas na rede pública da amostra estudada, enquanto que, para matemática, aprendem o equivalente a 7 meses escolares (ES 0.588) a mais. Esses são os efeitos médios isolados da rede privada, já controlado por todas as outras covariáveis que compõem a análise, apontando que as escolas privadas estudadas propiciam uma aprendizagem em linguagem e matemática mais acelerada para crianças de pré-escola. Porém, é importante ressaltar uma ameaça que foi pontuada ao longo de toda a dissertação. A amostra da rede privada está com maior número de escolas de alta mensalidade, o que pode gerar uma superestimação da magnitude média do efeito das escolas privadas sobre a aprendizagem. Se a amostra de escolas privadas tivesse sido aleatória e representativa do município estudado, é provável que a direção do efeito continuasse o mesmo, mas a magnitude fosse menor.

Outros fatores se apresentaram importantes para uma aprendizagem mais acelerada tanto em linguagem quanto em matemática, observados através da análise hierárquica de valor agregado e controlado por todas as outras covariáveis. Crianças um desvio padrão mais velha (em idade em meses), aprendem, em média, 3 meses (ES 0.184) a mais em linguagem e matemática do que seus pares. Crianças com indicador de Ambiente de Aprendizagem em Casa (AAC) um desvio padrão acima, têm um aprendizado em linguagem equivalente a 2 meses (ES 0.102) escolares. Para matemática, contudo, o AAC não é estatisticamente significativo, o que pode apontar para a ideia de que conceitos matemáticos são menos abordados no ambiente do lar, auxiliando compreender a maior variância e desigualdade em matemática do que em linguagem.

O nível socioeconômico (NSE) é significativo apenas para matemática, indicando um aprendizado acelerado em 2 meses (ES 0.139). Isso pode apontar que matemática é um elemento abordado majoritariamente por famílias mais bem estruturadas no contexto extraescolar, reforçando o fenômeno de uma desigualdade mais intensa para matemática do que para linguagem. Além disso, o nível socioeconômico médio da escola, variável utilizada como *proxy* de efeito composição social/"efeito par" da instituição, também apresenta impacto em matemática. Instituições escolares um desvio padrão abaixo da média no nível socioeconômico médio escolar apresentariam 4 meses (ES 0.293) a menos de aprendizado em matemática.

A composição do alunado é vastamente investigada na literatura educacional. A evidência internacional aponta que separar alunos em desvantagem potencial tende a ampliar as desigualdades, influenciando na forma com que são tratados, a qualidade do ensino, assim como as aspirações para etapas subsequentes do processo de escolarização (HAAHR et al., 2005; EUROPEAN GROUP, 2005; JENKS et al., 1972; ROSENTHAL; JACOBSON, 1968; BARTHOLO; COSTA; 2016; BARTHOLO et al., 2020). O cenário que se aponta nessa dissertação é uma intensa estratificação entre redes escolares, gerando não apenas uma desigualdade escolar no início da escolarização obrigatória como, também, criam contextos escolares pouco heterogêneos, alargando as desigualdades ao longo do tempo.

No que tange à influência do sexo na aprendizagem, os achados da dissertação reforçam o que foi observado pelo estudo Geres, assim como a diferença verificada em avaliações padronizadas, como o Saeb e PISA (BROOKE et al, 2014). Ainda que não haja diferenças em matemática entre meninos e meninas no início da escolarização, os meninos aprendem, em média, 2 meses escolares (ES 0.111) a mais de matemática do que meninas ao longo de um ano de escolarização na pré-escola. Ou seja, a evidência aqui aponta que parte do que podemos verificar como desigualdades de sexo ou gênero já se iniciam ainda no começo da vida escolar. É preciso compreender de forma mais profunda por estudos e revisões da literatura internacional e nacional quais os fatores explicariam que meninas tenham maiores resultados em linguagem – mas não há indicativo de maior aprendizagem - e meninos maiores aprendizagens em matemática. Haveria práticas pedagógicas diferenciadas entre meninos e meninas dentro de escolar que justificaria essa discrepância? A diferença é substancialmente cultural ou haveria explicação de cunho biológico para tal desigualdade? Essas são perguntas que a presente dissertação não busca responder, tampouco houve tempo hábil para investigar profundamente. Contudo, o que se evidencia é que parte da desigualdade educacional entre gêneros podem ser compreendidas ainda nos primeiros anos de vida e se destacam já no início da Educação Básica obrigatória.

Outro ponto a se jogar luz é a influência da idade. Crianças um desvio-padrão mais velhas do que a média, aprenderam o equivalente a 2 a 3 meses escolares a mais em linguagem e matemática do que seus pares. Ou seja, o que se observa aqui é uma evidência de maturação biológica como elemento importante da aprendizagem. Diferenças na idade de poucos meses

pode gerar importantes discrepâncias na velocidade de aprendizagem das crianças e evidenciam a janela de oportunidade biológica a que a pré-escola está lidando: o Período Sensível.

Se, por um lado, a primeira etapa da dissertação evidenciou o tamanho da desigualdade no início da vida escolar, por outro, a segunda etapa de análise apontou o seu alargamento, estratificando ainda mais as crianças matriculadas em rede pública e privada. Novamente, o que se vê é a importância de fatores extraescolares envolvidos no desenvolvimento cognitivo de crianças de 4 a 5 anos. Contudo, para além de estabelecerem um efeito de desigualdade de início, indicam, também, um efeito mais permanente e contínuo: o efeito sobre a aprendizagem.

Ao olhar, por exemplo, o efeito de 1 mês do Ambiente de Aprendizagem em Casa (AAC) e 2 meses do nível socioeconômico (NSE) é plausível que se pense em um primeiro momento enquanto um efeito pequeno na aprendizagem. Contudo, se esses efeitos forem acumulativos ao longo dos anos, serão fatores extraescolares de suma importância, impactando profundamente nas desigualdades educacionais ao passar dos anos.

Isso aponta para a necessidade de uma rede de proteção contra os efeitos perversos da vulnerabilidade social. Políticas intersetoriais devem dar continuidade na rede de proteção da primeira infância, ampliando as oportunidades de acesso a bens e serviços que dificilmente famílias vulneráveis pouco teriam possibilidade de experienciar por si.

Outro ponto que enfaticamente deve ser mirado é a estratificação educacional estabelecido pela divisão público e privada. É sabido que há elementos de estratificação educacional dentro da própria rede pública, contudo a adição de identificadores de rede privada aponta uma estratificação ainda mais intensa entre ambas as redes do que interna a elas (BARTHOLO et al., 2020; BARTHOLO; COSTA, 2016). A dissertação vai na mesma direção deste achado e na contramão de evidências que apontam que há apenas desigualdade entre a rede pública e privada (ENGEL DE ABREU et al, 2015).

O agrupamento de crianças do mesmo perfil cria um estágio de “bola de neve”, ao agregar, na média, crianças com vulnerabilidades na escola pública e crianças com vantagens sociais na rede privada. Esse agrupamento de “desvantagens” torna o ambiente escolar mais

desafiante para a rede pública, diminuindo o ritmo de aprendizado e, possivelmente, gerando uma desvantagem “permanente” ao longo dos anos para escolas da rede pública.

Uma vez que no Brasil, diferentemente de diversos outros países, a rede privada é grande, agregando aproximadamente um quarto das matrículas da pré-escola, as políticas de combate de desigualdade educacional devem levar em consideração o perfil de discente das escolas privadas. Como a escolha por instituições educacionais é de liberdade dos pais – e aqui defende-se que deve continuar sendo – é desafiante pensar soluções de políticas públicas que ampliem a heterogeneidade da rede pública. É primordial, porém, atrair o desejo dos pais em matricular seus filhos na rede pública, em especial daqueles que têm a possibilidade de matricular seus filhos em escolas privadas. Isso, contudo, passa por tornar as instituições públicas mais atrativas, ampliando mecanismos de segurança, estruturando políticas de envolvimento familiar com comunidade escolar e comunicação família-escola, delineando currículos claros e acessíveis e possibilitando que os pais compreendam as práticas pedagógicas realizadas na instituição. Um processo que aumente a confiança civil para com as escolas públicas. Isso passa por investimento em insumos e garantias de processos escolares de qualidade, indo desde a infraestrutura até a qualidade da formação docente. Um enorme desafio enquanto política pública para o cenário brasileiro.

É importante ressaltar que a interpretação dos resultados da dissertação *não* aponta para uma falta de importância das instituições escolares públicas frente às instituições da rede privada. Essa seria a interpretação mais rasa e incorreta possível desse estudo. Apesar do foco da dissertação ser impacto das redes na aprendizagem em linguagem e matemática, a rede pública deve ser compreendida como uma importante rede de proteção social e de garantia de direitos às crianças e adolescentes. A rede privada não agrega essa função e tampouco se propõe a tal.

O que os resultados aqui evidenciam é que, apesar das diversas vulnerabilidades sociais, um agregado de desafios socioculturais e de dificuldade de acesso a bens, serviços e políticas públicas, crianças matriculadas da rede pública *aprendem*. A escola pública faz diferença e tem importância primordial para garantia de um cenário republicano que visa distribuir de forma mais justa as oportunidades. Como a matrícula em instituições privadas está associada a uma aprendizagem mais acelerada, gera-se um cenário de ampliação das desigualdades educacionais.

Contudo, é plausível pensar que a inexistência de uma rede de proteção centrado na escola pública como vemos hoje, geraria desigualdades educacionais ainda mais profundas.

Esse último ponto é importante, pois a presente dissertação foi escrita ao longo da pandemia de Covid-19, entre 2020 e 2021. Pela primeira vez na história recente as escolas foram fechadas por todo o mundo, obrigando estudantes, sejam eles ricos ou pobres, a ficarem em suas casas e não terem acesso às instituições escolares presencialmente. Se as desigualdades já se ampliam mesmo em um cenário normal de acesso escolar, é provável que se exacerbem ainda mais sem a possibilidade de acesso a toda a rede de proteção gerada pelo acesso presencial às escolas públicas, como alimentação, proteções de saúde, ambiente de socialização e aprendizagem pedagogicamente elaboradas.

Apesar dos achados inéditos da dissertação para a literatura nacional, o estudo aqui desenvolvido não tem o intuito e sequer é capaz de esgotar o debate sobre as desigualdades educacionais no início da escolarização. Aconselha-se fortemente para que sejam empreendidos novos estudos, em especial em outras regiões do país que não as centralizadas no Sudeste, como o presente trabalho. Além disso, ressalta-se que há poucos dados sobre qualidade da Educação Infantil no Brasil, uma vez que, diferentemente do Ensino Fundamental e Médio, não há uma política nacional de dados robustos sobre qualidade educacional para a pré-escola. Isso impõe uma dificuldade para compreender os mecanismos de desigualdade educacional no início da vida, fator importante para compreender as desigualdades ao longo de toda a Educação Básica e, por conseguinte, nas desigualdades sociais e econômicas. Sendo assim, é importante que haja políticas nacionais de monitoramento da qualidade da pré-escola, indo para além de dados de insumos escolares, mas, também, de resultados de desenvolvimento escolar.

Diversos são os trabalhos futuros que são propostos a serem realizados utilizando os dados do estudo longitudinal executado pelo LaPOpE/UFRJ e que possibilitaram a realização da presente dissertação. O primeiro deles é de uma interpretação pedagógica dos resultados. Apesar de ter lançado mão de uma interpretação dos resultados de effect size em meses de aprendizagem, há pouca clareza do que isso signifique para a prática educacional. Contudo, os dados utilizados na dissertação permitem identificar quais saberes de matemática e linguagem as crianças

compreendem no início da educação básica obrigatória e quais os conhecimentos elas aprendem, como contagem de objetos, reconhecimento de formas, soma, riqueza de vocabulário, reconhecimento de letras e palavras. Dessa forma, é possível contribuir para um debate mais profundo do como se comportam as desigualdades no desenvolvimento escolar das crianças. Um segundo ponto é adicionar mais variáveis de nível dois nos modelos multiníveis, em especial o Ambiente de Aprendizagem em Casa médio de cada escola, possibilitando captar um efeito de composição discente que reflita influências de forma mais profundas dos fatores familiares no desenvolvimento infantil ao longo da pré-escola. Por fim, um último ponto proposto é o de incluir análises com maior recorte longitudinal e ir para além do primeiro ano da pré-escola, verificando, também, o como as diferentes redes impactam na desigualdade e aprendizagem ao longo do segundo ano da pré-escola, sendo essas informações também coletadas pelo estudo executado pelo LaPOE/UFRJ. Desse modo, os trabalhos futuros propostos podem auxiliar em uma melhor compreensão das desigualdades educacionais no início da vida escolar e sua produção.

BIBLIOGRAFIA

ABUCHAIM et al. Importância dos vínculos familiares na primeira infância. Comitê Científico Núcleo Ciência pela Infancia. 2016

AGUIAR, D. K. ; BARTHOLO, T. L. . A associação entre a aptidão física não-aeróbica e o desempenho cognitivo de crianças na pré-escola. Revista ciência & desenvolvimento, v. 12, p. 725-740, 2019.

ALVES, M. T. G., Caracterização das desigualdades educacionais com dados públicos: desafios para conceituação e operacionalização empírica. Lua Nova (Impresso), v. 110, p. 1-28, 2020.

ALVES, Maria Teresa Gonzaga; SOARES, José Francisco; XAVIER, Flavia Pereira. Desigualdades educacionais no ensino fundamental de 2005 a 2013: hiato entre grupos sociais. Revista Brasileira de Sociologia, v. 4, n. 7, p. 49-82, 2016.

AMORIM, A. N. G. F., DIAS, N. M., ALBUQUERQUE, E. X. D. S., SILVA, V. C. D., FALCÃO, A. C. G. P., GUERRA, V. G. R., SILVA, M. H. D. AND SANTOS, L. L. D. AMORIM, A., DIAS, N., ALBUQUERQUE, E., SILVA, V., FALCÃO, A., & GUERRA, V. ET AL. (2019). Initial reading and writing skills in childhood education: achievement sample in the Northeast of Brazil for obtaining specific regional performance standards. Revista CEFAC, 21(5). doi: 10.1590/1982-0216/201921513219

BARROS R. P., CARVALHO M., FRANCO S, MENDONÇA R, ROSALÉM, A. Uma avaliação do impacto da qualidade da creche no desenvolvimento infantil. Pesquisa Planejamento Econômico. 2011;41(2).

BARTHOLO, T. L.; COSTA, M. . Evidence of a school composition effect in Rio de Janeiro public schools. Ensaio (Rio de Janeiro. Online), v. 24, p. 498-521, 2016.

BARTHOLO, T. L.; KOSLINSKI, M. C ; COSTA, M. ; BARCELLOS, T. M. . What do children know upon entry to pre-school in Rio de Janeiro?. ENSAIO (RIO DE JANEIRO. ONLINE), p. 1-22, 2019.

BARTHOLO, T. L.; KOSLINSKI, M. C ; COSTA, M. ; TYMMS, P. B. ; MERRELL, C. ; BARCELLOS, T. M. . The use of cognitive instruments for research in early childhood

education: constraints and possibilities in the Brazilian context. *PRÓ-POSIÇÕES (UNICAMP. ONLINE)*, v. 31, p. 1-24, 2020.

BONAMINO, A. M. C. de, OLIVEIRA, L. H. G. de. "Estudos longitudinais e pesquisa na educação básica" v. 19, n. 38, p. 33–50, 2 maio 2013. DOI: 10.26512/lc.v19i38.4084.

BONAMINO, A. M. C. DE; OLIVEIRA, L. H. G. DE. Estudos longitudinais e pesquisa na educação básica. *Linhas Críticas*, v. 19, n. 38, p. 33–50.

BOONE, William J. Rasch analysis for instrument development: why, when, and how?. *CBE—Life Sciences Education*, v. 15, n. 4, p. rm4, 2016.

BOUDON, R., *A Desigualdade das Oportunidades. A Mobilidade Social nas Sociedades Industriais*. Editora Universidade de Brasília. 1981.

BRONFENBRENNER, U., & MORRIS, P. A. The Bioecological Model of Human Development. In R. M. Lerner & W. Damon (Eds.), *Handbook of child psychology: Theoretical models of human development* (pp. 793-828), 2006

BROOKE, N. et al. Modelagem do crescimento da aprendizagem nos anos iniciais com dados longitudinais da pesquisa GERESI. *Educacao e Pesquisa*, v. 40, n. 1, p. 77–94, 2014.

BRYK, A.S., RAUDENBUSH, S.W. *Hierarchical linear models (applications and data analysis methods)*. Sage Publications, Thousand Oaks. 1992.

BUUREN, S. V., *Flexible Imputation of Missing Data (2nd Edition)*. 2018

BUUREN, S. V.; GROOTHUIS-OUDSHOORN, K.. Mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3), 1-67. URL <https://www.jstatsoft.org/v45/i03/>, 2011

CAMPOS, M. M. et al. A contribuição da educação infantil de qualidade e seus impactos no início do ensino fundamental. *Educação e Pesquisa*, São Paulo, v. 37, n.1, p. 15-33, 2011.

CAMPOS, M., FÜLLGRAF, J., & WIGGERS, V. A qualidade da educação infantil brasileira: alguns resultados de pesquisa. *Cadernos De Pesquisa*, 36(127), 87-128. 2006. doi: 10.1590/s0100-15742006000100005

CAMPOS, Maria Malta et al. A qualidade da educação infantil: um estudo em seis capitais Brasileiras. *Cad. Pesqui.* [online]. 2011, vol.41, n.142, pp.20-54. ISSN 0100-1574. <https://doi.org/10.1590/S0100-15742011000100003>.

CANO, I.. *Introdução à Avaliação de Programas Sociais*. Rio de Janeiro. FGV Editora, 2004

Cognitive outcomes in year 6. 2008.

Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., & York, R. (1966). *Equality of educational opportunity*. Washington, DC.

CRAHAY, M., BAYE, A. "Existem escolas justas e eficazes?", *Cadernos de Pesquisa*, v. 43, n. 150, p. 858–883, 2013. DOI: 10.1590/S0100-15742013000300007

CURY, C. R. J. "Direito à educação: direito à igualdade, direito à diferença", *Cadernos de Pesquisa*, v. 116, n. 116, p. 245–262, jul. 2002. DOI: 10.1590/S0100-15742002000200010.

DALBEN, A. *Fatores associados à proficiência em leitura e matemática: uma aplicação do modelo linear hierárquico com dados longitudinais do projeto geres campinas*, 2014.

DAMIANI MF, DUMITH S, HORTA BL, GIGANTE D. Educação infantil e longevidade escolar: dados de um estudo longitudinal. *Estudos Avaliação Educacional*. 2011;22(50):515–532.

ENGEL DE ABREU P, TOURINHO C, PUGLISI M, NIKAEDO C, ABREU N, MIRANDA M ET AL. *A pobreza e a mente: perspectiva da ciência cognitiva*. University of Luxembourg. 2015.

EUROPEAN GROUP FOR RESEARCH ON EQUITY IN EDUCATIONAL SYSTEMS. *Equity in European educational systems: A set of indicators*. *European Educational Research Journal*, 4(2), 1215-1276. 2005.

EVANS, GW; KIM, P., Cumulative Risk Exposure and Stress Dysregulation. *Psychological Science*. 18(11):953–958. 2007

FRANCO, C.; BROOKE, N.; ALVES, F. Estudo longitudinal sobre qualidade e equidade no ensino fundamental brasileiro: GERES 2005. *Ensaio*, v. 16, n. 61, p. 625–637, out. 2008.

FRANKLIN, Tamara B.; SAAB, Bechara J.; MANSUY, Isabelle M. Neural mechanisms of stress resilience and vulnerability. *Neuron*, v. 75, n. 5, p. 747-761, 2012.

FUNDAÇÃO MARIA CECÍLIA SOUTO VIDIGAL. Os avanços da primeira infância no Brasil no últimos cinco anos. 2020

GIRALDI, L. P. B.; SIGOLO, S. R. R. L. PERSPECTIVA LONGITUDINAL DE PESQUISA EM EDUCAÇÃO: produções brasileiras e a organização e implementação de um estudo deste perfil. *Atos de Pesquisa em Educação*, v. 11, n. 1, p. 2, 2 maio 2016.

GOLDSTEIN, H. Methods in School Effectiveness Research. *School Effectiveness and School Improvement*, v. 8, n. 4, p. 369–395, 1997.

GRUND, S.; LUDTKE, O.; ROBITZSCH, A. Multiple imputation of missing data for multilevel models: Simulations and recommendations. *Organizational Research Methods*, 2018

HECKMAN, James J. Skill formation and the economics of investing in disadvantaged children. *Science*, v. 312, n. 5782, p. 1900-1902, 2006.

HIGGINS, S., KATSIPATAKI, M., COLEMAN, R., HENDERSON, P., MAJOR, L.E., & COE, R.. The Sutton Trust-Education Endowment Foundation Teaching and Learning Toolkit. October 2014. London: Education Endowment Foundation.

HIGGINS, Steve; KOKOTSAKI, Dimitra; COE, Robert. The teaching and learning toolkit. Education Endowment Foundation and Sutton Trust, 2012.

HONORATO, G., Problemas crônicos do Ensino Superior no Brasil e desafios para os próximos anos: desigualdade horizontal, análise de egressos e financiamento.. In: OLIVEIRA, A. J. B.; HONORATO, G.; Desafios para o ensino superior brasileiro no contexto contemporâneo. Coleção de Cadernos do LEPES. V.3. Faculdade de Educação/UFRJ. 2020

INEP – Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Censo Educacional (2010 - 2019). Brasil. 2021.

INFÂNCIA, NÚCLEO CIÊNCIA PELA. O impacto do desenvolvimento na primeira infância sobre a aprendizagem. Comitê Científico do Núcleo Ciência pela Infância, 2014.

IPIPS. Placing early childhood education at the heart of worldwide policy making. Disponível em: <<http://www.ipips.org/>>. Acesso em 15 de maio de 2021.

JENCKS, C. Inequality: a reassessment of the effect of family and schooling in America. London: Lowe & Brydone, 1972.

KHINE, M. S. Rasch Measurement. Springer Singapore, 2020.

KLEIN, R., A Pré-Escola no Brasil. Trabalho apresentado na II Reunião da Abave, Fortaleza, 16-18 de novembro, 2006

KOLB B, MYCHASIUK R, MUHAMMAD A, GIBB R. Brain plasticity in the developing brain. Progress in Brain Research. 2013;207:35–64. doi:10.1016/B978-0-444-63327-9.00005-9.

KOSLINSKI, M. C ; BARTHOLO, T. L. . Desigualdades de oportunidades educacionais no início da trajetória escolar no contexto brasileiro. Lua Nova (Impresso), p. 215-245, 2020.

KOSLINSKI, M. C. et al. Ambiente de aprendizagem em casa e o desenvolvimento cognitivo na educação infantil. Revista Educação e Sociedade. 2021 PRELO

KOSLINSKI, Mariane C.; BARTHOLO, Tiago Lisboa. Impacto dos espaços de desenvolvimento infantil no primeiro ano na pré-escola. Estudos em Avaliação Educacional, v. 30, n. 73, p. 280-311, 2019.

KRAMER, Sonia; KAPPEL, Maria Dolores Bombardelli; FUNDAÇÃO IBGE. Educação da criança de 0 a 6 anos. pesquisa sobre padrões de vida (PPV): primeira infância (1996/1997). Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística, 2000.

LEE, V. E. Dados longitudinais em educação: um componente essencial da abordagem de valor agregado no que se refere à avaliação de desempenho escolar. Estudos em Avaliação Educacional, v. 21, n. 47, p. 431, 30 dez. 2010.

LINACRE, J. M. Many-Facet Rasch Measurement. 1994

OBSERVATÓRIO DO PNE. Educação Infantil. Disponível em: <<http://www.observatoriodopne.org.br/metas-pne/1-educacao-infantil>>. Acesso em 18 de maio de 2021.

OLIVEIRA, A. J. B.; HONORATO, G.; Desafios para o ensino superior brasileiro no contexto contemporâneo. Coleção de Cadernos do LEPES. V.3. Faculdade de Educação/UFRJ. 2020

PEDRAS CTPDA, GERALDO T, CREMITTE PAP. Consciência fonológica em crianças de escola pública e particular. Revi Soc Bras Fonoaudiol. 2006;11(2):65-9.

PEISNER-FEINBERG, Ellen S. et al. The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. Child development, v. 72, n. 5, p. 1534-1553, 2001.

PICANÇO, Felícia; MORAIS, Juliana. Estudos sobre estratificação educacional: síntese dos principais argumentos e desdobramentos. Educação & Sociedade, v. 37, p. 391-406, 2016.

PROCÓPIO, I. V.; DA SILVA FREGUGLIA, R.; CHEIN, F. Desigualdade de oportunidades na formação de habilidades: Uma análise com dados longitudinais. Economia Aplicada, v. 19, n. 2, p. 326–348, 2015.

RAUDENBUSH, S. W.; WILLMS, JD. The Estimation of School Effects. Journal of Educational and Behavioral Statistics, v. 20, n. 4, p. 307–335, 23 dez. 1995.

RAUDENBUSH, Stephen W. Educational applications of hierarchical linear models: A review. journal of Educational Statistics, v. 13, n. 2, p. 85-116, 1988.

RAWLS JR, Thomas M.. Theroy of Justice. U. Pa. L. Rev., v. 121, p. 1020, 1972.

REARDON, S. F.; RAUDENBUSH, S. W. Assumptions of Value-Added Models for Estimating School Effects. Education Finance and Policy, v. 4, n. 4, p. 492–519, out. 2009.

RIBEIRO, Carlos Antônio Costa. Desigualdade de oportunidades e resultados educacionais no Brasil. Dados, v. 54, p. 41-87, 2011.

ROSENTHAL, Robert; JACOBSON, Lenore. Pygmalion in the classroom. The urban review, v. 3, n. 1, p. 16-20, 1968.

RUBIN, Donald B. The calculation of posterior distributions by data augmentation: Comment: A noniterative sampling/importance resampling alternative to the data augmentation algorithm for

creating a few imputations when fractions of missing information are modest: The SIR algorithm. *Journal of the American Statistical Association*, v. 82, n. 398, p. 543-546, 1987.

SAMMONS, P. et al. Influences on children's attainment and progress in key stage 2:

Shonkoff, J. P., & Phillips, D. A. (Eds.). (2000). *From neurons to neighborhoods: The science of early childhood development*. National Academy Press

SILVA, Nelson do Valle. Expansão escolar e estratificação educacional no Brasil. *Origens e destinos: desigualdades sociais ao longo da vida*, p. 105-146, 2003.

SIMÕES, Patrícia Maria Uchôa; LIMA, Juceli Bengert. Infância, educação e desigualdade no Brasil. *Revista Iberoamericana de educación*, v. 72, p. 45-64, 2016.

SINGER, Wolf. Development and plasticity of cortical processing architectures. *Science*, v. 270, n. 5237, p. 758-764, 1995.

SOARES, José Francisco; DELGADO, Victor Maia Senna. Medida das desigualdades de aprendizado entre estudantes de ensino fundamental. *Estudos em Avaliação Educacional*, v. 27, n. 66, p. 754-780, 2016.

SOARES, T. M. et al. Modelos de valor agregado para medir a eficácia das escolas Geres. *Ensaio*, v. 25, n. 94, p. 59-89, 2017.

SOARES, T. M. Utilização da teoria da resposta ao item na produção de indicadores econômicos. *Pesquisa Operacional*, Rio de Janeiro, v. 25, n. 1, p. 83-112, 2005.

SYLVA, Kathy et al. (Ed.). *Early childhood matters: Evidence from the effective pre-school and primary education project*. Routledge, 2010.

SYLVA, Kathy et al. Capturing quality in early childhood through environmental rating scales. *Early childhood research quarterly*, v. 21, n. 1, p. 76-92, 2006.

TAGGART, Brenda et al. O poder da pré-escola: evidências de um estudo longitudinal na Inglaterra. *Cadernos de Pesquisa*, v. 41, p. 68-99, 2011.

TYMMS, P. Effect sizes in multilevel models. En I. Schagen e K. Elliot (Orgs.), But what does it mean? The use of effect sizes in educational research (pp. 79-112). National Foundation for Educational Research. 2004

TYMMS, P., MERRELL, C., HAWKER, D., et al. "Performance Indicators in Primary Schools: A comparison of performance on entry to school and the progress made in the first year in England and four other jurisdictions", 2014.

TYMMS, Peter et al. The first seven years at school. Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education), v. 21, n. 1, p. 67-80, 2009

Veloso, F., 15 anos de avanços na educação no Brasil: onde estamos? In: Veloso, F.; Pessoa, S.; Henriques, R.; Giambiagi, F. Educação Básica no Brasil: construindo o país do futuro. Rio de Janeiro: Elsevier, p. 3-24. 2009

ANEXOS

ANEXO - DADOS OBSERVADOS - Regressão multinível - Fatores associados à estratificação no início da Educação Básica obrigatória (1º ano da pré-escola) considerando a rede pública e privada.

Variável dependente: Desenvolvimento em Linguagem no início do 1º ano da pré-escola.

	Linguagem Modelo Nulo	Linguagem Modelo 1	Linguagem Modelo 2	Linguagem Modelo 3	Linguagem Modelo 4	Linguagem Modelo 5	Linguagem Modelo 6
(Intercepto)	-0.14 * (-0.26 -0.01)	-0.45 *** (-0.54 -0.35)	-4.33 *** (-4.84 -3.83)	-4.22 *** (-4.73 -3.72)	-4.30 *** (-4.87 -3.72)	-4.27 *** (-4.89 -3.64)	-4.29 *** (-4.90 -3.68)
Escola:Privada		0.89 *** (0.72 1.06)	0.89 *** (0.72 1.06)	0.89 *** (0.73 1.06)	0.95 *** (0.78 1.12)	0.72 *** (0.55 0.90)	0.47 *** (0.29 0.64)
Idade (meses)			0.07 *** (0.06 0.08)	0.07 *** (0.06 0.08)	0.07 *** (0.06 0.08)	0.07 *** (0.06 0.08)	0.07 *** (0.06 0.09)
Sexo:Masculino				-0.16 *** (-0.23 -0.10)	-0.18 *** (-0.25 -0.10)	-0.19 *** (-0.27 -0.11)	-0.21 *** (-0.29 -0.12)
Creche:Frequentou					0.11 ** (0.03 0.20)	0.06 (-0.04 0.15)	0.05 (-0.04 0.15)
Indicador de Ambiente de Aprendizagem em Casa						0.12 *** (0.09 0.15)	0.07 *** (0.04 0.10)
Indicador de Nível socioeconômico (Indivíduo)							0.14 *** (0.11 0.17)
Random Effects							
σ^2	0.97	0.97	0.90	0.90	0.89	0.83	0.80
τ_{00}	0.27 _{ESC}	0.09 _{ESC}	0.08 _{ESC}	0.08 _{ESC}	0.04 _{ESC}	0.04 _{ESC}	0.03 _{ESC}
ICC	0.22	0.08	0.08	0.08	0.04	0.04	0.03
N	75 _{ESC}	75 _{ESC}	75 _{ESC}	75 _{ESC}	69 _{ESC}	69 _{ESC}	69 _{ESC}
Observations	3261	3261	3156	3156	2380	1922	1910
Marginal R ² /	0.000 /	0.096 /	0.153 /	0.159 /	0.145 /	0.190 /	0.225 /
Conditional R ²	0.216	0.169	0.220	0.227	0.183	0.225	0.250

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

ANEXO - DADOS OBSERVADOS - Regressão multinível - Fatores associados à estratificação no início da Educação Básica obrigatória (1º ano da pré-escola) considerando a rede pública e privada.
Variável dependente: Desenvolvimento em Matemática no início do 1º ano da pré-escola.

	Matemática Modelo Nulo	Matemática Modelo 1	Matemática Modelo 2	Matemática Modelo 3	Matemática Modelo 4	Matemática Modelo 5	Matemática Modelo 6
(Intercepto)	-0.21 (-0.43 0.00)	-0.81 *** (-0.94 -0.68)	-6.71 *** (-7.39 -6.03)	-6.70 *** (-7.38 -6.02)	-6.45 *** (-7.22 -5.68)	-6.18 *** (-7.03 -5.33)	-6.20 *** (-7.04 -5.36)
Escola:Privada		1.69 *** (1.45 1.92)	1.70 *** (1.47 1.92)	1.70 *** (1.47 1.92)	1.72 *** (1.47 1.97)	1.44 *** (1.18 1.70)	1.09 *** (0.84 1.35)
Idade (meses)			0.11 *** (0.10 0.12)	0.11 *** (0.10 0.12)	0.10 *** (0.09 0.12)	0.10 *** (0.08 0.12)	0.10 *** (0.09 0.12)
Sexo:Masculino				-0.01 (-0.10 0.08)	-0.05 (-0.15 0.05)	-0.05 (-0.17 0.06)	-0.07 (-0.18 0.04)
Creche:Frequentou					0.14 ** (0.04 0.25)	0.10 (-0.03 0.23)	0.09 (-0.04 0.22)
Indicador de Ambiente de Aprendizagem em Casa						0.15 *** (0.11 0.19)	0.09 *** (0.05 0.13)
Indicador de Nível socioeconômico (Indivíduo)							0.18 *** (0.14 0.23)
Random Effects							
σ^2	1.80	1.80	1.64	1.64	1.59	1.53	1.49
τ_{00}	0.82 _{ESC}	0.17 _{ESC}	0.13 _{ESC}	0.13 _{ESC}	0.12 _{ESC}	0.11 _{ESC}	0.09 _{ESC}
ICC	0.31	0.08	0.08	0.08	0.07	0.07	0.06
N	75 _{ESC}	75 _{ESC}	75 _{ESC}	75 _{ESC}	69 _{ESC}	69 _{ESC}	69 _{ESC}
Observations	3251	3251	3146	3146	2372	1917	1905
Marginal R ² /	0.000 /	0.170 /	0.237 /	0.237 /	0.194 /	0.232 /	0.263 /
Conditional R ²	0.314	0.240	0.295	0.295	0.249	0.284	0.304

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

ANEXO - DADOS OBSERVADOS - Regressão multinível - Fatores associados à estratificação no início da Educação Básica obrigatória (1º ano da pré-escola) considerando os estratos baixa, média e alta mensalidade.

Variável dependente: Desenvolvimento em Linguagem no início do 1º ano da pré-escola.

	Linguagem Modelo Nulo	Linguagem Modelo 1	Linguagem Modelo 2	Linguagem Modelo 3	Linguagem Modelo 4	Linguagem Modelo 5	Linguagem Modelo 6
(Intercepto)	-0.14 * (-0.26 -0.01)	-0.45 *** (-0.51 -0.38)	-4.29 *** (-4.78 -3.79)	-4.18 *** (-4.68 -3.68)	-4.27 *** (-4.84 -3.70)	-4.23 *** (-4.85 -3.61)	-4.26 *** (-4.87 -3.65)
Escola:Baixa mens.		0.28 * (0.03 0.53)	0.23 (-0.01 0.48)	0.23 (-0.02 0.47)	0.34 * (0.05 0.62)	0.15 (-0.13 0.43)	-0.03 (-0.30 0.24)
Escola:Média mens.		0.58 *** (0.36 0.79)	0.61 *** (0.40 0.83)	0.63 *** (0.41 0.84)	0.80 *** (0.53 1.07)	0.62 *** (0.35 0.90)	0.41 ** (0.14 0.68)
Escola:Alta mens.		1.33 *** (1.16 1.50)	1.31 *** (1.15 1.48)	1.31 *** (1.15 1.48)	1.33 *** (1.13 1.53)	1.05 *** (0.85 1.26)	0.75 *** (0.55 0.96)
Idade (meses)			0.07 *** (0.06 0.08)	0.07 *** (0.06 0.08)	0.07 *** (0.06 0.08)	0.07 *** (0.06 0.08)	0.07 *** (0.06 0.09)
Sexo:Masculino				-0.16 *** (-0.23 -0.09)	-0.18 *** (-0.26 -0.11)	-0.19 *** (-0.28 -0.11)	-0.21 *** (-0.29 -0.13)
Creche:Frequentou					0.11 ** (0.03 0.19)	0.04 (-0.05 0.14)	0.04 (-0.06 0.13)
Indicador de Ambiente de Aprendizagem em Casa						0.12 *** (0.09 0.15)	0.07 *** (0.04 0.10)
Indicador de Nível socioeconômico (Indivíduo)							0.13 *** (0.10 0.17)
Random Effects							
σ^2	0.97	0.97	0.90	0.89	0.89	0.83	0.80
τ_{00}	0.27 ESC	0.03 ESC	0.03 ESC	0.03 ESC	0.02 ESC	0.02 ESC	0.01 ESC
ICC	0.22	0.03	0.03	0.03	0.02	0.02	0.01
N	75 ESC	75 ESC	75 ESC	75 ESC	69 ESC	69 ESC	69 ESC
Observations	3261	3261	3156	3156	2380	1922	1910
Marginal R ² /	0.000 /	0.126 /	0.186 /	0.192 /	0.167 /	0.213 /	0.244 /
Conditional R ²	0.216	0.154	0.211	0.217	0.185	0.229	0.255

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

ANEXO - DADOS OBSERVADOS - Regressão multinível - Fatores associados à estratificação no início da Educação Básica obrigatória (1º ano da pré-escola) considerando os estratos baixa, média e alta mensalidade.

Variável dependente: Desenvolvimento em Matemática no início do 1º ano da pré-escola.

	Matemática Modelo Nulo	Matemática Modelo 1	Matemática Modelo 2	Matemática Modelo 3	Matemática Modelo 4	Matemática Modelo 5	Matemática Modelo 6
(Intercepto)	-0.21 (-0.43 0.00)	-0.81 *** (-0.91 -0.71)	-6.65 *** (-7.32 -5.97)	-6.64 *** (-7.31 -5.96)	-6.41 *** (-7.17 -5.64)	-6.14 *** (-6.98 -5.29)	-6.16 *** (-7.00 -5.32)
Escola:Baixa mens.		0.99 *** (0.62 1.36)	0.97 *** (0.61 1.32)	0.96 *** (0.61 1.31)	1.06 *** (0.62 1.50)	0.84 *** (0.41 1.28)	0.59 ** (0.17 1.01)
Escola:Média mens.		1.20 *** (0.88 1.51)	1.29 *** (0.98 1.59)	1.29 *** (0.99 1.59)	1.40 *** (1.00 1.80)	1.17 *** (0.76 1.57)	0.88 *** (0.48 1.28)
Escola:Alta mens.		2.27 *** (2.01 2.52)	2.21 *** (1.98 2.45)	2.21 *** (1.98 2.45)	2.22 *** (1.91 2.53)	1.88 *** (1.56 2.20)	1.47 *** (1.15 1.79)
Idade (meses)			0.11 *** (0.10 0.12)	0.11 *** (0.10 0.12)	0.10 *** (0.09 0.12)	0.10 *** (0.08 0.12)	0.10 *** (0.09 0.12)
Sexo:Masculino				-0.01 (-0.10 0.07)	-0.05 (-0.15 0.05)	-0.06 (-0.17 0.05)	-0.08 (-0.19 0.03)
Creche:Frequentou					0.14 * (0.03 0.25)	0.09 (-0.05 0.22)	0.07 (-0.06 0.20)
Indicador de Ambiente de Aprendizagem em Casa						0.15 *** (0.11 0.19)	0.09 *** (0.05 0.13)
Indicador de Nível socioeconômico (Indivíduo)							0.18 *** (0.14 0.22)
Random Effects							
σ^2	1.80	1.79	1.63	1.63	1.59	1.53	1.48
τ_{00}	0.82 _{ESC}	0.08 _{ESC}	0.07 _{ESC}	0.07 _{ESC}	0.08 _{ESC}	0.08 _{ESC}	0.06 _{ESC}
ICC	0.31	0.04	0.04	0.04	0.05	0.05	0.04
N	75 _{ESC}	75 _{ESC}	75 _{ESC}	75 _{ESC}	69 _{ESC}	69 _{ESC}	69 _{ESC}
Observations	3251	3251	3146	3146	2372	1917	1905
Marginal R ² / Conditional R ²	0.000 / 0.314	0.195 / 0.230	0.261 / 0.290	0.261 / 0.290	0.213 / 0.250	0.253 / 0.288	0.279 / 0.308

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

ANEXO - DADOS OBSERVADOS - Regressão multinível - Fatores associados à aprendizagem no início da Educação Básica obrigatória (1º ano da pré-escola) considerando a rede pública e privada.

Variável dependente: Desenvolvimento em Linguagem no início do 2º ano da pré-escola.

	Linguagem Modelo Nulo	Linguagem Modelo 1	Linguagem Modelo 2	Linguagem Modelo 3	Linguagem Modelo 4	Linguagem Modelo 5	Linguagem Modelo 6	Linguagem Modelo 7	Linguagem Modelo 8	Linguagem Modelo 9
(Intercepto)	0.55 *** (0.40 0.70)	0.15 ** (0.05 0.25)	-3.55 *** (-4.03 -3.07)	-3.48 *** (-3.96 -3.00)	-3.49 *** (-4.03 -2.94)	-3.25 *** (-3.83 -2.68)	-3.27 *** (-3.83 -2.70)	-3.15 *** (-3.72 -2.59)	-0.65 ** (-1.10 -0.20)	-0.08 (-0.51 0.36)
Escola:Privada		1.14 *** (0.97 1.31)	1.13 *** (0.97 1.29)	1.13 *** (0.97 1.29)	1.07 *** (0.88 1.25)	0.86 *** (0.68 1.05)	0.65 *** (0.47 0.83)	-0.00 (-0.31 0.31)	0.00 (-0.21 0.22)	-0.05 (-0.25 0.14)
Idade (meses)			0.07 *** (0.06 0.08)	0.07 *** (0.06 0.08)	0.07 *** (0.06 0.08)	0.07 *** (0.06 0.08)	0.07 *** (0.06 0.08)	0.07 *** (0.06 0.08)	0.02 *** (0.01 0.03)	0.01 *** (0.01 0.02)
Sexo:Masculino				-0.11 *** (-0.18 -0.05)	-0.13 *** (-0.20 -0.05)	-0.11 ** (-0.18 -0.03)	-0.12 ** (-0.19 -0.04)	-0.12 ** (-0.19 -0.04)	0.00 (-0.05 0.06)	-0.00 (-0.06 0.05)
Creche:Frequentou					0.01 (-0.07 0.09)	-0.02 (-0.11 0.07)	-0.02 (-0.11 0.07)	-0.03 (-0.12 0.06)	-0.06 (-0.13 0.00)	-0.06 * (-0.12 -0.00)
Indicador de Ambiente de Aprendizagem em Casa						0.11 *** (0.08 0.14)	0.07 *** (0.04 0.10)	0.07 *** (0.04 0.10)	0.03 * (0.00 0.05)	0.02 (-0.00 0.04)
Indicador de Nível socioeconômico (Indivíduo)							0.11 *** (0.08 0.14)	0.10 *** (0.07 0.13)	0.03 * (0.00 0.05)	0.01 (-0.01 0.03)
Indicador de Nível socioeconômico (Escola)								0.28 *** (0.17 0.39)	0.16 *** (0.08 0.24)	0.12 *** (0.05 0.20)
Desempenho em linguagem início do ano									0.58 *** (0.55 0.61)	0.47 *** (0.43 0.50)
Desempenho em matemática início do ano										0.18 *** (0.15 0.21)
Random Effects										
σ^2	0.74	0.74	0.68	0.68	0.68	0.65	0.64	0.64	0.37	0.33
τ_{00}	0.40 ESC	0.09 ESC	0.08 ESC	0.08 ESC	0.07 ESC	0.06 ESC	0.04 ESC	0.02 ESC	0.01 ESC	0.01 ESC
ICC	0.35	0.11	0.10	0.10	0.09	0.08	0.06	0.04	0.02	0.02
N	75 ESC	75 ESC	75 ESC	75 ESC	69 ESC	69 ESC	69 ESC	69 ESC	69 ESC	69 ESC
Observations	2712	2712	2641	2641	2084	1793	1781	1781	1780	1774
Marginal R ² / Conditional R ²	0.000 / 0.348	0.189 / 0.280	0.248 / 0.324	0.252 / 0.327	0.187 / 0.263	0.228 / 0.288	0.256 / 0.303	0.288 / 0.314	0.592 / 0.602	0.637 / 0.644

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

ANEXO - DADOS OBSERVADOS - Regressão multinível - Fatores associados à aprendizagem no início da Educação Básica obrigatória (1º ano da pré-escola) considerando a rede pública e privada.

Variável dependente: Desenvolvimento em Matemática no início do 2º ano da pré-escola.

	Matemática Modelo Nulo	Matemática Modelo 1	Matemática Modelo 2	Matemática Modelo 3	Matemática Modelo 4	Matemática Modelo 5	Matemática Modelo 6	Matemática Modelo 7	Matemática Modelo 8	Matemática Modelo 9
(Intercepto)	0.96 *** (0.69 1.22)	0.20 * (0.04 0.37)	-6.19 *** (-7.00 -5.39)	-6.22 *** (-7.02 -5.41)	-6.25 *** (-7.16 -5.35)	-5.95 *** (-6.92 -4.98)	-5.96 *** (-6.91 -5.01)	-5.85 *** (-6.80 -4.90)	-1.08 ** (-1.81 -0.35)	-0.64 (-1.36 0.09)
Escola:Privada		2.11 *** (1.81 2.40)	2.09 *** (1.82 2.36)	2.09 *** (1.82 2.36)	2.12 *** (1.83 2.41)	1.77 *** (1.49 2.05)	1.35 *** (1.07 1.62)	0.67 ** (0.19 1.15)	0.39 * (0.06 0.72)	0.42 * (0.08 0.75)
Idade (meses)			0.12 *** (0.10 0.13)	0.12 *** (0.10 0.13)	0.12 *** (0.10 0.14)	0.12 *** (0.10 0.13)	0.12 *** (0.10 0.14)	0.12 *** (0.10 0.14)	0.04 *** (0.02 0.05)	0.03 *** (0.02 0.04)
Sexo:Masculino				0.03 (-0.07 0.14)	-0.00 (-0.12 0.12)	0.04 (-0.09 0.17)	0.01 (-0.11 0.14)	0.01 (-0.11 0.14)	0.08 (-0.01 0.17)	0.12 ** (0.03 0.21)
Creche:Frequentou					0.04 (-0.10 0.17)	-0.01 (-0.16 0.15)	-0.01 (-0.16 0.13)	-0.02 (-0.17 0.12)	-0.07 (-0.17 0.04)	-0.07 (-0.17 0.04)
Indicador de Ambiente de Aprendizagem em Casa						0.19 *** (0.14 0.23)	0.12 *** (0.07 0.17)	0.11 *** (0.06 0.16)	0.04 * (0.01 0.08)	0.03 (-0.00 0.07)
Indicador de Nível socioeconômico (Indivíduo)							0.22 *** (0.17 0.27)	0.21 *** (0.16 0.26)	0.08 *** (0.04 0.11)	0.06 *** (0.03 0.10)
Indicador de Nível socioeconômico (Escola)								0.29 ** (0.12 0.47)	0.05 (-0.07 0.17)	0.04 (-0.09 0.16)
Desempenho em matemática início do ano									0.77 *** (0.73 0.81)	0.68 *** (0.64 0.72)
Desempenho em linguagem início do ano										0.23 *** (0.17 0.29)
Random Effects										
σ^2	2.07	2.07	1.89	1.89	1.91	1.88	1.81	1.81	0.95	0.92
τ_{00}	1.31 ESC	0.28 ESC	0.21 ESC	0.21 ESC	0.15 ESC	0.10 ESC	0.07 ESC	0.04 ESC	0.02 ESC	0.02 ESC
ICC	0.39	0.12	0.10	0.10	0.07	0.05	0.04	0.02	0.02	0.02
N	75 ESC	75 ESC	75 ESC	75 ESC	69 ESC	69 ESC	69 ESC	69 ESC	69 ESC	69 ESC
Observations	2712	2712	2641	2641	2084	1793	1781	1781	1775	1774
Marginal R ² /	0.000 /	0.221 /	0.284 /	0.284 /	0.229 /	0.269 /	0.306 /	0.321 /	0.647 /	0.658 /
Conditional R ²	0.387	0.313	0.355	0.355	0.284	0.307	0.332	0.337	0.654	0.665

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

ANEXO - DADOS OBSERVADOS - Regressão multinível - Fatores associados à aprendizagem no início da Educação Básica obrigatória (1º ano da pré-escola) considerando os estratos baixa, média e alta mensalidade.

Variável dependente: Desenvolvimento em Linguagem no início do 2º ano da pré-escola.

	Linguagem Modelo Nulo	Linguagem Modelo 1	Linguagem Modelo 2	Linguagem Modelo 3	Linguagem Modelo 4	Linguagem Modelo 5	Linguagem Modelo 6	Linguagem Modelo 7	Linguagem Modelo 8	Linguagem Modelo 9
(Intercepto)	0.55 *** (0.40 0.70)	0.15 *** (0.08 0.23)	-3.53 *** (-4.00 -3.05)	-3.45 *** (-3.93 -2.97)	-3.47 *** (-4.01 -2.92)	-3.24 *** (-3.81 -2.67)	-3.25 *** (-3.82 -2.69)	-3.18 *** (-3.74 -2.62)	-0.66 ** (-1.11 -0.21)	-0.08 (-0.51 0.35)
Escola:Baixa mens.		0.62 *** (0.35 0.90)	0.57 *** (0.32 0.83)	0.57 *** (0.32 0.82)	0.51 ** (0.20 0.82)	0.34 * (0.05 0.63)	0.18 (-0.10 0.46)	-0.17 (-0.49 0.16)	-0.04 (-0.28 0.20)	-0.08 (-0.30 0.14)
Escola:Média mens.		0.88 *** (0.65 1.11)	0.92 *** (0.70 1.14)	0.93 *** (0.71 1.14)	0.87 *** (0.58 1.16)	0.72 *** (0.44 1.00)	0.57 *** (0.29 0.84)	0.14 (-0.22 0.49)	0.02 (-0.25 0.28)	-0.05 (-0.29 0.20)
Escola:Alta mens.		1.51 *** (1.32 1.69)	1.47 *** (1.30 1.64)	1.47 *** (1.30 1.64)	1.43 *** (1.22 1.65)	1.18 *** (0.96 1.39)	0.93 *** (0.72 1.14)	0.37 * (0.01 0.74)	0.13 (-0.14 0.41)	0.03 (-0.22 0.28)
Idade (meses)			0.07 *** (0.06 0.08)	0.07 *** (0.06 0.08)	0.07 *** (0.06 0.08)	0.07 *** (0.06 0.08)	0.07 *** (0.06 0.08)	0.07 *** (0.06 0.08)	0.02 *** (0.01 0.03)	0.01 *** (0.01 0.02)
Sexo:Masculino				-0.11 *** (-0.18 -0.05)	-0.13 *** (-0.20 -0.06)	-0.11 ** (-0.19 -0.04)	-0.12 ** (-0.20 -0.05)	-0.12 ** (-0.20 -0.05)	0.00 (-0.06 0.06)	-0.00 (-0.06 0.05)
Creche:Frequentou					0.01 (-0.07 0.09)	-0.03 (-0.12 0.06)	-0.03 (-0.12 0.05)	-0.04 (-0.13 0.05)	-0.07 (-0.13 0.00)	-0.06 * (-0.13 -0.00)
Indicador de Ambiente de Aprendizagem em Casa						0.11 *** (0.08 0.14)	0.07 *** (0.04 0.10)	0.07 *** (0.04 0.10)	0.03 * (0.00 0.05)	0.02 (-0.00 0.04)
Indicador de Nível socioeconômico (Indivíduo)							0.11 *** (0.08 0.14)	0.10 *** (0.07 0.13)	0.03 * (0.00 0.05)	0.01 (-0.01 0.03)
Indicador de Nível socioeconômico (Escola)								0.20 *** (0.09 0.32)	0.13 ** (0.05 0.22)	0.11 ** (0.03 0.18)
Desempenho em linguagem início do ano									0.58 *** (0.55 0.61)	0.46 *** (0.43 0.50)
Desempenho em matemática início do ano										0.18 *** (0.15 0.21)
Random Effects										
σ^2	0.74	0.74	0.68	0.68	0.68	0.65	0.64	0.64	0.37	0.33
τ_{00}	0.40 ESC	0.05 ESC	0.04 ESC	0.04 ESC	0.04 ESC	0.03 ESC	0.02 ESC	0.02 ESC	0.01 ESC	0.01 ESC
ICC	0.35	0.07	0.05	0.05	0.06	0.05	0.04	0.02	0.02	0.02
N	75 ESC	75 ESC	75 ESC	75 ESC	69 ESC	69 ESC	69 ESC	69 ESC	69 ESC	69 ESC
Observations	2712	2712	2641	2641	2084	1793	1781	1781	1780	1774
Marginal R ² /	0.000 /	0.216 /	0.277 /	0.281 /	0.214 /	0.255 /	0.279 /	0.293 /	0.593 /	0.638 /
Conditional R ²	0.348	0.267	0.316	0.319	0.258	0.288	0.305	0.310	0.602	0.644

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

ANEXO - DADOS OBSERVADOS - Regressão multinível - Fatores associados à aprendizagem no início da Educação Básica obrigatória (1º ano da pré-escola) considerando os estratos baixa, média e alta mensalidade.

Variável dependente: Desenvolvimento em Matemática no início do 2º ano da pré-escola.

	Matemática Modelo Nulo	Matemática Modelo 1	Matemática Modelo 2	Matemática Modelo 3	Matemática Modelo 4	Matemática Modelo 5	Matemática Modelo 6	Matemática Modelo 7	Matemática Modelo 8	Matemática Modelo 9
(Intercepto)	0.96 *** (0.69 1.22)	0.21 ** (0.07 0.34)	-6.16 *** (-6.96 -5.37)	-6.18 *** (-6.98 -5.38)	-6.23 *** (-7.13 -5.32)	-5.93 *** (-6.89 -4.96)	-5.95 *** (-6.90 -5.00)	-5.87 *** (-6.82 -4.92)	-1.08 ** (-1.81 -0.36)	-0.63 (-1.36 0.09)
Escola:Baixa mens.		1.37 *** (0.90 1.85)	1.27 *** (0.85 1.69)	1.27 *** (0.85 1.69)	1.46 *** (0.96 1.96)	1.17 *** (0.70 1.64)	0.85 *** (0.40 1.29)	0.48 (-0.04 0.99)	0.38 * (0.00 0.75)	0.45 * (0.06 0.83)
Escola:Média mens.		1.65 *** (1.24 2.05)	1.73 *** (1.36 2.09)	1.72 *** (1.36 2.09)	1.89 *** (1.42 2.36)	1.59 *** (1.13 2.05)	1.27 *** (0.83 1.71)	0.80 ** (0.23 1.38)	0.40 (-0.01 0.82)	0.40 (-0.02 0.82)
Escola:Alta mens.		2.69 *** (2.36 3.01)	2.62 *** (2.34 2.90)	2.62 *** (2.34 2.90)	2.55 *** (2.20 2.90)	2.13 *** (1.79 2.47)	1.63 *** (1.30 1.97)	1.04 *** (0.45 1.62)	0.38 (-0.04 0.81)	0.36 (-0.08 0.79)
Idade (meses)			0.12 *** (0.10 0.13)	0.12 *** (0.10 0.13)	0.12 *** (0.10 0.14)	0.12 *** (0.10 0.13)	0.12 *** (0.10 0.14)	0.12 *** (0.10 0.14)	0.04 *** (0.02 0.05)	0.03 *** (0.02 0.04)
Sexo:Masculino				0.03 (-0.07 0.14)	-0.01 (-0.13 0.11)	0.03 (-0.10 0.16)	0.01 (-0.12 0.13)	0.01 (-0.12 0.13)	0.08 (-0.01 0.17)	0.12 ** (0.03 0.21)
Creche:Frequentou					0.03 (-0.10 0.16)	-0.02 (-0.17 0.13)	-0.02 (-0.17 0.12)	-0.03 (-0.17 0.12)	-0.07 (-0.17 0.04)	-0.07 (-0.17 0.04)
Indicador de Ambiente de Aprendizagem em Casa						0.19 *** (0.14 0.23)	0.12 *** (0.07 0.16)	0.11 *** (0.06 0.16)	0.04 * (0.01 0.08)	0.03 (-0.00 0.07)
Indicador de Nível socioeconômico (Indivíduo)							0.22 *** (0.17 0.27)	0.21 *** (0.16 0.26)	0.08 *** (0.04 0.12)	0.06 *** (0.03 0.10)
Indicador de Nível socioeconômico (Escola)								0.22 * (0.04 0.40)	0.05 (-0.08 0.18)	0.05 (-0.09 0.18)
Desempenho em matemática início do ano									0.77 *** (0.73 0.81)	0.68 *** (0.64 0.73)
Desempenho em linguagem início do ano										0.23 *** (0.17 0.29)
Random Effects										
σ^2	2.07	2.08	1.89	1.89	1.91	1.88	1.81	1.81	0.95	0.92
τ_{00}	1.31 ESC	0.16 ESC	0.11 ESC	0.11 ESC	0.10 ESC	0.06 ESC	0.05 ESC	0.03 ESC	0.02 ESC	0.02 ESC
ICC	0.39	0.07	0.05	0.05	0.05	0.03	0.02	0.02	0.02	0.02
N	75 ESC	75 ESC	75 ESC	75 ESC	69 ESC	69 ESC	69 ESC	69 ESC	69 ESC	69 ESC
Observations	2712	2712	2641	2641	2084	1793	1781	1781	1775	1774
Marginal R2 /	0.000 /	0.245 /	0.309 /	0.309 /	0.245 /	0.284 /	0.316 /	0.323 /	0.647 /	0.658 /
Conditional R2	0.387	0.300	0.346	0.346	0.282	0.307	0.333	0.335	0.654	0.665

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

ANEXO - Rubin's Rule - Regressão multinível - Fatores associados à estratificação no início da Educação Básica obrigatória (1º ano da pré-escola) considerando a rede pública e privada. Linguagem 1ª Onda

term	estimate	std.error	statistic	df	p.value
(Intercept)	-4.25	0.26	-16.56	813.47	0.00
Escola:Privada	0.49	0.09	5.64	90.73	0.00
Idade (meses)	0.07	0.00	15.57	1299.29	0.00
Sexo:Masculino	-0.16	0.03	-4.63	1023.73	0.00
Creche:Frequentou	0.10	0.05	1.96	13.75	0.07
AAC	0.07	0.02	3.78	11.35	0.00
NSE	0.13	0.02	7.86	32.30	0.00

ANEXO - Rubin's Rule - Regressão multinível - Fatores associados à estratificação no início da Educação Básica obrigatória (1º ano da pré-escola) considerando a rede pública e privada. Matemática 1ª Onda

term	estimate	std.error	statistic	df	p.value
(Intercept)	-6.73	0.34	-19.54	1431.21	0.00
Escola:Privada	1.18	0.12	10.12	90.95	0.00
Idade (meses)	0.11	0.01	17.56	1718.53	0.00
Sexo:Masculino	-0.01	0.05	-0.22	1365.04	0.83
Creche:Frequentou	0.14	0.05	2.56	76.56	0.01
AAC	0.09	0.03	2.79	8.26	0.02
NSE	0.17	0.03	5.39	8.61	0.00

ANEXO - Rubin's Rule - Regressão multinível - Fatores associados à aprendizagem no início da Educação Básica obrigatória (1º ano da pré-escola) considerando a rede pública e privada.. Linguagem 2ª Onda

term	estimate	std.error	statistic	df	p.value
(Intercept)	-0.18	0.19	-0.95	1999.10	0.34
Escola:Privada	0.13	0.10	1.22	83.23	0.22
Idade (meses)	0.01	0.00	4.27	2598.16	0.00
Sexo:Masculino	-0.02	0.02	-0.91	2550.39	0.36
Creche:Frequentou	-0.04	0.03	-1.31	23.85	0.20
AAC	0.01	0.01	1.18	34.99	0.25
NSE	0.01	0.01	1.11	174.90	0.27
NSE (Escola)	0.11	0.04	2.71	136.94	0.01
Matemática 1ª Onda	0.18	0.01	16.57	2614.64	0.00
Linguagem 1ª Onda	0.45	0.01	30.86	2601.83	0.00

ANEXO - Rubin's Rule - Regressão multinível - Fatores associados à aprendizagem no início da Educação Básica obrigatória (1º ano da pré-escola) considerando a rede pública e privada.. Matemática 2ª Onda

term	estimate	std.error	statistic	df	p.value
(Intercept)	-0.57	0.31	-1.87	2519.08	0.06
Escola:Privada	0.42	0.20	2.12	35.15	0.04
Idade (meses)	0.03	0.01	4.87	2593.53	0.00
Sexo:Masculino	0.10	0.04	2.67	2522.50	0.01
Creche:Frequentou	-0.03	0.05	-0.68	62.14	0.50
AAC	0.02	0.02	1.06	61.64	0.29
NSE	0.05	0.02	2.75	52.87	0.01
NSE (Escola)	0.09	0.08	1.19	44.64	0.24
Matemática 1ª Onda	0.67	0.02	37.13	2526.80	0.00
Linguagem 1ª Onda	0.23	0.02	9.25	2452.39	0.00